PARAMETER ESTIMATION IN 2D FIELDS *

Nikhil Balram *José M. F. Moura*

LASIP, Depart. Elect. and Comp. Eng., Carnegie Mellon University, Pittsburgh, PA, 15213-3890.

Abstract

In this paper, we deal with the problem of estimating the parameters of noncausal finite lattice Gauss Markov random fields. We show how the structure of the potential matrix (the inverse of the field covariance matrix) can be used to specify the valid parameter space and formulate a computationally practical Maximum Likelihood estimation procedure. We also provide a modification that enables this to generate accurate parameter estimates from noisy data.

1. Introduction

An important class of statistical models for 2D fields of spatially distributed data is the family of noncausal finite lattice Gauss Markov random fields (GMRF's). A GMRF may be represented on a finite $N \times M$ lattice through the minimum mean square error representation [1] written compactly as

$$A\vec{X} = \vec{e}, \qquad \vec{e} \sim \mathcal{N}(\vec{0}, \sigma^2 A), \tag{1}$$

where $\vec{X}$ and $\vec{e}$ are the lexicographically ordered vectors of the NM variables in respectively, the field X, and the correlated driving noise field e, and A collects the AR field coefficients. It is apparent that A, which we call the *potential matrix*, is the (scaled) inverse of the field covariance matrix. In [2], the structure and properties of the potential matrix were used to derive a framework for recursive processing of noncausal fields. This has been applied, for example, in image processing, to do optimal recursive image enhancement and image compression, e.g., [3]. Such applications require as a preliminary step the estimation of the parameters of the noncausal GMRF. This is the problem we are concerned with here.

Optimal parameter estimation in the case of finite lattice noncausal GMRF's is handicapped by two problems: the constrained nature of the parameter space, and the large cost of computing the likelihood function and its gradient. Apart from special cases such as fields defined on periodic lattices [4], the specification of the parameter space is difficult. The likelihood function and its derivatives are extremely costly to compute because of the intractable nature of the partition function which takes the form of $|A|$ for GMRF's. This problem has lead to the use of suboptimal techniques such as the coding method [5] and pseudolikelihood [6] which are not guaranteed to produce a result that lies in the valid parameter space.

In this paper, we use the structure of the potential matrix A to specify the valid parameter space, in section 2, and to obtain a computationally practical Maximum Likelihood (ML) estimation procedure, in section 3. In section 4, we modify the ML procedure to enable us to obtain accurate estimates from noisy observations. We present experimental results in section 5, and in section 6, the conclusions. For proofs and a more detailed discussion of the results presented here, as well as for additional results, see [7].

2. Valid parameter space

The valid parameter space for a nondegenerate GMRF is defined as the region in which the field covariance matrix is positive definite. From (1), this translates to the condition $\sigma^2 > 0$, which we assume henceforth, and the requirement that the potential matrix A be positive definite. The latter condition may be expressed through the constraint on the smallest eigenvalue of A

$$\lambda_{min}(A) > 0 \tag{2}$$

which defines the valid region for the field potentials or parameters. When the eigenstructure of A is available as a function of the field parameters, (2) provides the exact description of the parameter space. When this is not available, we decompose A into symmetric components with known eigenvalues and bound the space using these.

*The work reported here was partially supported by ONR grant # N0014-91-J-1001.

First order fields: In the case of first order (nearest neighbor) fields, with parameters β_h and β_v representing, respectively, the strength of the horizontal and vertical nearest neighbor interactions, the potential matrix A can be expressed as

$$A = I_N \otimes B + S_N \otimes C \tag{3}$$

where $\otimes$ represents the Kronecker product,

$$B = I_M - \beta_h S_M, \qquad C = -\beta_v I_M, \tag{4}$$

I_K is the $K \times K$ identity matrix, and $S_K, (K = N, M)$, depends on the boundary conditions (b.c.). The eigenvalues of A are given by

$$\lambda_{i,j}(A) = 1 - \beta_v \lambda_i(S_N) - \beta_h \lambda_j(S_M), \tag{5}$$

for $1 \leq i \leq N$, $1 \leq j \leq M$, where $\{\lambda_i(S_K)\}$ are the eigenvalues of $S_K, (K = N, M)$. Examples of b.c., usually from the PDE literature, are Dirichlet (Free), Asymmetric Neumann, Symmetric Neumann, and Periodic, see [2]. We call the resulting fields, *Dirichlet*, *variational*, *symmetric*, and *periodic*, respectively, and consider all but the last named here. The eigenvalues of S_K are given (for $1 \leq k \leq K$) by

$$\lambda_k(S_K) = \begin{cases} 2\cos\frac{k\pi}{K+1} & \text{for Dirichlet fields} \\ 2\cos\frac{(k-1)\pi}{K} & \text{for variational fields} \end{cases} \tag{6}$$

and for symmetric fields,

$$\lambda_k(S_K) = \begin{cases} 2\cosh\theta_{k1} & \text{when } k = k1 \\ 2\cosh\theta_{k2} & \text{when } k = k2 \\ -2\cosh\theta_{k1} & \text{when } k = k3 \\ -2\cosh\theta_{k2} & \text{when } k = k4 \\ 2\cos\theta_k & \text{for all other } k \end{cases} \tag{7}$$

where $\{\theta_k : 1 \leq k \leq K, \; k \neq k1, k2, k3, k4\}$ are the $K-4$ solutions of

$$2[(K+1)\theta - k\pi] + 4\tan^{-1}\frac{3\sin 2\theta}{1 - 3\cos 2\theta} = 0, \tag{8}$$

and $\pm\theta_{k1}$, $\pm\theta_{k2}$ are the 4 roots of

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

Using (6) and (7) in (5) and applying (2), we obtain the region in (β_h, β_v) space that corresponds to the valid parameter space for each of these fields.

1. The parameter space for the first order Dirichlet field is defined by

$$|\beta_v|\cos\frac{\pi}{N+1} + |\beta_h|\cos\frac{\pi}{M+1} < \frac{1}{2}. \tag{10}$$

2. The parameter space for the first order variational field is defined by the 4 inequalities:

$$\beta_v + \beta_h < \frac{1}{2}, \tag{11}$$

$$\beta_v - \beta_h\cos\frac{\pi}{M} < \frac{1}{2}, \tag{12}$$

$$\beta_h - \beta_v\cos\frac{\pi}{N} < \frac{1}{2}, \tag{13}$$

$$-\beta_v\cos\frac{\pi}{N} - \beta_h\cos\frac{\pi}{M} < \frac{1}{2}. \tag{14}$$

3. The parameter space for the first order symmetric field is defined by

$$|\beta_v|\cosh\theta_N^* + |\beta_h|\cosh\theta_M^* < \frac{1}{2}, \tag{15}$$

with $\theta_K^* = \max\{\theta_{k1}, \theta_{k2}\}$, where $\pm\theta_{k1}$, $\pm\theta_{k2}$ are the 4 solutions of (9).

Higher order fields: In [7], we show that the potential matrix A_p for a pth order field may be defined recursively as

$$A_p = A_{p-1} + \Delta_p, \tag{16}$$

where A_{p-1} is the potential matrix for the $(p-1)$th order field, and Δ_p contains the coefficients for the new interactions and can be decomposed using a special class of matrices known as *interaction matrices* whose eigenvalues are derived analytically. For general fields, where $\lambda_{min}(A_p)$ may not be available, (16) enables us to derive a recursively generated bound on the parameter space through the relationship

$$\lambda_{min}(A_p) \geq \lambda_{min}(A_{p-1}) + \lambda_{min}(\Delta_p) \tag{17}$$

and the eigenvalues of the interaction matrices, see [7].

3. Parameter estimation

Let θ denote a vector containing the field potentials. Maximum Likelihood (ML) estimation requires the minimization, with respect to (θ, σ^2), of the negative log likelihood

$$L(X/\theta, \sigma^2) = \frac{1}{2}[\ln\sigma^2 - \frac{1}{NM}\ln|A(\theta)| + \frac{1}{NM\sigma^2}\vec{X}^T A(\theta)\vec{X},$$

which is obtained by scaling $-\ln P(X)$ by $(1/NM)$ and removing the constants.

First order fields: For a first order field, $\theta = (\beta_h, \beta_v)$. Using the structure of A from (3), (4), the quadratic $\vec{X}^T A \vec{X}$ can be expressed as

$$\frac{1}{NM}\vec{X}^T A \vec{X} = C_0^x - 2(\beta_h C_h^x + \beta_v C_v^x), \tag{18}$$

where

$$C_0^x = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} x_{i,j}^2 \tag{19}$$

$$C_h^x = \frac{1}{NM}\left(\sum_{i=1}^{N}\sum_{j=1}^{M-1} x_{i,j}x_{i,j+1} + C_{h.b.}^x\right) \tag{20}$$

$$C_v^x = \frac{1}{NM}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{M} x_{i,j}x_{i+1,j} + C_{v.b.}^x\right) \tag{21}$$

with correction terms $C^x_{h.b.}$ and $C^x_{v.b.}$ given by

$$C^x_{h.b.} = \begin{cases} 0 & \text{Dirichlet} \\ \frac{1}{2}\sum_{i=1}^{N}(x_{i,1}^2 + x_{i,M}^2) & \text{variational} \\ \sum_{i=1}^{N}(x_{i,1}x_{i,2} + x_{i,M-1}x_{i,M}) & \text{symmetric} \end{cases}$$

$$C^x_{v.b.} = \begin{cases} 0 & \text{Dirichlet} \\ \frac{1}{2}\sum_{j=1}^{M}(x_{1,j}^2 + x_{N,j}^2) & \text{variational} \\ \sum_{j=1}^{M}(x_{1,j}x_{2,j} + x_{N-1,j}x_{N,j}) & \text{symmetric} \end{cases}$$

C^x_0 may be interpreted as the sample power, C^x_h and C^x_v as, respectively, the sample horizontal and vertical nearest neighbor correlations.

The partition function term in $L(.)$, $\ln|A|$, can be computed at a cost of $O(NM)$ using the eigenvalues of A from (5) since the determinant of a matrix is given by the product of its eigenvalues. Using this fact along with the relationship in (18), we can express $L(.)$ as

$$L(X/\beta_h, \beta_v, \sigma^2) = \tfrac{1}{2}\ln\sigma^2 + \tfrac{1}{2\sigma^2}[C^x_0 - 2\beta_h C^x_h - 2\beta_v C^x_v] - \tfrac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln(1 - \beta_v\lambda_i(S_N) - \beta_h\lambda_j(S_M)). \quad (22)$$

For an efficient search of the parameter space, we need the gradient of $L(.)$,

$$\frac{\partial L}{\partial \beta_h} = \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\lambda_j(S_M)}{(1 - \beta_v\lambda_i(S_N) - \beta_h\lambda_j(S_M))} - \frac{C^x_h}{\sigma^2}$$

$$\frac{\partial L}{\partial \beta_v} = \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\lambda_i(S_N)}{(1 - \beta_v\lambda_i(S_N) - \beta_h\lambda_j(S_M))} - \frac{C^x_v}{\sigma^2}$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}[C^x_0 - 2\beta_h C^x_h - 2\beta_v C^x_v].$$

The partial derivative with respect to σ^2 provides an explicit solution for the ML estimate,

$$\widehat{\sigma^2} = C^x_0 - 2(\widehat{\beta}_h C^x_h + \widehat{\beta}_v C^x_v), \quad (23)$$

which substituted into (22) reduces $L(.)$ to

$$L(X/\beta_h, \beta_v, \sigma^2(\beta_h, \beta_v)) = \frac{1}{2}\ln(C^x_0 - 2\beta_h C^x_h - 2\beta_v C^x_v) - \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln(1 - \beta_v\lambda_i(S_N) - \beta_h\lambda_j(S_M)),$$

a function of β_h, β_v, only, where for convenience we have removed the constant term.

Conjugate gradient search: The ML estimates are obtained using the Polak–Ribiere conjugate gradient search method, e.g., [8]. The parameter space constraints are incorporated in the bracketing procedure that precedes each line minimization. Since the computation associated with each evaluation of $L(.)$ or its gradient is only $O(NM)$, the ML estimation procedure is computationally practical.

Higher order fields: For higher order fields, the same conjugate gradient search procedure may be used with the appropriate parameter space constraints. However, the computation of $\ln|A|$ in evaluating $L(.)$ is a major concern. A further complication is the gradient of L, with the partial derivative with respect to each field parameter β_τ having the form

$$\frac{\partial L}{\partial \beta_\tau} = \frac{1}{2NM}\text{trace}(A^{-1}\Delta^\tau_p) - \frac{C^x_\tau}{\sigma^2}, \quad (24)$$

where C^x_τ is the sample covariance of lag τ. This expression is computationally intractable because it requires the inversion of a $NM \times NM$ matrix. We solve both these problems by exploiting the recursive framework derived in [2]. In particular, the recursive framework is used to compute the partition function expeditely and to provide the basis for a fast Monte Carlo sampling procedure that estimates the gradient. The gradient estimation is made possible by reformulating (24) as the difference between an expectation and a sample value:

$$\frac{\partial L}{\partial \beta_\tau} = \frac{E\{C^x_\tau\} - C^x_\tau}{\sigma^2}, \quad (25)$$

where $E\{.\}$ is the expectation operator. At any point in the parameter space, an estimate of the gradient is obtained from (25) by applying the recursive structure in [2] to rapidly synthesize a set of field samples which are then used to estimate $E\{C^x_\tau\}$. Consequently, we are able to formulate and implement a computationally practical and accurate conjugate gradient search procedure to provide the ML parameter estimates for noncausal fields of arbitrary order, see [7] for details.

4. Estimation from noisy data

In many applications, e.g., image enhancement [3], the field parameters have to estimated from noisy data. We present below a modification to the ML estimation procedure from section 3 that provides accurate parameter estimates from observations

$$y_{i,j} = x_{i,j} + n_{i,j}, \qquad 1 \le i \le N,\ 1 \le j \le M \quad (26)$$

contaminated by additive white Gaussian noise (AWGN) $n_{i,j}$ that is independent of the field variables $x_{i,j}$. The noise variance σ_n^2 is assumed to be known or estimated separately from the flat or less busy regions of the field.

The terms C^y_0, C^y_h, C^y_v, are defined for the observed field Y using the appropriate definitions in (19)–(21). Substituting (26) into (19)–(21), and approximating the summations as expectations, we get for first order fields:

$$C^y_0 \approx C^x_0 + \sigma_n^2, \quad (27)$$

$$C^y_h \approx C^x_h + \begin{cases} 0 & \text{Dirichlet or symmetric} \\ \frac{\sigma_n^2}{M} & \text{variational} \end{cases} \quad (28)$$

Table 1: Results for ML estimation from 30 samples of 1st order Dirichlet field with $N = M = 32$, $\beta_h = 0.27$, $\beta_v = 0.225$, $\sigma^2 = 1.0$, and MML estimation from samples corrupted by AWGN with $\sigma_n^2 = 0.196$ (SNR $\approx$ 10db).

	noiseless (ML)	noisy (MML)
$<\widehat{\beta_h}>$	0.268026	0.264472
$<\widehat{\beta_v}>$	0.225250	0.228506
$<\widehat{\sigma^2}>$	0.995672	0.999266
var($\widehat{\beta_h}$)	0.000302	0.000338
var($\widehat{\beta_v}$)	0.000317	0.000328
var($\widehat{\sigma^2}$)	0.001871	0.003625

$$C_v^y \approx C_v^x + \begin{cases} 0 & \text{Dirichlet or symmetric} \\ \frac{\sigma_n^2}{N} & \text{variational} \end{cases} \quad (29)$$

The approximations in (27)–(29) should be used with caution at low SNR (0db and lower) where instability may occur as the noise variance becomes the dominant component in the observation sample power.

The modified Maximum Likelihood (MML) procedure for estimation from the noisy data consists of two steps. In the first step, the noiseless data terms C_0^x, C_h^x, C_v^x, in the negative log likelihood function (22) and its gradient are approximated from their noisy data counterparts using (27)–(29). In the second step, the ML procedure from section 3 is used as before to obtain the parameter estimates. Note that in the case of Dirichlet fields the approximations in (27)–(29) are valid for arbitrary order, see [7].

5. Experimental results

Thirty 32×32 samples of a first order Dirichlet field were generated recursively using the framework in [2]. We computed ML estimates from each sample using the procedure described in section 3. The mean ($< . >$) and variance (var(.)) of the estimates for each parameter are provided in Table 1. As a comparison, the pseudolikelihood (PL) estimate [6] was computed for each sample. In more than a quarter of the cases (8 out of 30), the PL estimates were outside the valid region defined by (10). This emphasises the fact that the pseudolikelihood estimate is not guaranteed to produce valid parameters. This is a major concern when the fields are highly correlated, as is often the case with images, and the true parameters are close to the boundary of the valid region.

Continuing the experiment, white Gaussian noise with variance $\sigma_n^2 = 0.196$ (chosen to get SNR $\approx$ 10db) was added to each sample and the MML procedure was used to estimate the parameters from each noisy sample. The mean and variance of the estimates are given in Table 1. The MML procedure is shown to work well, providing accurate estimates of the parameters at moderately low SNR (10db). In [3], MML is applied as part of an image enhancement algorithm.

6. Conclusions

The structure of the potential matrix provides the means for specification of the valid parameter space and computationally feasible expressions for the likelihood function and its gradient, leading to the formulation of a computationally practical ML estimation procedure. A simple approximation allows this to be used with noisy data. For clarity, most of the results presented here are stated explicitly in terms of first order fields. See [7] for the extension to fields of arbitrary order.

References

[1] J.W. Woods. Two-dimensional discrete Markovian fields. *IEEE Trans. Inform. Theory*, IT-18:232–240, 1972.

[2] J. M. F. Moura and N. Balram. Recursive structure of noncausal Gauss Markov random fields. *IEEE Trans. Inform. Theory*, IT-38(2), March 1992.

[3] N. Balram and J. M. F. Moura. Rapid enhancement and compression of image data. In *Proceedings Visual Communications and Image Processing'91*, volume 1606, pages 374–385. SPIE, November 1991.

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

[5] J. E. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Stat. Soc., B*, 36(2):192–236, 1974.

[6] J. E. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.

[7] N. Balram and J. M. F. Moura. Noncausal Gauss Markov random fields: Parameter structure and estimation. Technical Report 91-03, LASIP, ECE, Carnegie Mellon University, April 1991. 43 pages.

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