NISS

Hierarchical Image Reconstruction Using Markov Random Fields

Nancy J. McMillan and L. Mark Berliner

Technical Report Number 15
May, 1994

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Hierarchical Image Reconstruction Using Markov Random Fields

NANCY J. MCMILLAN and L. MARK BERLINER
National Institute of Statistical Sciences, USA
and Ohio State University, USA

Abstract

At least since the seminal work of Geman and Geman (1984), Markov random fields have served as Bayesian models for images in computational reconstruction from degraded, observational data. We consider fully Bayesian hierarchical models, in which one stage of the hierarchy is a Markov random field. This field is parameterized by a scalar λ which in principle controls the degree of smoothness of random images generated by the model. We then develop a class of conjugate priors for λ. Based on data, posterior inferences are developed employing familiar versions of Gibbs' Sampling. The basic hierarchical model is extended to an exchangeable model in which several "similiar" images are to be reconstructed. Artificial examples, motivated by imaging problems arising in materials science and stereology, are presented.

KEY WORDS: Bayesian analysis; Conjugate priors; Exchangeability; Gibbs sampling.

1 Introduction

Our focus is on image reconstruction via classes of Bayesian smoothing models. The Bayesian paradigm provides a convenient mechanism for balancing faithfulness to observational data with prior beliefs, incorporated via Markov random field priors, in the underlying smoothness of images. Specifically, the true image, $\boldsymbol{\theta}$, is modeled as a realization of some Markov random field. We observe X, a degraded version of $\boldsymbol{\theta}$, where the degradation process is modeled via some probability density, f, given $\boldsymbol{\theta}$. Inference regarding $\boldsymbol{\theta}$ is based on the posterior distribution of $\boldsymbol{\theta}$. Many other researchers have considered such models: Geman and Geman, [7], Marroquin Mittier and Poggio, [10], Besag, [3], and several articles from [13], to cite only a few.

We consider three extensions to the above, previously considered, classes of models. The first is incorporation of uncertainty regarding the signal to noise ratio of our degradation

model. We assume the observed data, X, follows the model, $f(X|\boldsymbol{\theta}, \sigma^2)$ where parameter σ^2 controls this degradation rate. Most previous work has assumed σ^2 to be known, or estimated in some ad hoc fashion. The second extension is the addition of a second stage prior. (For a general review of hierarchical models, see [2].) Moving from a Bayesian image reconstruction model to a hierarchical Bayesian model is a natural step which has been suggested, but not pursued, in [10]. Towards this end, we assume that the aforementioned prior and degradation model are specified given λ, a hyperparameter. Specifically, we assume $\boldsymbol{\theta}$ is a realization of some Markov random field, denoted by $\pi_1(\boldsymbol{\theta}|\lambda)$. The hyperparameter λ roughly controls the amount of smoothness to be expected in images generated from this model.

Substantial flexibility in Bayesian modeling accrues from the use of hierarchical models. One interesting possibility, pursued here, involves the simultaneous estimation of several images deemed to be a priori "similiar" or *exchangeable*. This is the third extension we suggest to previous work. Specifically, assume that our data comprises n degraded images, $X_1, \ldots, X_n$. Given true images, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$, the X_i are independent and each has distribution $f(\cdot|\boldsymbol{\theta}_i, \sigma^2)$.

In the simple hierarchical model for a single image, the complete joint prior distribution to be used here is of the form

$$\pi(\boldsymbol{\theta}, \lambda, \sigma^2) = \pi_{1.1}(\boldsymbol{\theta}|\lambda)\pi_{1.2}(\sigma^2)\pi_2(\lambda). \tag{1}$$

The components $\pi_2(\lambda)$ and $\pi_{1.2}(\sigma^2)$ are chosen to be appropriate conjugate priors, as described in Section 2. For the exchangeable case, we simply assume that the $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n$ are viewed as an *i.i.d.* sample, conditioned on λ, from prior $\pi_{1.1}(\boldsymbol{\theta}|\lambda)$.

Hierarchical models introduce additional computational complexities into an already difficult problem. In the hierarchical classes of models proposed, a computationally infeasible normalizing constant present in the first stage Markov random field prior, $\pi_{1.1}(\boldsymbol{\theta}|\lambda)$, is actually used in the definition of the conjugate prior for λ. To deal with this problem, we suggest use of a pseudo-posterior based on approximating this normalizing constant. Importance sampling is employed to obtain this approximation. (See Section 3.) Fortunately, this costly step in the analysis procedure depends only on the particular model proposed, $\pi(\boldsymbol{\theta}|\lambda)$, not on the images observed. This idea is akin to the Griddy Gibbs' algorithm proposed in [15].

We emphasize estimation of the expected posterior image rather than the more commonly used maximum a posteriori image or "MAP" estimate. This permits the natural use of posterior variance estimates to reflect the credibility of the estimated image. Another argument, though less compelling perhaps, is based on decision theoretic concerns. The loss function corresponding to Bayesian optimality of a maximum a posteriori estimator is 0-1 loss. This is an unattractive loss function for a visually-oriented problem such as image reconstruction where a one pixel error and a 100 pixel error are not equally detrimental. Of course, we do not necessarily wish to imply that "sum of pixel-wise squared errors" loss is clearly appropriate in general either.

The size of typical imaging problems forces numerically approximate Bayesian analyses. The analysis here will involve Markov Chain Monte Carlo; Specifically, we employ Gibbs' Sampling. A complete review of this method is omitted for brevity. The reader is referred to Gelfand and Smith (1990). The Gibbs' Sampler produces a sequence of dependent, simulated observations from the joint posterior distribution. In the simple hierarchical model case here this posterior is denoted by $\pi(\boldsymbol{\theta}, \lambda, \sigma^2|x)$. Note that the formulation implies that

$$\pi(\boldsymbol{\theta}, \lambda, \sigma^2|x) \propto f(x|\boldsymbol{\theta}, \sigma^2)\pi_{1.1}(\boldsymbol{\theta}|\lambda)\pi_{1.2}(\sigma^2)\pi_2(\lambda). \tag{2}$$

The key to implementing a Gibbs' Sampler is the construction of the so-called "full conditionals;" Namely, the conditional distributions of each variable of the posterior given all other variables. The purpose of Section 2 is to develop the components of the model and display the construction of the required full conditionals. (For simplicity, the formulation is presented in the simple hierarchical case. The corresponding representations for the exchangeable model are simple modifications.) Concepts needed in this development include graphs, Gibbs' distributions, Markov random fields, and appropriate conjugate priors for Markov random fields.

Section 4 contains the results of computer experiments based on both simple hierarchical and exchangeable model examples. Mean and variance estimates presented were obtained via Gibbs' sampling, as described by Geman and Geman in [7]. We compare estimates from hierarchical and non-hierarchical models. Section 5 presents some concluding remarks.

2 Model Formulation

There are two main tasks in this section. The first is the development of the four components, $f, \pi_{1.1}, \pi_{1.2}$, and π_2, on the righthand side of (2). These components are explicitly given in Models 2.1, 2.3, 2.4, and 2.5, respectively. The second task is description of the resulting "full conditionals" used by the Gibbs' Sampler. Basic representations of these conditionals are readily obtained from (2). Employing the notation popularized by Gelfand and Smith (1990), we see that

$$[\boldsymbol{\theta}|\lambda, \sigma^2, x] \propto f(x|\boldsymbol{\theta}, \sigma^2)\pi_{1.1}(\boldsymbol{\theta}|\lambda). \tag{3}$$

$$[\lambda|\boldsymbol{\theta}, \sigma^2, x] \propto \pi_{1.1}(\boldsymbol{\theta}|\lambda)\pi_2(\lambda). \tag{4}$$

$$[\sigma^2|\boldsymbol{\theta}, \lambda, x] \propto f(x|\boldsymbol{\theta}, \sigma^2)\pi_{1.2}(\sigma^2). \tag{5}$$

We assume that the true image $\boldsymbol{\theta}$ is a pixel image representable as $\boldsymbol{\theta} = \{\theta_s | s \in S\}$ where S is an index set for the pixels, possibly, the $r \times c$ pixel lattice, described by

$$S = \{(i, j) | 1 \leq i \leq r, 1 \leq j \leq c\}. \tag{6}$$

Each individual pixel, θ_s is an element of the discrete set $\Theta = \{0, 1, \cdots, L-1\}$ where L is some fixed constant. Generally, conditional on $\boldsymbol{\theta}$, our observed image, X, will be an observation from a random process, also indexed by S, $X = \{X_s | s \in S\}$. The probability density function (probability mass function for the discrete case) of X given $\boldsymbol{\theta}$ is denoted by $f(X|\boldsymbol{\theta})$. A convenient and standard assumption on the degradation process is that conditional on the parameter $\boldsymbol{\theta}$ and for each $s \in S$, X_s is independent of $X_t, t \neq s$ and the distribution of X_t depends only on θ_t. Hence, the density $f(X|\boldsymbol{\theta})$ can be expressed as $f(X|\boldsymbol{\theta}) = \prod_{s \in S} f_s(X_s|\theta_s)$. For our examples, we will also assume that $f_s(\cdot|\cdot)$ is a normal density and does not depend on s. This is summarized as follows:

Model 2.1 ***Gaussian Error Model***

Let the pixel degradation model be

$$X_s|\theta_s, \sigma^2 \sim N(\theta_s, \sigma^2) \tag{7}$$

where the X_s's are assumed independent.

2.1 Graphs, Gibbs' Distributions, and Markov Random Fields

As stated in Section 1, $\boldsymbol{\theta}$ will be modeled via a Markov random field. We briefly introduce a variety of necessary definitions. The reader is refered to [7] for more complete discussion.

Let $S = \{s_1, ..., s_N\}$ be a finite set of sites. $\mathcal{N} = \{\eta_s \subset S | s \in S\}$ is a *neighborhood system* for S provided the following two conditions hold:

1. $s \notin \eta_s$, $\forall s \in S$ and

2. $t \in \eta_s \Longrightarrow s \in \eta_t$, $\forall t, s \in S$.

If these conditions hold then η_s is the *neighborhood* of $s \in S$. The pair $G = (S, \mathcal{N})$ is a *graph.* We note the correspondence between this definition of a graph and the more standard edge and vertex definition.

To complete the notion of a graph, we define cliques. The set $C \subset S$ is a *clique* if for all $t, s \in C$ such that $t \neq s$, $s \in \eta_t$. Singleton sets, $C = \{s\}$ for all $s \in S$, are also assumed to be cliques. Let $\mathcal{C}$ be the set of all cliques. Consider the following example which conveniently suggests the role of S as an index set for the pixels of an image.

Example 2.2 ***(2-dimensional lattice, nearest neighbors)***
Let $S = \{(i,j) | 1 \leq i \leq r, 1 \leq j \leq c\}$ be the $r \times c$ two-dimensional lattice. For all $s = (i,j) \in S$ let

$$\eta_s = \{(k,l) \in S | 0 < (k-i)^2 + (l-j)^2 \leq 1\}. \tag{8}$$

Constructing $\mathcal{N} = \{\eta_s | s \in S\}$, the pair $G = (S, \mathcal{N})$ is a graph. Under this neighborhood structure, all cliques are one of three types, singletons, horizontally adjacent doubletons, or vertically adjacent doubletons.

To define joint Gibbs' distributions on graphs, let $\boldsymbol{\theta} = \{\theta_s | s \in S\}$ be a family of discrete random variables having some joint probability mass function, denoted P. Assume that all the θ_s have a common finite state space. That is let $\Theta = \{0, 1, ..., L-1\}$ be the possible values of θ_s for all $s \in S$. This implies that the state space (or configuration space) of the family of random variables on the graph can be defined as:

$$\Omega = \{\boldsymbol{\omega} = (\omega_{s_1}, \ldots, \omega_{s_N}) | \omega_s \in \Theta, \forall s \in S\}. \tag{9}$$

This construction of the joint configuration space as the cross product of the *identical, finite* state spaces of the individual random variables, θ_s, is in general overly restrictive in the definition of a Gibbs' distribution. It will, however, be sufficient and notationally convenient for our problem.

A joint Gibbs distribution is defined with respect to a graph structure, $G = (S, \mathcal{N})$. The neighborhood structure of the graph places a restriction on the form of the probability mass function. By definition, P is a Gibbs' distribution with respect to G, if it can be written in the form:

$$P(\boldsymbol{\theta} = \boldsymbol{\omega}) = Z^{-1} \exp\{\sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega})\} \tag{10}$$

for all $\boldsymbol{\omega} \in \Omega$.

The functions $V_C(\cdot)$, $C \in \mathcal{C}$ are called potentials. Each $V_C(\boldsymbol{\omega})$ is written as if the function depended upon the entire vector $\boldsymbol{\omega}$, but this is not the case. Each one is only allowed to depend on the part of the vector $\boldsymbol{\omega}$ contained in C. The constant Z is called the partition function. It is the normalizing constant, defined by

$$Z = \sum_{\boldsymbol{\omega} \in \Omega} \exp\{\sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega})\}. \tag{11}$$

This is an extremely flexible model for the distribution of a family of discrete random variables. Redefining the set of sites, S, and/or the neighborhood structure, $\mathcal{N}$, enables one to write any discrete joint distribution as a Gibbs' distribution with respect to some graph, $G = (S, \mathcal{N})$.

By definition, this same family of random variables, $\boldsymbol{\theta}$, is a *discrete Markov random field* on the graph G if the following two conditions hold:

1. $P\{\boldsymbol{\theta} = \boldsymbol{\omega}\} > 0 \ \forall \boldsymbol{\omega} \in \Omega$ and

2. $P(\theta_t = \omega_t | \{\theta_s = \omega_s : s \neq t\}) = P(\theta_t = \omega_t | \{\theta_s = \omega_s : s \in \eta_t\})$.

These conditions are known as the positivity condition and the Markov condition, respectively. For more information on discrete Markov random fields in this context see [4], [5], [7], and [9]. The crucial relationship between Gibbs' distributions and Markov random fields with respect to graphs is known as the *Hammersly-Clifford Theorem*: Let G be a

graph. Then $\boldsymbol{\theta}$ is a discrete Markov random field with respect to G *if and only if* the joint distribution of $\boldsymbol{\theta}$ is a discrete Gibbs distribution with respect to G. For further discussion see [4] or [14]. The form of the conditional distributions resulting from a Gibbs' distribution with respect to a graph, G, as defined by (10), is

$$P(\theta_t = \omega_t | \{\theta_s = \omega_s : s \neq t\}) = Z_t^{-1} \exp\{\sum_{C \in \mathcal{C}(t)} V_C(\boldsymbol{\omega})\}, \tag{12}$$

where

$$\mathcal{C}(t) = \{C \in \mathcal{C} | t \in C\} \tag{13}$$

and

$$Z_t = \sum_{\omega_t \in \Theta} \exp\{\sum_{C \in \mathcal{C}(t)} V_C(\boldsymbol{\omega})\}. \tag{14}$$

Note that (12) implies the full conditional distribution of θ_t depends only on the neighbors of t,

$$P(\theta_t = \omega_t | \{\theta_s = \omega_s : s \neq t\}) = P(\theta_t = \omega_t | \{\theta_s = \omega_s : s \in \eta_t\}). \tag{15}$$

Two key uses of the Hammersly-Clifford Theorem to Bayesian image reconstruction are as follows. First, the dual interpretation of joint Gibbs' distributions as Markov random fields provides a convenient mechanism for modeling local dependence structures among a large collection of variables. Second, the high-dimensional joint distribution of $\boldsymbol{\theta}$ can be readily represented in a fashion, based on (12), permitting rapid simulation as needed in Gibbs' Sampling.

Consider the following model which will serve as both our example Markov random field/Gibbs' distribution and as the prior distribution for our image reconstruction model.

Model 2.3 ***First Stage Prior for*** $\boldsymbol{\theta}$

Let $G = (S, \mathcal{N})$ *be as defined in Example 2.2. Let* $\boldsymbol{\theta} = \{\theta_s | s \in S\}$ *be a family of discrete random variables with p.m.f.* $\pi_{1.1}$. *Let* $\Theta = \{0, 1, \cdots, L-1\}$ *be the sample space of* θ_s *for all* $s \in S$. *The p.m.f.* $\pi_{1.1}$ *is defined to be*

$$\pi_{1.1}(\boldsymbol{\theta} = \boldsymbol{\omega} | \lambda) = Z(\lambda)^{-1} \exp\{\lambda \sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega})\} \tag{16}$$

where

$$V_C(\boldsymbol{\omega}) = \begin{cases} 0 & \text{if } |C| = 1 \\ 1 & \text{if } \omega_s = \omega_t, s, t \in C \\ -1 & \text{if } \omega_s \neq \omega_t, s, t \in C \end{cases} \tag{17}$$

and

$$Z(\lambda) = \sum_{\boldsymbol{\omega} \in \Omega} \exp\{\lambda \sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega})\}. \tag{18}$$

For notational convenience in subsequent formulae we define

$$U(\boldsymbol{\omega}) = \sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega}). \tag{19}$$

$U(\cdot)$ *is known as the energy function. The full conditionals are:*

$$\pi_{1.1}(\theta_t = \omega_t | \{\theta_s = \omega_s : s \neq t\}, \lambda) = Z_t^{-1} \exp\{\lambda \sum_{s' \in \eta_t} V_{\{s',t\}}(\boldsymbol{\omega})\}, \tag{20}$$

where

$$Z_t = \sum_{\omega_t \in \Theta} \exp\{\lambda \sum_{s' \in \eta_t} V_{\{s',t\}}(\boldsymbol{\omega})\}. \tag{21}$$

Combination of this Markov random field prior and the likelihood resulting from conditionally independent data, such as that suggested by Model 2.1, via Bayes theorem results in a posterior which is again a Markov random field with respect to the same graph, $G = (S, \mathcal{N})$. (It should be noted that, in general, the neighborhood structure of the posterior is not necessarily the same as the neighborhood structure of the prior.) First, combination of the Gaussian likelihood and the first stage prior results in

$$\pi_G(\boldsymbol{\theta} = \boldsymbol{\omega} | \lambda, \sigma^2, x) = Z(\lambda, \sigma^2, x)^{-1} \exp\{\lambda \sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega}) - \frac{1}{2\sigma^2} \|x - \boldsymbol{\omega}\|^2\} \tag{22}$$

where $V_C(\boldsymbol{\omega})$ is as defined in Model 2.3 and

$$Z(\lambda, \sigma^2, x) = \sum_{\boldsymbol{\omega} \in \Omega} \exp\{\lambda \sum_{C \in \mathcal{C}} V_C(\boldsymbol{\omega}) - \frac{1}{2\sigma^2} \|x - \boldsymbol{\omega}\|^2\}. \tag{23}$$

The Hammersly-Clifford theorem is employed to obtain the Markov random field description of this model,

$$\pi_G(\theta_t = \omega_t | \lambda, \sigma^2, x, \{\theta_s = \omega_s, s \neq t\}) = Z_t^{-1}(\lambda, \sigma^2, x_t) \exp\{\lambda \sum_{C \in \mathcal{C}(t)} V_C(\boldsymbol{\omega}) - \frac{1}{2\sigma^2}(x_t - \omega_t)^2\}, \tag{24}$$

where

$$Z_t(\lambda, \sigma^2, x_t) = \sum_{\omega_t \in \Theta} \exp\{\lambda \sum_{C \in \mathcal{C}(t)} V_C(\boldsymbol{\omega}) - \frac{1}{2\sigma^2}(x_t - \omega_t)^2\}. \tag{25}$$

2.2 Conjugate Priors for λ and σ^2

In this section we introduce appropriate conjugate priors for λ and σ^2. Note that the conditional independence assumptions implicit in (1) imply that choices of a conjugate prior for λ depends only on the selected conditional distribution for $\boldsymbol{\theta}$, and not on the degradation model. On the other hand, the conjugate prior for σ^2 depends only on the degradation model used.

The conjugate prior for σ^2 based on Model 2.1 is familiar:

Model 2.4 ***Prior for*** σ^2

Assume the variance, σ^2, of the degradation model, Model 2.1 is distributed according to an inverted gamma density, denoted $\sigma^2 \sim InvGamma(\alpha, \beta)$ with density function

$$\pi_{1.2}(\sigma^2|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha(\sigma^2)^{\alpha+1}} \exp\{-\frac{1}{\sigma^2\beta}\}, \tag{26}$$

where $\alpha > 0$ and $\beta > 0$. The corresponding prior mean (if $\alpha > 1$) and variance (if $\alpha > 2$) of σ^2 are

$$\begin{aligned} E_{\pi_{1.2}}(\sigma^2) &= \frac{1}{\beta(\alpha-1)} \\ Var_{\pi_{1.2}}(\sigma^2) &= \frac{1}{\beta^2(\alpha-1)^2(\alpha-2)}. \end{aligned} \tag{27}$$

We proceed with a definition of an appropriate conjugate prior for λ based on $\pi_{1.1}(\boldsymbol{\theta}|\lambda)$ given in Model 2.3. When the suggested, second stage conjugate prior is combined with the first stage prior, the hyperparameters of the conjugate prior have a more concrete interpretation than λ in terms of certain properties of the expected true image, $\boldsymbol{\theta}$. This is critical since the direct specification of λ itself is difficult. This development is a special case of the general analyses of conjugate priors for exponential families given by Diaconis and Ylvisaker (1979).

Model 2.5 ***Prior for*** λ

For parameters $n_0 > 0$ and $u_0 \in (\min_{\boldsymbol{\omega}\in\Omega} U(\boldsymbol{\omega}), \max_{\boldsymbol{\omega}\in\Omega} U(\boldsymbol{\omega}))$, define

$$\pi_2(\lambda|n_0,u_0) = h^{-1}(n_0,t_0)\exp\{n_0u_0\lambda - n_0\mathcal{Q}(\lambda)\} \tag{28}$$

where

$$\mathcal{Q}(\lambda) = \log Z(\lambda) = \log \sum_{\boldsymbol{\omega} \in \Omega} \exp\{\lambda U(\boldsymbol{\omega})\} \tag{29}$$

and

$$h(n_0, u_0) = \int_{\Re} \left[\frac{e^{u_0 a}}{Z(\lambda)}\right]^{n_0} da. \tag{30}$$

First note that it is easy to check (or see Theorem 1 in Diaconis and Ylvisaker, 1979) that these densities are proper under the assumptions on n_0 and u_0. Next, by construction, the posterior density of $\lambda|\boldsymbol{\theta}$ is in the same family as the prior; simple calculation confirms that the updated parameters are $n_0' = n_0 + 1$ and $u_0' = \frac{n_0 u_0 + U(\boldsymbol{\theta})}{n_0+1}$.

To understand implications of the use of these priors, we investigate the interpretation of u_0 and n_0 in terms of expected features of the image. First, note that since Model 2.3 is a member of the exponential family, it follows that conditional on λ, $E_{\pi_{1.1}}(U(\boldsymbol{\theta})|\lambda) = \mathcal{Q}'(\lambda)$ and $Var_{\pi_{1.1}}(U(\boldsymbol{\theta})|\lambda) = \mathcal{Q}''(\lambda)$. Indeed, $U(\boldsymbol{\theta})$ is a sufficient statistic for λ.

Proposition 2.1 *Under the formulation in Model 2.5, the implied marginal mean of $U(\boldsymbol{\theta})$ is*

$$E[U(\boldsymbol{\theta})] = u_0. \tag{31}$$

Proof: Since $E_{\pi_{1.1}}(U(\boldsymbol{\theta})|\lambda) = \mathcal{Q}'(\lambda)$, iterated conditional expectation yields $E[U(\boldsymbol{\theta})] = E_{\pi_2}[\mathcal{Q}'(\lambda)]$. The result is then obtained readily by application of the usual integration by parts; see Theorem 2 in Diaconis and Ylvisaker (1979). Details are omitted for brevity.

Next, note that the usual interpretation of n_0 as the "equivalent sample size" of the prior is valid; see Diaconis and Ylvisaker (1979, p. 275). Finally, note that in Model 2.3, the quantity U has the following interpretation:

$$\begin{aligned} U(\boldsymbol{\theta}) &= [\#\text{of adjacent pairs that are the same level}] \\ &- [\#\text{of adjacent pairs that are different levels}] \end{aligned}$$

Hence, we have established rationales for choosing the parameters u_0 and n_0 to reflect prior information concerning the global degree of smoothness of the image.

2.3 Summary

Combination of all the model specifications described leads to the following joint posterior distribution,

$$\begin{aligned}\pi(\boldsymbol{\theta}=\boldsymbol{\omega},\lambda,\sigma^2|x) \;\propto\; & (\sigma^2)^{-\frac{rc}{2}-\alpha-1} \\ & \exp\{\lambda\,(n_0u_0+U(\boldsymbol{\omega}))-\frac{1}{\sigma^2}\left(\frac{1}{\beta}+\frac{1}{2}\|\boldsymbol{\theta}-x\|^2\right)- \\ & (n_0+1)\mathcal{Q}(\lambda)\}. \end{aligned} \tag{32}$$

First, note that the full conditionals for θ_t for all $t \in S$ are exactly as described in (24). Next, the full conditional for λ is

$$\pi(\lambda|\boldsymbol{\theta},\sigma^2,x) \propto \exp\{\lambda\,(n_0u_0+U(\boldsymbol{\theta}))-(n_0+1)\mathcal{Q}(\lambda)\}. \tag{33}$$

Finally, the full conditional distribution of σ^2 is an inverted gamma with parameters, $\alpha' = \alpha + rc/2$ and $\beta' = (1/\beta + 1/2\|\boldsymbol{\theta}-x\|)^{-1}$. Hence, the required inputs for (3), (4), and (5) are specified.

3 Estimating the partition function

We now consider the computational problem introduced due to the hierarchical structure of the model as alluded to in Section 1. Specifically, the posterior, (32), depends upon $\mathcal{Q}(\lambda)$. Evaluation of this function is typically infeasible. We suggest the following procedure for dealing with this problem. First, estimate $\mathcal{Q}(\lambda)$ via importance sampling on a grid of values to obtain $\hat{\mathcal{Q}}(\lambda)$. Then, substitute the estimate into the joint posteriors to obtain pseudo-posteriors from which Bayesian inference will be based. Inference from the pseudo-posteriors proceeds via Gibbs sampling. We chose to use the Markov mesh (see [1]) importance sampling function for discrete, finite lattice Gibbs distributions as described in [11].

Figure 1 displays the estimate of $\mathcal{Q}(\lambda)$ used in the examples described in the next section. To obtain this estimate we defined a grid of 3000 λ values between 0 and 3. At each grid point, standard importance sampling estimates of $\mathcal{Q}(\cdot)$ where obtained based in independent samples, each of size 50, from the importance sampler. The final estimated function is a smoothed estimate based on the 3000 grid points.

Approximation was restricted to $\lambda \in [0, 3]$. This choice was based on rough investigations of the posterior resulting from combination of the first and second stage priors; The "action" of the posterior was mostly contained in this region when $\boldsymbol{\theta}$ was assumed to be a true image such as those we will reconstruct in the next section.

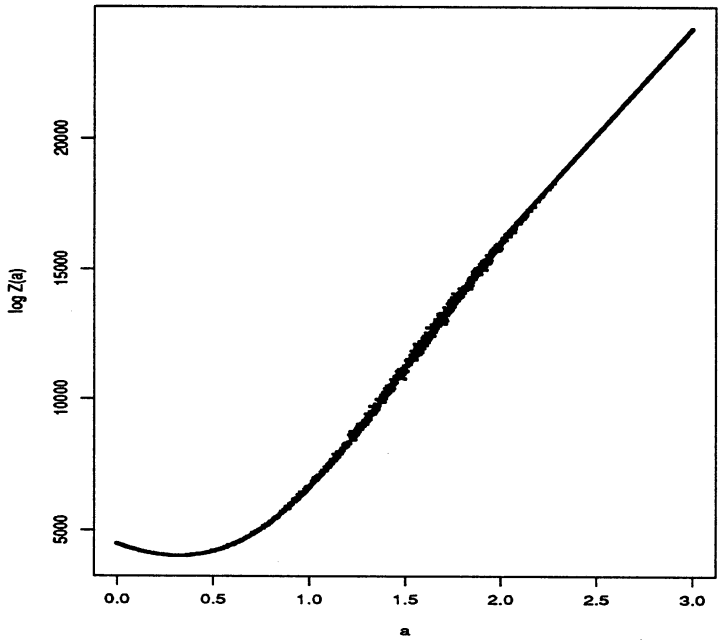

Figure 1: Markov mesh importance sampler estimate of $\mathcal{Q}(\lambda^1)$ from the first stage prior. The $n \times m$ lattice assumed is 64×64. L is assumed to be 3.

4 Example: Exchangeable Images

The example is motivated by problems arising in materials science, [8]. Consider the observation of multiple fractures of a composite material. A composite material consists of at least two distinct component materials not combined at the chemical level but making up one body of matter; concrete is an example. The component materials are commonly known as phases. From each fracture, we observe, with error, an intensity level, pixel image. The distinct phases of the material produce different mean intensity levels in the corresponding pixels of the observed image. The observed images are assumed to be degraded to an extent which makes classification of individual pixels into constituent phases difficult. Our goal is to produce a relevant model for improving the quality of this sequence of related images. The exchangeable model suggested earlier in this work will serve as the vehicle for enhancement.

Observed fracture surfaces, $X_1, \cdots, X_n$, are each assumed to be conditionally independent observations from Model 2.1, where each X_i has mean $\boldsymbol{\theta}_i$ and variance σ^2. A common error variance assumption is natural when all images are obtained via the same sensing device.

The composite material is assumed to consist of L distinct phases. Each phase produces a unique intensity level for pixels belonging to that phase from the set $\{0, 1, \cdots, L-1\}$. A priori, we believe that the fracture surfaces will be composed of several contiguous regions where each region contains only one phase. To encourage the formation of single phase, local regions, the true underlying images, $\boldsymbol{\theta}_i$, are modeled as i.i.d. observations from Model 2.3. The hyperparameter λ is unknown but common to all the images as the same degree of "smoothness" is expected in multiple fractures of the same composite.

As discussed in Section 2, the parameters, λ and σ^2, are modeled as independent variables from the appropriate conjugate priors, Models 2.5 and 2.4, respectively. In real applications the hyperparameter values would be specified based on prior knowledge of the images we expect and the sensing device which was employed. In our example we used the following specifications: $\alpha = 10$ and $\beta = 1.0$ so that $E(\sigma^2) = 0.11$ and $Var(\sigma^2) = 0.00012$ and $u_0 = 5890$ and $n_0 = 1.0$ so that $E(U(\boldsymbol{\theta}_i)) = 5890$ for all the underlying images.

Figure 2 displays three artifically constructed, underlying fracture surfaces, along with their observed, degraded counterparts. All the images are on a 64×64 pixel lattice; The degraded images were constructed by adding simulated, independent Gaussian noise with a standard deviation of 0.8 to the fracture surfaces. Note that this standard deviation is significantly larger than would be expected under the specified prior for σ^2. Furthermore, the degrees of smoothness in these images, as reflected through their values of U, are 7292, 7374, and 7104, respectively. The possible range on U for a 64×64 pixel image is $[-8064, 8064]$. This suggests that, though our prior mean for U, $u_0 = 5890$, is not "right on target," the actual values of U do not appear to contradict the exchangeability assumption.

Relevant full conditionals for the exchangeable model, necessary for implementation of Gibbs Sampling, are

1. $[\sigma^2 | \boldsymbol{\theta}, \lambda, x_1, \cdots, x_n]$, an inverted gamma density with $\alpha' = \sum_{i=1}^{n} |x_i|/2 + \alpha$ and $\beta' = [1/\beta + \sum_{i=1}^{n} \|x_i - \boldsymbol{\theta}_i\|^2/2]^{-1}$,

2. $[\lambda|\boldsymbol{\theta}, \sigma^2, x_1, \cdots, x_n]$ from the conjugate family of densities with u_0'$= (n_0 u_0 + \sum_{i=1}^{n} U(\boldsymbol{\theta}_i))/(n_0 + n)$ and $n_0' = n_0 + n$,

3. $[\theta_{i,s}|\lambda, \sigma^2, x_1, \cdots, x_n]$, as described in (24) for $i = 1, \cdots, n$ and $s \in S$.

To estimate selected posterior quantities, a single run of the Gibbs Sampler was run for 21000 iterations. The first 1000 iterations were discarded as a burn-in phase. Estimates of posterior expectations given below are based on ergodic averages of functions of the remaining 20000 iterations. Figure 3 displays posterior means, $\hat{E}(\boldsymbol{\theta}_i)$ and posterior variances $\widehat{Var}(\boldsymbol{\theta}_i)$ for $i = 1, 2, 3$. As ancillary information we also estimated the posterior means and variances of σ^2 and λ: $\hat{E}(\lambda) = 0.9427$, $\widehat{Var}(\lambda) = 0.000031$, $\hat{E}(\sigma^2) = 0.6643$, and $\widehat{Var}(\sigma^2) = 0.000112$.

We next describe a comparative analysis intended to evaluate whether the exchangeable model leads to any "sharing" of information across images (i.e., "borrowing strength"). Specifically, using the identical prior specifications used above, Image 1 was restored based on a simple hierarchical model. That is, Images 2 and 3 were not used in the analysis. A separate Gibbs' Sampler, again of length 20000, after discarding 1000 iterates, was run for this set-up. Estimates of $\hat{E}(\boldsymbol{\theta}_1)$ and $\widehat{Var}(\boldsymbol{\theta}_1)$ are displayed in Figure 4. Estimates for λ and σ^2 are $\hat{E}(\lambda) = 0.9202$, $\widehat{Var}(\lambda) = 0.0000536$, $\hat{E}(\sigma^2) = 1.022$, and $\widehat{Var}(\sigma^2) = 0.000568$. Comparing quantities of primary interest in this analysis we find the estimate of $\hat{E}(\boldsymbol{\theta}_1)$ under the exchangeable model superior to the reconstruction under the hierarchical model.

5 Comments

1. We have only briefly described the specific procedure for estimating the partition function primarily for brevity. However, though such estimation would typically be necessary for analysis of hierarchical models, this problem is of secondary importance relative to the main purpose of this article.

2. Turning to the results of Section 4, we hope the reader agrees that the results are encouraging. Though the examples presented are artifical and involve simulated "observation errors," the inference we take is that the quality of reconstructions obtained indicate positive potential for the basic approach in real applications.

3. As indicated in the Introduction, we have emphasized posterior mean estimation,

rather than MAP estimation. Our primary reason for suggesting this direction is the possibility for obtaining natural variance estimates. The behavior of the variance estimates presented in Figure 3 is quite interesting. As we might anticipate, posterior variances are large at what our "eye" would perceive as boundaries between phases within an image. In a sense, the posterior variance "images" act as simple "segmentation algorithms" or "edge-detectors." (See Nadler and Smith, 1993, Chapter 3.)

It is generally possible the MAP estimates may provide reconstructions which are more faithful to the data than corresponding posterior means. Indeed, we will not speculate on any general tendencies for which of these two estimates should be recommended. In fact we believe that which should be recommended will typically depend on the modeling strategy employed.

References

[1] K. Abend, T. J. Harley, and L. N. Kanal, *Classification of binary random patterns*, IEEE Trans. Inf. Theory **IT-11** (1965), 538–544.

[2] J. O. Berger, *Statistical decision theory and bayesian analysis*, second edition ed., Springer-Verlag, 1985.

[3] J. E. Besag, *On the statistical analysis of dirty pictures*, J. R. Statist. Soc. B **48** (1986), 259–302.

[4] J.E. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. R. Statist. Soc. B **36** (1974), 192–236.

[5] R. C. Dubes and A. K. Jain, *Random field models in image analysis*, J. App. Statist. **16** (1989), 131–163.

[6] A. E. Gelfand and A. F. M. Smith, *Sampling-based approaches to calculating marginal densities*, J. Amer. Statist. Assoc. **85** (1990), 398–409.

[7] S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, IEEE Trans. Patt. Anal. Mach. Intell. **PAMI-6** (1984), 721–741.

[8] A. F. Karr, *Statistics and materials science*, Tech. Report 4, National Institute of Statistical Sciences, 1994.

[9] R. Kinderman and J. L. Snell, *Markov random fields and their application*, American Mathematical Society, 1980.

[10] J. Marroquin, S. Mitter, and T. Poggio, *Probabilistic solution of ill-posed problems in computational vision*, J. Amer. Statist. Assoc. **82** (1987), 76–89.

[11] N. J. McMillan, *Computational methods for spatial statistics and image data*, Ph.D. thesis, Ohio State University, 1993.

[12] M. Nadler and E. P. Smith, *Pattern recognition engineering*, Wiley, New York, 1993.

[13] A. Possolo, *Spatial statistics and imaging*, Institute of Mathematical Statistics Lecture Notes-Monograph Series, 90.

[14] C. J. Preston, *Generalized gibbs states and markov random fields*, Adv. App. Prob. **5** (1973), 242–261.

[15] C. Ritter and M. A. Tanner, *Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler*, J. Amer. Statist. Assoc. **87** (1992), 861–868.

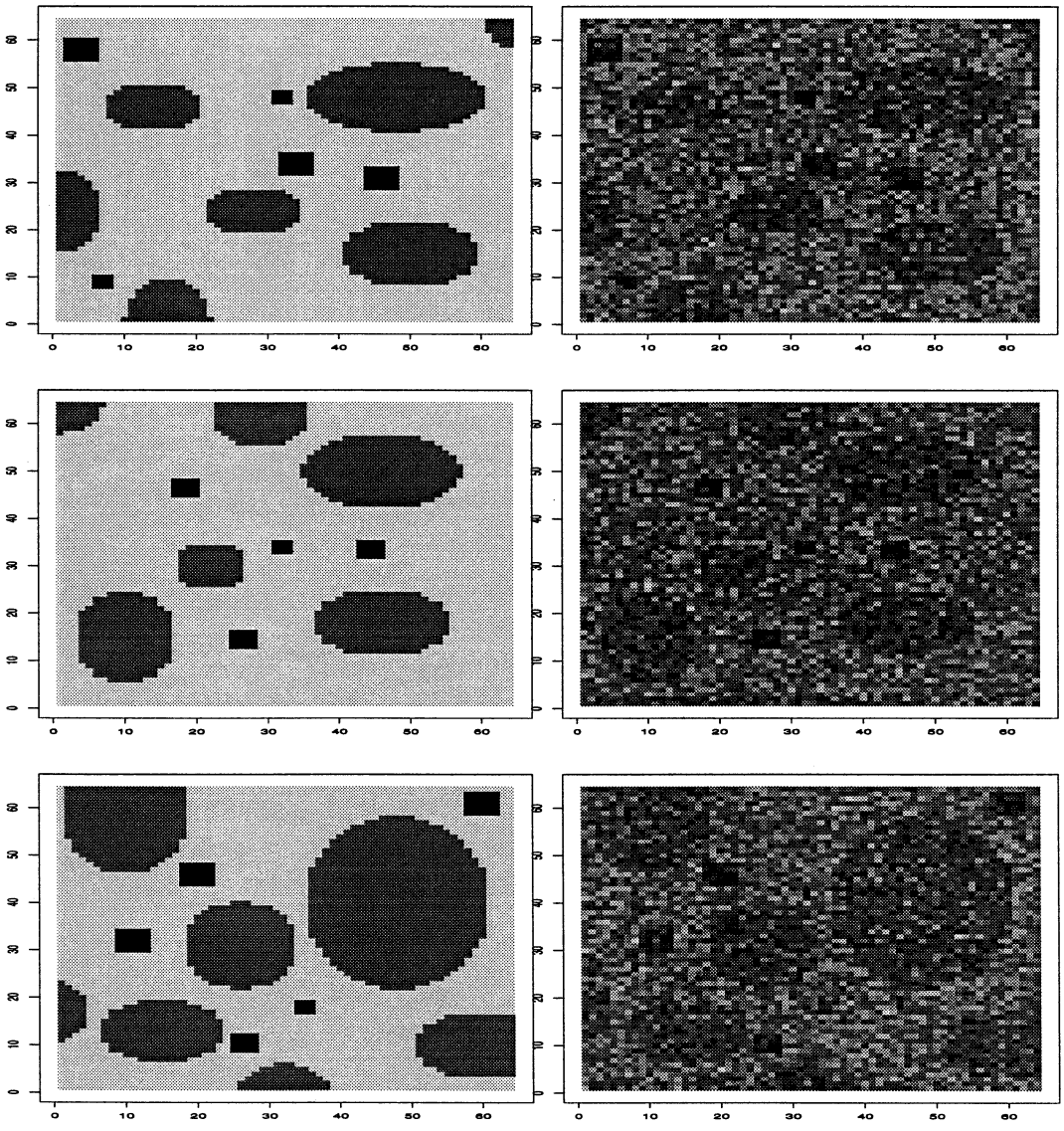

Figure 2: The left column of images are the unobserveable, underlying fracture surfaces. The right column of images are the corresponding observed images.

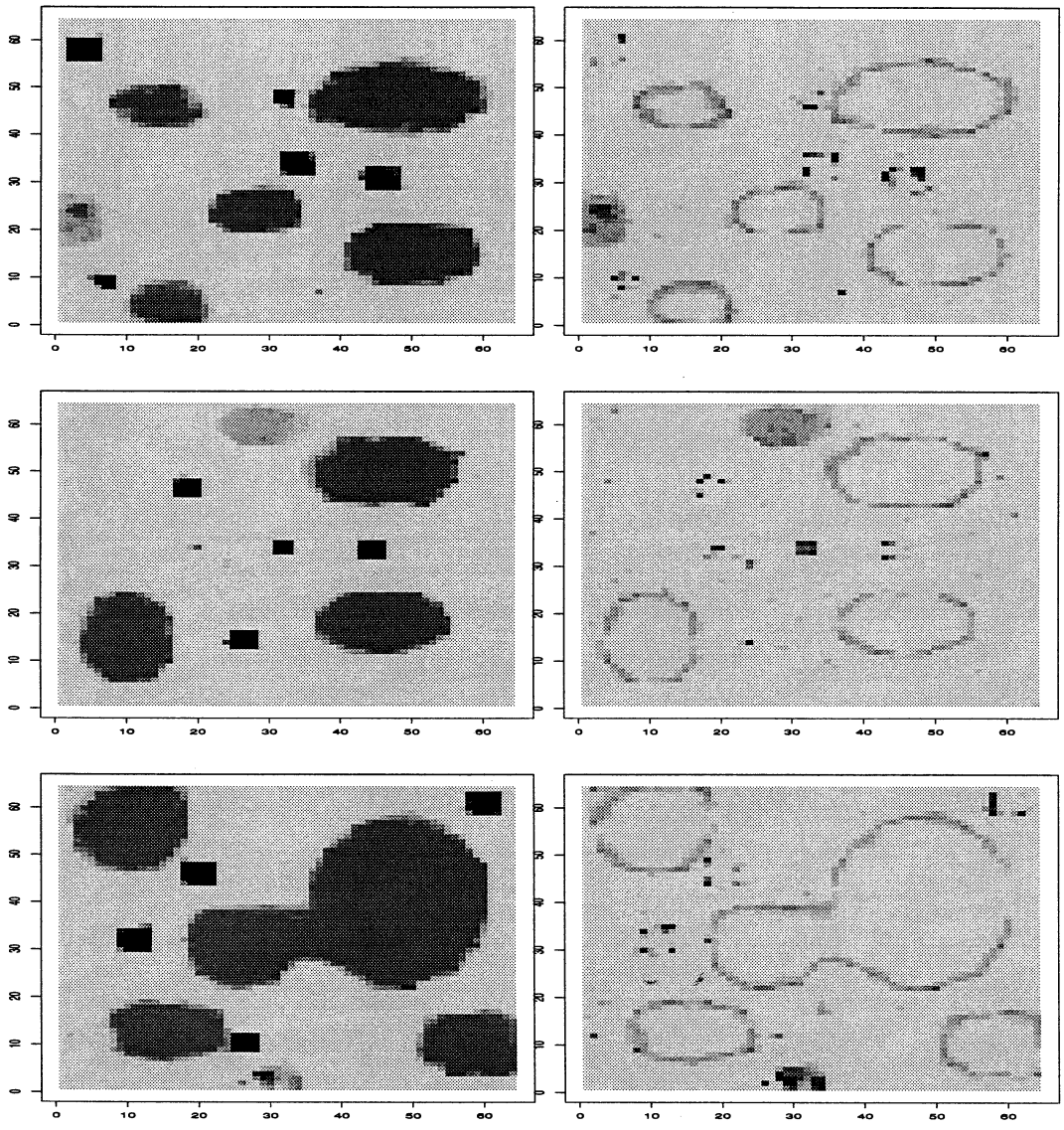

Figure 3: The left column contains reconstructions of the three observed images based on the exchangeable model. The right column contains corresponding variance estimates.

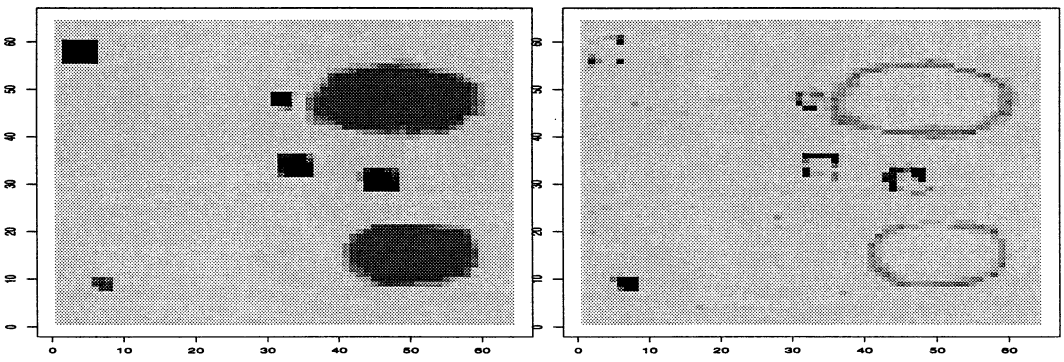

Figure 4: The image on the left is an estimate of the first image based on the hierarchical model. The right image is the corresponding variance estimate.