An Information Theoretic Proof of the Maximum Entropy Spectrum in n Dimensions

by

Dimitris N. Politis
Purdue University

Technical Report #92-10

Department of Statistics
Purdue University

February 1992

△

An Information Theoretic Proof of the Maximum Entropy Spectrum in n Dimensions

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Abstract

A well-known result of Burg (1969) and Woods (1976) identifies a Gaussian discrete Markov random field with autocovariance specified on a finite part L of the 2-dimensional integer lattice, as the random field with Maximum Entropy among all discrete random fields with same autocovariance values on L. In this correspondence, an intuitive information theoretic proof of a version of the Maximum Entropy theorem for random fields in n dimensions is presented, where n is any positive integer.

Index Terms. Autoregression, Gaussian random fields, Maximum Entropy Spectrum.

I. Introduction

Suppose $\{X(\mathbf{t}), \mathbf{t} = (t_1, t_2, \ldots, t_n) \in \mathbf{Z}^n\}$ is a homogeneous (strictly stationary, shift invariant) random field in n dimensions, with $n \in \mathbf{N}$, i.e., a collection of real-valued random variables $X(\mathbf{t})$, defined on a probability space $(\Omega, \mathcal{A}, P)$, and indexed by the variable $\mathbf{t} \in \mathbf{Z}^n$. The random field $\{X(\mathbf{t})\}$ will be assumed to possess finite second moments, and its autocovariance function will be denoted as $\gamma_X(\mathbf{s}) = Cov\{X(\mathbf{t}), X(\mathbf{t}+\mathbf{s})\}$, for any $\mathbf{t}, \mathbf{s} \in \mathbf{Z}^n$.

It will be assumed that the marginal distribution function of $X(\mathbf{t})$ possesses a density f with respect to Lebesgue measure ν on $\mathbf{R}$. In general, it will be assumed that for any finite set $A \subset \mathbf{Z}^n$, with cardinality m, the marginal distribution function of $\{X(\mathbf{t}), \mathbf{t} \in A\}$ possesses a density with respect to the product Lebesgue measure ν^m on $\mathbf{R}^m$, and the generic notation for this density will be f. The entropy of the set of variables $\{X(\mathbf{t}), \mathbf{t} \in A\}$ is defined by [4]

$$H(X(\mathbf{t}), \mathbf{t} \in A) = -\int f(X(\mathbf{t}), \mathbf{t} \in A) \log f(X(\mathbf{t}), \mathbf{t} \in A) d\nu^m \tag{1}$$

Similarly, the conditional entropy of the random variable $X(\mathbf{u})$ given $\{X(\mathbf{t}), \mathbf{t} \in A\}$ is defined by

$$H(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A) = -\int f(X(\mathbf{u}), X(\mathbf{t}), \mathbf{t} \in A) \log f(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A) d\nu^{m+1} \tag{2}$$

where $f(X(\mathbf{u})|X(\mathbf{t}), \mathbf{t} \in A)$ is the conditional density of $X(\mathbf{u})$ given the values of $\{X(\mathbf{t}), \mathbf{t} \in A\}$.

The random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is said to have an entropy rate

$$h_X = \lim_{N\to\infty} \frac{H(X(\mathbf{t}), \mathbf{t} \in C_N)}{|C_N|} \tag{3}$$

provided the limit exists, where C_N is the cube of points $\mathbf{t} \in \mathbf{Z}^n$ whose coordinates satisfy $|t_k| \leq N$, and has cardinality $|C_N| = (2N+1)^n$. For homogeneous random fields it can be shown [7] that the limit exists in $[-\infty, \infty]$, and can alternatively be calculated as

$$h_X = \lim_{N\to\infty} \frac{H(X(\mathbf{t}), \mathbf{t} \in C_N^+)}{N^n} \tag{4}$$

where C_N^+ is the cube of points $\mathbf{t} \in \mathbf{Z}^n$ whose coordinates satisfy $0 < t_k \leq N$.

Following [11], we define *quarter-plane purely non-deterministic* (q.p.n.d.) zero-mean, homogeneous random fields $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$, as the fields that possess the causal-type unilateral innovations representation

$$X(\mathbf{t}) = Z(\mathbf{t}) + \sum_{\mathbf{u} \in Past(\mathbf{t})} h(\mathbf{u}-\mathbf{t}) Z(\mathbf{u}) \tag{5}$$

where $Past(\mathbf{t}) \equiv \{\mathbf{u} : u_j \leq t_j, \forall j\} - \{\mathbf{t}\}$, and $\sum_{\mathbf{u} \in Past(\mathbf{0})} |h(\mathbf{u})|^2 < \infty$. The random field $\{Z(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a zero-mean 'white noise', i.e., $E[Z(\mathbf{t})Z(\mathbf{u})] = 0$, if $\mathbf{t} \neq \mathbf{u}$, and $EZ(\mathbf{t})^2 = \sigma^2$. The term 'quarter-plane' is chosen because, in the case $n = 2$, $Past(\mathbf{t})$ is the quarter-plane 'south-west' of point $\mathbf{t}$.

Now with $\mathbf{u} = (u_1, \ldots, u_n)$, define a notion of the 'past' of point $\mathbf{t}$ in the direction n by

$$Past_n(\mathbf{t}) = \cup_{j=1}^{n} \{\mathbf{u} : u_j < t_j, u_k = t_k, j < k \leq n\} \tag{6}$$

$$= \{\mathbf{u} : u_n < t_n\} \cup \{\mathbf{u} : u_n = t_n, u_{n-1} < t_{n-1}\} \cup \cdots \cup \{\mathbf{u} : u_n = t_n, u_{n-1} = t_{n-1}, \ldots, u_2 = t_2, u_1 < t_1\}$$

The $Past_n(\mathbf{t})$ can be thought of as an 'augmented' half-space in the direction n, because it mainly consists of the half-space $\{\mathbf{u} : u_n < t_n\}$. Of course, because the labelling and numbering of coordinates is arbitrary, it is straightforward to define the 'past' $Past_j(\mathbf{t})$ in the direction j, for any $j = 1, \ldots, n$ (cf. [10]).

By analogy to [8], we define *half-space purely non-deterministic* (h.p.n.d.) zero-mean, homogeneous random fields $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$, as the fields that possess the causal-type innovations representation

$$X(\mathbf{t}) = Z(\mathbf{t}) + \sum_{\mathbf{u} \in Past_n(\mathbf{t})} h(\mathbf{u}-\mathbf{t}) Z(\mathbf{u}) \tag{7}$$

where $\{Z(\mathbf{t})\}$ is a zero-mean, white noise, and $\sum_{\mathbf{u} \in Past_n(\mathbf{0})} |h(\mathbf{u})|^2 < \infty$.

It is apparent that the class of q.p.n.d. random fields is a subset of the class of h.p.n.d. random fields. In other words, a random field possessing the q.p.n.d. representation (5), also possesses the h.p.n.d. representation (7), with $h(\mathbf{u}) = 0$, for $\mathbf{u} \in Past_n(\mathbf{0}) - Past(\mathbf{0})$.

Now let

$$\Pi_p^{(r)}(\mathbf{t}) = \{\mathbf{u} : 0 < d_p(\mathbf{t}, \mathbf{u}) \leq r\} \tag{8}$$

be the set of r-close neighbors of point $\mathbf{t}$, and

$$L_p^{(r)}(\mathbf{t}) = \Pi_p^{(r)}(\mathbf{t}) \cap Past_n(\mathbf{t}) \tag{9}$$

where 'closeness' is measured by the l_p distance in $\mathbf{Z}^n$ given by $d_p(\mathbf{t}, \mathbf{u}) = (\sum_j |t_j - u_j|^p)^{1/p}$, for $1 \leq p < \infty$, and $d_\infty(\mathbf{t}, \mathbf{u}) = \sup_j |t_j - u_j|$. For simplicity, we will focus on convex neighborhoods of the $\Pi_p^{(r)}$ type, although more general neighborhoods are possible.

The Burg-Woods theorem [2], [12] indicates that, at least for the case $n = 2$, a Gaussian discrete Markov random field with autocovariance specified on the set $L = L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$ is the random field with Maximum Entropy among all discrete random fields with same autocovariance values on L. Note that, due to the symmetry of the autocovariance ($\gamma_X(\mathbf{s}) = \gamma_X(-\mathbf{s})$), if γ_X is specified on $L = L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$, then it is also specified on $\Pi_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$. In the next section, an information theoretic proof of the Maximum Entropy theorem for random fields in n dimensions is presented, where n is any positive integer.

The method of proof used will require to limit the search for the Maximum Entropy Spectrum to the class of half-space purely non-deterministic homogeneous random fields, and the process with Maximum Entropy will be found to be a Gaussian, h.p.n.d., linear unilateral autoregression $\{W(\mathbf{t})\}$ satisfying for $\mathbf{t} \in \mathbf{Z}^n$,

$$W(\mathbf{t}) + \sum_{\mathbf{u} \in L_p^{(r)}(\mathbf{t})} a(\mathbf{t} - \mathbf{u}) W(\mathbf{t} - \mathbf{u}) = \tilde{Z}(\mathbf{t}) \tag{10}$$

where $\{\tilde{Z}(\mathbf{t})\}$ is some zero-mean, Gaussian, white noise field, and $a(\mathbf{t})$ are some constants.

It is important to point out however that the classes of q.p.n.d. or h.p.n.d. homogeneous random fields satisfying equation (5) or (7) are quite large, and both contain *all* linear autoregressive stationary processes in n dimensions that are given by a recursion of the type

$$U(\mathbf{t}) + \sum_{\mathbf{u} \in \Pi_p^{(r)}(\mathbf{t})} a(\mathbf{t} - \mathbf{u}) U(\mathbf{t} - \mathbf{u}) = \tilde{Z}(\mathbf{t}) \tag{11}$$

for any $p \geq 1$ and $r \in \mathbf{N}$, under the usual condition that the spectral density of the process $\{U(\mathbf{t})\}$ is bounded. The Gaussian Markov random field that is the solution of the Burg-Woods Maximum Entropy theorem turns out [12] to be a Gaussian linear autoregression $\{U(\mathbf{t})\}$ satisfying (11).

In general, a sufficient condition for a homogeneous random field to be quarter-plane purely non-deterministic is [11] that the random field possesses a continuous and strictly positive spectral density on the compact n-torus $(-\pi, \pi]^n$. This is also a sufficient condition for a homogeneous random field to be half-space purely non-deterministic, since a q.p.n.d. random field is also a h.p.n.d. random field. For the case $n = 2$, a necessary and sufficient condition for a homogeneous random field to be half-space purely non-deterministic, is [8] that the random field possesses a spectral density whose logarithm is absolutely integrable over the compact n-torus $(-\pi, \pi]^n$. This latter result shows the direct analogy of h.p.n.d. random fields to stationary sequences ($n = 1$) that are purely non-deterministic, that is, sequences whose Wold decomposition has no deterministic component.

II. Information Theoretic Proof of the Maximum Entropy Spectrum in n Dimensions

The proof given below is patterned after the information theoretic proof of the maximum entropy spectrum in one dimension by Choi and Cover [3], to which it reduces in the case $n = 1$.

Theorem 1 *Consider a h.p.n.d., zero-mean, Gaussian homogeneous random field* $\{W(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$*, that has the unilateral autoregressive representation (10) for some* $p \geq 1$ *and* $r \in \mathbf{N}$.

If the random field $\{W(\mathbf{t})\}$ *satisfies the constraints*

$$\gamma_W(\mathbf{t}) = E[W(\mathbf{u})W(\mathbf{u}+\mathbf{t})] = \beta(\mathbf{t}) \tag{12}$$

for $\mathbf{t} \in L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$*, and some constants* $\beta(\mathbf{t})$*, then* $\{W(\mathbf{t})\}$ *has maximum entropy rate among all h.p.n.d. homogeneous random fields* $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ *that have the following unilateral representation, where* $\{Z(\mathbf{t})\}$ *is a zero-mean, white noise, with variance* $EZ(\mathbf{t})^2 = \sigma^2$, $\sum_{\mathbf{u}\in Past_n(\mathbf{0})} |h(\mathbf{u})|^2 < \infty$*, and* C *is some constant,*

$$X(\mathbf{t}) = C + Z(\mathbf{t}) + \sum_{\mathbf{u}\in Past_n(\mathbf{t})} h(\mathbf{u}-\mathbf{t})Z(\mathbf{u}) \tag{13}$$

and whose autocovariance satisfies the constraints

$$\gamma_X(\mathbf{t}) = Cov\{X(\mathbf{u}), X(\mathbf{u}+\mathbf{t})\} = \beta(\mathbf{t}) \tag{14}$$

for $\mathbf{t} \in L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$.

Proof. Let $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ be a h.p.n.d. homogeneous random field possessing the representation (13), with autocovariance satisfying (14). Also let $\{Y(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ be given by

$$Y(\mathbf{t}) = Z^*(\mathbf{t}) + \sum_{\mathbf{u}\in Past_n(\mathbf{t})} h(\mathbf{u}-\mathbf{t})Z^*(\mathbf{u}) \tag{15}$$

where the $h(\mathbf{u})$ coefficients are exactly the ones appearing in (13), and $\{Z^*(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ is a zero-mean, *Gaussian* white noise, with $E[Z^*(\mathbf{t})Z^*(\mathbf{u})] = 0$, if $\mathbf{t} \neq \mathbf{u}$, and $EZ^*(\mathbf{t})^2 = \sigma^2$. It is apparent that the autocovariance $\gamma_Y(\mathbf{t}) = E[Y(\mathbf{u})Y(\mathbf{u}+\mathbf{t})]$ is identical to $\gamma_X(\mathbf{t})$, for all $\mathbf{t} \in \mathbf{Z}^n$.

Let an integer $N > r$, and consider the positive cube C_N^+. Looking at the multivariate probability densities of $(X(\mathbf{t}), \mathbf{t} \in C_N^+)$, and $(Y(\mathbf{t}), \mathbf{t} \in C_N^+)$, it is immediate [4] that

$$H(X(\mathbf{t}), \mathbf{t} \in C_N^+) \leq H(Y(\mathbf{t}), \mathbf{t} \in C_N^+) \tag{16}$$

since the mean-zero, multivariate normal distribution of $(Y(\mathbf{t}), \mathbf{t} \in C_N^+)$ has maximum entropy among all multivariate distributions with same covariance matrix.

It is apparent that an ordering (sometimes called the 'lexicographical' ordering [6], [7]) of the elements of C_N^+ is induced by defining $\mathbf{s} << \mathbf{t}$ if $|C_N^+ \cap Past_n(\mathbf{s})| < |C_N^+ \cap Past_n(\mathbf{t})|$, for $\mathbf{s}, \mathbf{t} \in C_N^+$, where $|A|$ denotes the cardinality of set A. Using this ordering, the density $f(Y(\mathbf{t}), \mathbf{t} \in C_N^+)$ can be expanded in a chain rule as follows.

$$f(Y(\mathbf{t}), \mathbf{t} \in C_N^+) = \prod_{\mathbf{t} \in C_N^+} f(Y(\mathbf{t}) | Y(\mathbf{u}), \mathbf{u} \in C_N^+ \cap Past_n(\mathbf{t})) \tag{17}$$

This immediately yields the following chain rule for entropies

$$H(Y(\mathbf{t}), \mathbf{t} \in C_N^+) = \sum_{\mathbf{t} \in C_N^+} H(Y(\mathbf{t}) | Y(\mathbf{u}), \mathbf{u} \in C_N^+ \cap Past_n(\mathbf{t})) \tag{18}$$

Using the fact that $H(A|B, C) \leq H(A|B)$ for any three random elements A, B, C ('conditioning reduces entropy' [4]), we have that

$$\sum_{\mathbf{t} \in C_N^+} H(Y(\mathbf{t}) | Y(\mathbf{u}), \mathbf{u} \in C_N^+ \cap Past_n(\mathbf{t})) \leq \sum_{\mathbf{t} \in C_N^+} H(Y(\mathbf{t}) | Y(\mathbf{u}), \mathbf{u} \in C_N^+ \cap L_p^{(r)}(\mathbf{t})) \tag{19}$$

Note that, due to homogeneity of the $\{Y(\mathbf{t})\}$ field, the computation of the right-hand-side of (19) requires knowledge only of the multivariate distribution of the set of variables $\{Y(\mathbf{0})\} \cup \{Y(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(\mathbf{0})\}$. However, the distribution of $\{Y(\mathbf{0})\} \cup \{Y(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(\mathbf{0})\}$ is zero-mean, multivariate normal, with covariance matrix completely determined by (14). By constraint (12), this multivariate normal distribution is *identical* to the distribution of the set of variables $\{W(\mathbf{0})\} \cup \{W(\mathbf{t}), \mathbf{t} \in L_p^{(r)}(W(\mathbf{0}))\}$.

It follows that

$$\sum_{\mathbf{t} \in C_N^+} H(Y(\mathbf{t}) | Y(\mathbf{u}), \mathbf{u} \in C_N^+ \cap L_p^{(r)}(\mathbf{t})) = \sum_{\mathbf{t} \in C_N^+} H(W(\mathbf{t}) | W(\mathbf{u}), \mathbf{u} \in C_N^+ \cap L_p^{(r)}(\mathbf{t})) \tag{20}$$

Observe that by a chain rule expansion of the density $f(W(\mathbf{t}), \mathbf{t} \in C_N^+)$ one obtains

$$H(W(\mathbf{t}), \mathbf{t} \in C_N^+) = \sum_{\mathbf{t} \in C_N^+} H(W(\mathbf{t}) | W(\mathbf{u}), \mathbf{u} \in C_N^+ \cap Past_n(\mathbf{t})) \tag{21}$$

Note also that due to (10) and the h.p.n.d. assumption, the random field $\{W(\mathbf{t})\}$ possesses the following Markov-type property (see also [10])

$$f(W(\mathbf{t}) | W(\mathbf{u}), \mathbf{u} \in Past_n(\mathbf{t})) = f(W(\mathbf{t}) | W(\mathbf{u}), \mathbf{u} \in L_p^{(r)}(\mathbf{t})) \tag{22}$$

for any $\mathbf{t} \in \mathbf{Z}^n$. Hence, the sum in the right-hand-side of (21) differs from that in the right-hand-side of (20) only in the contributions of the $r(2N)^{n-1}$ points that are r-close to the sides (boundary) of C_N^+ that are parallel to the direction n, i.e., the points whose $L_p^{(r)}$ neighborhood is not completely included in C_N^+. However, the entropy rate h_W can be explicitly calculated (cf. [1], [6], [7]) as $h_W = H(W(\mathbf{0}) | W(\mathbf{u}), \mathbf{u} \in L_p^{(r)}(\mathbf{0}))$ and is finite; thus the contribution to the entropy rate of each of these $r(2N)^{n-1}$ points must be bounded. It follows that

$$H(Y(\mathbf{t}), \mathbf{t} \in C_N^+) \leq H(W(\mathbf{t}), \mathbf{t} \in C_N^+) + O(rN^{n-1}) \tag{23}$$

Combining (23) with (16) we can write

$$H(X(\mathbf{t}), \mathbf{t} \in C_N^+) \leq H(Y(\mathbf{t}), \mathbf{t} \in C_N^+) \leq H(W(\mathbf{t}), \mathbf{t} \in C_N^+) + O(rN^{n-1}) \tag{24}$$

Dividing by N^n and taking limits in the above asymptotic inequality it is seen that $h_X \leq h_W$, and the theorem is proven. □

To apply the theorem in practice, one could possibly use data to estimate the autocovariance $\gamma_X(\mathbf{s})$ of $X(\mathbf{t})$, for $\mathbf{s}$ in the set $L = L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$, (for some chosen p and r), and then extrapolate the remaining autocovariance values by invoking the Maximum Entropy principle and fitting a unilateral autoregression of the form (10) to the data by Yule-Walker type equations or some other method. However, this intuitive approach might fail for dimensions $n > 1$ (cf. [5]), although it is well known that it works fine in the case $n = 1$. In general, the abovementioned estimation and fitting procedure will be valid provided the estimated autocovariance values coincide with those of some true positive definite autocovariance function [12], that is, if the estimated autocovariance admits a positive definite extension to the whole of $\mathbf{Z}^n$. For a review of algorithms related to the practical problem of multidimensional spectral estimation, see [9] and the references therein.

As a further point, consider the possibility that $\gamma_X(\mathbf{s})$ is known (measured) only in a subset $L^* \subset L = L_p^{(r)}(\mathbf{0}) \cup \{\mathbf{0}\}$. The following theorem can be proved by the same arguments used in the proof of Theorem 1.

Theorem 2 *Consider a h.p.n.d., zero-mean, Gaussian homogeneous random field* $\{W(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$*, that has, for any* $\mathbf{t} \in \mathbf{Z}^n$*, the unilateral autoregressive representation*

$$W(\mathbf{t}) + \sum_{\mathbf{u}} a(\mathbf{t} - \mathbf{u}) W(\mathbf{t} - \mathbf{u}) = \tilde{Z}(\mathbf{t}) \tag{25}$$

where the summation is over all $\mathbf{u}$ *such that* $\mathbf{u} - \mathbf{t} \in L^*$*, and* $\{\tilde{Z}(\mathbf{t})\}$ *is a zero-mean, Gaussian, white noise field.*

If the random field $\{W(\mathbf{t})\}$ *satisfies the constraints (12) for* $\mathbf{t} \in L^*$*, and some constants* $\beta(\mathbf{t})$*, then it has maximum entropy rate among all h.p.n.d. homogeneous random fields* $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{Z}^n\}$ *that have the unilateral representation (13), and whose autocovariance satisfies the constraints (14) for* $\mathbf{t} \in L^*$. □

III. Conclusions

An information theoretic proof of a general Maximum Entropy theorem (Theorem 2) for random fields in n dimensions was presented, where n is any positive integer.

The interpretation of Theorems 1 and 2 is that a Gaussian, half-space purely non-deterministic (h.p.n.d.), linear unilateral autoregression $\{W(\mathbf{t})\}$, has maximum entropy among all h.p.n.d. homogeneous random fields with autocovariance which is the same as the autocovariance of $\{W(\mathbf{t})\}$ on a finite part L of the discrete lattice $\mathbf{Z}^n$.

The restriction of the search to the class of h.p.n.d. homogeneous random fields makes it possible to have a unilateral autoregressive representation of the form (10) for $\{W(\mathbf{t})\}$.

Acknowledgement

The author is grateful to Professor Tom Cover of Stanford University for many helpful discussions.

References

[1] Anastassiou, D. and Sakrison, D. J. (1982), Some results regarding the entropy rate of random fields, *IEEE Trans. Inform. Theory*, 28, 340-343.

[2] Barnard, T. and Burg, J.P. (1969), Analytical studies of techniques for the computation of high-resolution wavenumber spectra, *Advanced Array Research Report # 9*, Texas Instruments, Inc..

[3] Choi, B.S. and Cover, T.M. (1984), An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum, *Proc. IEEE*, vol. 72, 8, 1094-1095.

[4] Cover, T.M. and Thomas, J. (1991), *Elements of Information Theory*, John Wiley, New York.

[5] Dickinson, B.W. (1980), Two-dimensional Markov spectral estimates need not exist, *IEEE Trans. Inform. Theory*, 26, 120-121.

[6] Föllmer, H. (1973), On entropy and information gain in random fields, *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, 26, 207-217.

[7] Georgii, H.-O. (1988), *Gibbs measures and phase transitions,* DeGruyter Studies in Mathematics, vol. 9, Berlin.

[8] Korezlioglu, H. and Loubaton, P. (1986), Spectral Factorization of Wide Sense Stationary Processes on $\mathbf{Z}^2$, *J. Multivariate Analysis*, 19, 24-47.

[9] McClellan, J.H. (1982), Multidimensional spectral estimation, *Proc. IEEE*, vol. 70, 9, 1029-1039.

[10] Politis, D.N. (1992), Markov chains in many dimensions, Technical Report No. 92-02, Department of Statistics, Purdue University.

[11] Tjøstheim, D. (1978), Statistical spatial series modelling, *Adv. Applied Prob.*, 10, 130-154.

[12] Woods, J.W. (1976), Two-dimensional Markov spectral estimation, *IEEE Trans. Inform. Theory*, 22, 552-559.