

Bayesian 2D Deconvolution: A Model for Diffuse Ultrasound Scattering*

ODDVAR HUSBY†, TORGRIM LIE‡, THOMAS LANGØ‡, JØRN HOKLAND§ and HÅVARD RUE†

Keywords: *Medical ultrasound, restoration, Markov random fields, diffuse scattering, Markov chain Monte Carlo*

Observed medical ultrasound images are degraded representations of the true acoustic tissue reflectance. The degradation is due to blur and speckle, and significantly reduces the diagnostic value of the images. In order to remove both blur and speckle we have developed a new statistical model for diffuse scattering in 2D ultrasound radio-frequency images, incorporating both spatial smoothness constraints and a physical model for diffuse scattering. The modeling approach is Bayesian in nature, and we use Markov chain Monte Carlo methods to obtain the restorations. The results from restorations of some real and simulated radio-frequency ultrasound images are presented, and compared with results produced by Wiener filtering.

1. Introduction

Ultrasound B-scan images are degraded representations of the true acoustical scatterers in the imaged tissue. The degradation introduces image artifacts (speckle) and limits the spatial resolution, hence reducing the diagnostic value of the images. The dominant degradation effects are blur and speckle due to what may be approximated as a convolution of the true reflectance field r with the imaging system point spread function h , with the addition of noise introduced in the image formation process. We assume the point spread function to be spatially invariant (Taxt, 1995; Iraca *et al.*, 1989), and thus the observed signal y can be modeled as

$$y_{i,j} = \{h * r\}_{i,j} + \eta_{i,j} = \sum_{k,l} h_{k,l} r_{i-k,j-l} + \eta_{i,j}, \quad (i,j) \in \mathcal{J} \quad (1)$$

where $\eta_{i,j}$ are independent Gaussian variables with mean zero and common variance τ^2 , and \mathcal{J} is a rectangular lattice with $n_r \times n_l = n$ sites. The indices i and j represent radial and lateral location, respectively, while k and l are local coordinates for the point spread function centered at (i,j) . Note that the assumption of spatially invariant point spread function is made out of computational convenience only, since our focus is on the modeling of the image formation process. We study the effect of this assumption in Langø *et al.* (1999), and refer to e.g. Taxt and Frolova (1999); Jensen and Leeman (1994), for other discussions of the effect of the point spread function.

Due to the limited bandwidth of the point spread function and the presence of

*© 2001 IEEE. Reprinted, with permission, from IEEE Transactions on Ultrasonics, ferroelectrics, and frequency control.

†Department of Mathematical Sciences, NTNU, N-7491 Trondheim, Norway. E-mail: Oddvar.Husby@math.ntnu.no

‡SINTEF Unimed, Ultrasound, 7465 Trondheim, Norway

§Department of Informatics, NTNU, 7491 Trondheim, Norway

additive noise, perfect deconvolution is impossible. As an alternative, statistical restoration of ultrasound images attempts to remove the effect of the degradation by estimating the true reflectance image r . A simple and commonly used technique is Wiener filtering (Jain, 1989; Tact, 1995); other methods include multiscale wavelet analysis (Zong *et al.*, 1998), but even with these methods the restored images suffer from blur and speckle.

We propose a new method for deconvolution of ultrasound images, based on Markov random field models (Winkler, 1995) and Bayesian statistical methods (Geman and Geman, 1984; Besag *et al.*, 1995). These methods have become increasingly popular over the last decade or two, and represent a general methodology for restoration of indirectly observed measurements. The most important feature is the ability to model the randomness and uncertainty inherent in the system. Specifically, in dealing with ultrasound the methodology allows us to use stochastic models of the imaged tissue and the image formation process, thus incorporating *a priori* knowledge and measurement error in a way not possible in deterministic restoration methods. Also, we are able to quantitatively assess the errors in our restorations, producing a *distribution* over the image space rather than a single filter-image restoration.

Our Markov random field model is based on the standard model for diffuse scattering (Goodman, 1975; Wagner *et al.*, 1987), where a point scatterer within one resolution cell is modeled to be Gaussian with mean zero and variance depending on the tissue type within the cell. Our idea is to let a spatially varying variance field represent the *underlying structures* of the imaged region; based on the assumption that this field is the *sole* descriptor needed to obtain faithful image restorations. The argument for the scattering model in Wagner *et al.* (1987) goes as follows: Assuming that the concentration of scatterers is large relative to the wavelength of the pulse, the reflectance $r_{i,j}$ at each location (i,j) is a Gaussian random variable with mean zero and variance $\sigma_{i,j}^2$. The reflectances at different points are assumed to be uncorrelated, and thus the reflectances have the conditional distributions (i.e. the distribution of $r_{i,j}$ given that the value of $\sigma_{i,j}^2$ is known)

$$p(r_{i,j}|\sigma_{i,j}^2) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left\{-\frac{r_{i,j}^2}{2\sigma_{i,j}^2}\right\}, \quad \forall (i,j) \in \mathcal{J} \quad (2)$$

Note that the radio frequency echo in general consists of a diffuse and a coherent scatter component (Cohen *et al.*, 1997), and that our model is valid only when the coherent component is negligible and when the number of diffuse scatterers is large enough for the law of large numbers to apply. This might not be the case for all tissue types; however, there exists statistical tests for identifying the regions of an image for which the diffuse scattering assumptions are true (Georgiou and Cohen, 1998). For the remainder of this paper we will assume these assumptions to hold. We see that the estimation of the variance field is essential; given the variance field, the reflectance can be simulated by drawing from the distribution equation (2).

The restoration can now be reformulated as a Bayesian estimation problem. We define an appropriate prior distribution $p(\sigma^2)$ for the variance field, trying to model the spatial properties inherent in the imaged regions. Using Bayes theorem, the prior distribution is then combined with the likelihood model $p(y|\sigma^2)$ to obtain the posterior distribution $p(\sigma^2|y, h)$ from which inferences can be made:

$$p(\sigma^2|y, h) \propto p(y|\sigma^2, h)p(\sigma^2) \quad (3)$$

We define the prior model in Section II, selecting a distribution from the family suggested in Geman and McClure (1987), where it was used for tomographic reconstruction. These models have the property that they are able to recover discontinuities in piecewise smooth images. This property is essential in our application, as discontinuities may represent interfaces between regions of different tissue types. Following the definition of the prior we compute the full posterior distribution. The distribution is analytically intractable; hence we have to rely on iterative stochastic techniques to make inferences. Markov chain Monte Carlo methods (Geman and Geman, 1984; Hastings, 1970; Green, 1995) are well suited for sampling Markov random field models, but we argue that when the model is formulated in terms of σ^2 only, these techniques can be very inefficient. However, in Section III we show that this problem can be solved by augmenting the model with the reflectance field r . We show that y and σ^2 are conditionally independent given r , and thus that r can be viewed as an auxiliary variable (Besag and Green, 1993; Higdon, 1998) decoupling y and σ^2 . When sampling from the joint distribution of r and σ^2 this fact can be utilized to construct an algorithm with better computational properties.

A related approach can be found in Hokland and Kelly (1996), who use a discrete Markov model for the variance field. However, their discrete model made the method unstable. Furthermore, it is unclear how well a discrete model describes the underlying structure of e.g. human tissue, and also how well it can account for the radial reduction in intensity seen in ultrasound images. Our main contribution is the formulation of the continuous variance field, which from a computational viewpoint is far better, making sampling and estimation easier, faster and more robust. Specifically, we avoid the problems of choosing the number of levels in the discrete variance field, and of estimating the variance at each level. We also feel that a continuous model is a better descriptor of the imaged regions, since it is reasonable to assume that scattering properties varies slightly within tissue regions. Finally, radial reduction in intensity can only be explained by a continuous model.

We conclude the paper with examples and a brief discussion in Sections IV and V.

2. Bayesian Model Formulation

2.1. Introduction

With a slight abuse of notation we will use $p(\cdot)$ to represent both the distribution and the density of the random variables. Recall that the images are defined on the rectangular lattice \mathcal{J} , using the polar coordinate representation of B-scan ultrasound images. All images are defined on the domain \mathcal{J} , letting e.g. $\sigma^2 = (\sigma_{1,1}^2, \dots, \sigma_{n_r, n_l}^2)^T$, where n_r and n_l represent the number of samples in the radial and lateral directions, respectively. For any subset $\mathcal{A} \subset \mathcal{J}$ we write $r_{\mathcal{A}} = \{r_{i,j}; (i,j) \in \mathcal{A}\}$ and $r_{-\mathcal{A}} = \{r_{i,j}; (i,j) \in \mathcal{J} \setminus \mathcal{A}\}$ and write $p(r_{\mathcal{A}})$ and $p(r_{-\mathcal{A}})$ for the corresponding densities. Our approach is based on Markov random field (MRF) models. One of the key features of Markov random fields is that each site (i,j) has associated with it a neighborhood $\partial(i,j)$, such that the conditional probability of the field at the site (i,j) , given the values at all other sites, only depend on the values of the field in (i,j) (Winkler, 1995). We write $(i,j) \sim (k,l)$ if the sites (i,j) and (k,l) are neighbors.

2.2. Prior distribution for the variance field

The crucial part of our model is the formulation of the prior distribution for the variance field σ^2 . We let the prior be of the form

$$p(\sigma^2) \propto \exp \left\{ -\beta \sum_{(i,j) \sim (k,l)} w_{ij,kl} \phi \left(\ln \frac{\sigma_{i,j}^2}{\sigma_{k,l}^2} \right) - \sum_{(i,j)} \psi(\sigma_{i,j}^2) \right\} \quad (4)$$

where the first sum is over all pairs of neighboring pixels in the lattice \mathcal{I} . $\phi(\cdot)$ is a spatially structured interaction term, and $\psi(\cdot)$ is an unstructured level term. The choice of the prior model is justified from physical considerations about the scattering medium, and is explained below.

In the diffuse scattering model the variance depends on the acoustical properties of the scattering medium (Goodman, 1975), and it is thus reasonable to assume the following:

- Variances tend to be approximately constant in regions of homogeneous tissue.
- Abrupt changes in variance may occur at interfaces between different tissue regions, as the scattering intensity varies between tissue types.

In other words, we believe the field to be piecewise planar with each subregion corresponding to the different tissue types in the imaged region. Hence prior distributions with convex interaction functions will be unsuited for our purpose, since they tend to over-smooth images, favoring gradual changes in intensity. We will instead use an interaction function of the form

$$\phi(u) = \frac{\left| \frac{u}{\delta} \right|^\gamma}{1 + \left| \frac{u}{\delta} \right|^\gamma} \quad (5)$$

as suggested in Geman and McClure (1987); this model is known to provide for recovery of discontinuities in blurred and noisy images. The non-interpolating property of the model is due to its concavity and finite limit as $u \rightarrow \infty$, favoring large intensity jumps instead of gradual changes; see Figure 1. We refer to Geman and Reynolds (1992) and Charbonnier *et al.* (1997) for a further discussion of the properties of this class of models.

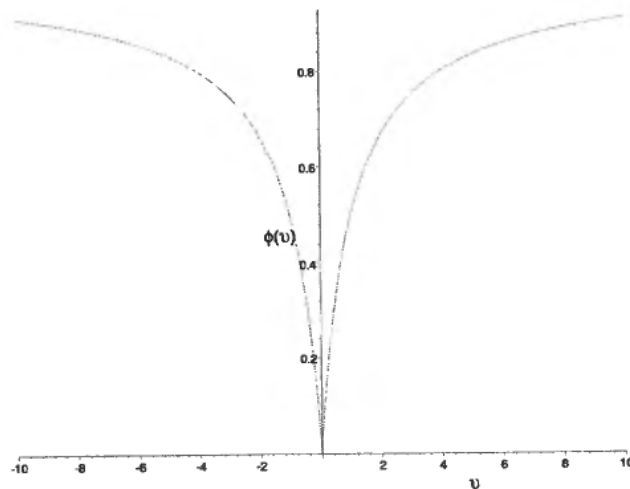


Figure 1. Plot of the interaction function $\phi(u)$ with $\delta = 1$ and $\gamma = 1$.

In addition to the spatially structured model equation (5), the unstructured term is chosen from the conjugate family of inverse Gamma distributions; and can be seen as a prior on the level of the variances. The parameter of the inverse Gamma distribution is chosen small to make the distribution sufficiently vague. Combining the two terms, we get the following prior density:

$$p(\sigma^2) \propto \exp \left\{ -\beta \sum_{i,j \sim k,l} w_{ij,kl} \phi \left(\ln \frac{\sigma_{i,j}^2}{\sigma_{k,l}^2} \right) - \sum_{i,j} \frac{\varepsilon}{\sigma_{i,j}^2 + (\varepsilon + 1) \ln \sigma_{i,j}^2} \right\} \quad (6)$$

$$= \exp \{ -\Phi(\sigma^2) - \Psi(\sigma^2) \}$$

where β and ε are positive parameters, and $w_{ij,kl}$ is a weight depending on the neighborhood system. In the experiments we use a second order neighborhood system with $w_{ij,kl} = 1$ for orthogonal nearest neighbors, and $w_{ij,kl} = 1/\sqrt{2}$ for diagonal nearest neighbors. The parameter β determines the strength of the interaction, while δ in equation (5) determines the size of the intensity jumps allowed. The parameter γ is set equal to 1 in all experiments, since for $\gamma \leq 1$ sharp boundaries are favored over gradual transitions (Geman and Reynolds, 1992), while larger values of γ permit more variations within homogeneous regions.

2.3. Posterior distribution

As noted in Section I, it is possible to formulate the model in terms of the variance field σ^2 only. Given an estimate $\hat{\sigma}^2$, the reflectance field r can be simulated readily using equation (2). However, in Appendix A we show that the posterior density $p(\sigma^2|y)$ has poor sampling properties, leading to computational problems. This deficiency is amended by considering the augmented model (σ^2, r) as follows.

2.3.1. Model augmentation The idea behind model augmentation is to introduce auxiliary variables that allow one to construct sampling algorithms that in some sense are easier to work with. The variable z of interest is augmented by a variable u , and to generate realizations from $p(z)$ one specifies the conditional distribution $p(u|z)$ and builds $p(z, u) = p(u|z)p(z)$, which is then sampled from. The idea is to specify u in such a way that the joint sampling of (z, u) is in some sense easier than the sampling of z alone. Note that $p(z, u)$ maintains the marginal distribution $p(z)$.

In our setting we let r be the auxiliary variable. The augmented model is then (σ^2, r) with distribution

$$p(\sigma^2, r|y, h) \propto p(y|r, h)p(r, \sigma^2) \propto p(y|r, h)p(r|\sigma^2)p(\sigma^2) \quad (7)$$

Note that σ^2 and y are conditionally independent given r , thus r decouples σ^2 and y , making the variance field σ^2 conditionally independent of the blur. This is important because the blur causes long-range dependency, making unconditional sampling of σ^2 very difficult.

2.3.2. Posterior density We now proceed to find the joint density for σ^2 and r . According to equation (1), the density for the observed image y can be written as

$$p(y|r, h) = \prod_{(i,j) \in \mathcal{J}} p(y_{i,j}|r, h) \propto \prod_{(i,j) \in \mathcal{J}} \exp \left\{ -\frac{1}{2\tau^2} (y_{i,j} - \{h * r\}_{i,j})^2 \right\} \quad (8)$$

where the convolution between h and r can be considered a constant shifting of the expectation of y , and is as defined in equation (3).

By inserting the expression for $p(\sigma^2)$, $p(r|\sigma^2)$, and $p(y|r, h)$ in equation (7), the joint density for σ^2 and r can be written as

$$\begin{aligned} p(\sigma^2, r|y, h) &\propto p(y|r, h)p(r|\sigma^2)p(\sigma^2) \\ &\propto \exp\left\{-\sum_{i,j}\left(\frac{1}{2\tau^2}(y_{i,j}-\{h*r\}_{i,j})^2 + \frac{r_{i,j}^2 + 2\varepsilon}{2\sigma_{i,j}^2} + \left(\varepsilon + \frac{3}{2}\right)\ln\sigma_{i,j}^2\right)\right. \\ &\quad \left.-\beta\sum_{i,j\sim k,l}w_{ij,kl}\phi\left(\ln\frac{\sigma_{i,j}^2}{\sigma_{k,l}^2}\right)\right\} \end{aligned} \quad (9)$$

We will later show that our sampling algorithm will use the conditional distribution of $r_{i,j}$ and $\sigma_{i,j}^2$ at single pixels (i,j) , and from equation (9) we arrive at

$$\begin{aligned} p(\sigma_{i,j}^2|\sigma_{\partial(i,j)}^2, r) &\propto \exp\left\{-\left(\varepsilon + \frac{r_{i,j}^2}{2}\right)\frac{1}{\sigma_{i,j}^2} - \left(\varepsilon + \frac{3}{2}\right)\ln\sigma_{i,j}^2\right. \\ &\quad \left.-\beta\sum_{(k,l)\in\partial(i,j)}w_{ij,kl}\phi\left(\ln\sigma_{i,j}^2/\sigma_{k,l}^2\right)\right\} \end{aligned} \quad (10)$$

$$\begin{aligned} p(r_{i,j}|r_{-(i,j)}, \sigma^2, y, h) \\ \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{m,n}\left(y_{m,n} - \sum_{(k,l)\neq(m-i,n-j)}h_{k,l}r_{m-k,n-l} - h_{m-i,n-j}r_{i,j}\right)^2 - \frac{r_{i,j}^2}{2\sigma_{i,j}^2}\right\} \end{aligned} \quad (11)$$

By performing the multiplication of the quadratic parenthesis and rearranging the terms, equation (11) can be written as:

$$\begin{aligned} p(r_{i,j}|r_{-(i,j)}, \sigma^2, y, h) &\propto \exp\left\{-r_{i,j}^2\left(\frac{1}{2\tau^2}\sum_{m,n}h_{m-i,n-j}^2 + \frac{1}{2\sigma_{i,j}^2}\right)\right. \\ &\quad \left.+ 2r_{i,j}\left(\frac{1}{2\tau^2}\sum_{m,n}(h_{m-i,n-j}Y_{m,n})\right)\right\} \\ &\propto \exp\{-r_{i,j}^2g_1(\sigma^2, h, \tau) + 2r_{i,j}g_2(r_{-(i,j)}, y, h, \tau)\} \end{aligned} \quad (12)$$

where

$$Y_{m,n} = y_{m,n} - \sum_{(k,l)\neq(m-i,n-j)}h_{k,l}r_{m-k,n-l} = y_{m,n} - \sum_{k,l}h_{k,l}r_{m-k,n-l} + h_{m-i,n-j}r_{i,j} \quad (13)$$

Concerning implementation, note that we can store the values of $Y_{m,n}$ in a two-dimensional table, reducing the the double summation in equation (11) to a single sum over the values of $Y_{m,n}$. For every pixel updated, we have to update as many values in the table as there are elements in the point spread function h . Thus we

reduce the effective neighborhood size of r from the support of h convolved with h to only two times the support of h .

From equation (12) one can recognize that the conditional distribution for $r_{i,j}$ given y , $r_{-(i,j)}$ and σ^2 to be a Gaussian distribution with mean

$$\mu_{i,j} = \frac{g_2}{g_1} = \frac{\frac{1}{\tau^2} \sum_{m,n} h_{m-i, n-j} Y_{m,n}}{\frac{1}{\tau^2} \sum_{m,n} h_{m-i, n-j}^2 + \frac{1}{\sigma_{i,j}^2}} \quad (14)$$

and variance

$$\kappa_{i,j}^2 = \frac{1}{2g_1} = \left[\frac{1}{\tau^2} \sum_{m,n} h_{m-i, n-j}^2 + \frac{1}{\sigma_{i,j}^2} \right]^{-1} \quad (15)$$

and thus the conditional sampling of $r_{i,j}$ is straightforward using one of the standard algorithms for sampling Gaussian random variables.

3. The Sampling Algorithm

3.1. Metropolis-Hastings algorithm

Given the above model we want to use the posterior means as estimators of the variance and reflectance fields. The densities equation (9) are analytically intractable, hence we have to rely on Markov chain Monte Carlo methods as follows: We construct an ergodic Markov chain $\sigma^2(0); r(0); \sigma^2(1); r(1), \dots$ converging in distribution to $p(\sigma^2, r|y, h)$. This is done as follows: For each $k=1, 2, \dots$ we generate $\sigma^2(k)$ from $r(k-1)$ using $p(\sigma^2(k)|r(k-1))$, and then we generate $r(k)$ from $\sigma^2(k)$ using $p(r(k)|\sigma^2(k), y, h)$. After a burn-in of t_0 iterations an estimate of the posterior means can be found as

$$\widehat{E(\sigma^2|r)} = \frac{1}{t} \sum_{i=t_0+1}^{t_0+t} \sigma^2(i) \quad \widehat{E(r|\sigma^2, y, h)} = \frac{1}{t} \sum_{i=t_0+1}^{t_0+t} r(i) \quad (16)$$

Several methods exist for constructing the Markov chain $\sigma^2(0); r(0); \sigma^2(1); r(1), \dots$; we refer to Gilks *et al.* (1996) for a overview of the theory and methodology related to Markov chain Monte Carlo methods. Among the most common methods is the Metropolis-Hastings algorithm (Hastings, 1970), which is the one we have used in the experiments. Assuming we want to sample from a distribution $p(z)$, the algorithm is as follows: If the current state of the Markov chain is z , a new state z' is drawn from a proposal distribution with density $q(z, z')$. The new state is then accepted with probability

$$\alpha(z, z') = \min \left\{ 1, \frac{p(z')q(z', z)}{p(z)q(z, z')} \right\} \quad (17)$$

or otherwise the old state is retained. This algorithm is usually based on a succession of random local changes, since global changes tend to have small probability of being accepted. Usually, updating is most conveniently done at one site at the time, requiring access to the full conditionals $p(z_{i,j}|z_{-(i,j)})$ and relying on the Markov property $p(z_{i,j}|z_{-(i,j)}) = p(z_{i,j}|z_{\partial(i,j)})$ to do efficient updating. The attractiveness of

this approach comes from the fact that $p(z_{i,j}|z_{\partial(i,j)})$ usually has a small support, and hence can be computed efficiently.

We will follow this approach, sampling from the conditional densities given in equations (10) and (11). This sampling scheme is efficient for the following reasons.

- $p(\sigma_{i,j}^2|\sigma_{-(i,j)}^2, r)$ is a distribution with small support: The neighborhood is of the same size as specified in the prior equation (6). Because of this local Markov property, sampling is eased. We again stress that *given* r , the variance field σ^2 is conditionally independent of the point spread function h .
- $p(r_{i,j}|r_{-(i,j)}, \sigma^2, y, h)$ is a Gaussian distribution, which is straightforward to sample from.

Details of the sampling are given in the next section.

3.2. Sampling algorithm

We now proceed to describe some details related to the sampling scheme given in section III-A; this part may be skipped at a first reading.

3.2.1. Sampling the variance field. The choice of proposal distribution $q(\cdot)$ poses a small problem. Since variance is a scale parameter, we chose to propose a new state by scaling the existing state with a random number $u \sim U[1/2, 2]$. However, when working with the model, we found that the restorations often contained outlying pixel values in otherwise smooth regions. While not discussed in the original paper by Geman and Reynolds (1992), this is a feature of the model, as shown by Hurn and Jennison (1996). The reason is as follows: Let $r_{i,j}^0$ be the true pixel value at site (i,j) , and assume that $\tilde{r}_{i,j}$ is the marginal minimizer of equation (8). Then Hurn and Jennison (1996) show that the difference $\tilde{r}_{i,j} - r_{i,j}^0$ is Gaussian distributed with variance $(1/\tau^2 \sum_{k,l} h_{k,l}^2)^{-1}$. This variance can get extremely large when the blurring is severe, or when there is much noise. Then because of the concavity of the roughness penalty $\phi(\cdot)$, outlying pixel values are allowed to occur. Geman and Reynolds (1992) does not encounter this problem because they use a modified Gibbs sampler supported on a interval around the current values of the state and its neighbors. The algorithm is formalized in Geman *et al.* (1993), restricting the image space to those images in which no pixel values differs from its neighbors by more than a threshold γ_0 . However, as noted in Hurn and Jennison (1996), this restriction can be removed by embedding the modified Gibbs sampler in a Hastings algorithm.

In the present context, with σ^2 taking values on the whole of \mathbb{R}^+ , Gibbs sampling is costly to implement. Since the unsatisfactory reconstructions represent local, coordinate-wise minima, a better approach would be to design the algorithm to effectively escape such local minima. We have chosen a new proposal in which, letting $\tilde{\sigma}_{i,j}^2$ be the median of the neighbors of the current state, a new state is proposed by scaling $\tilde{\sigma}_{i,j}^2$ with a random number $u \sim U[1/2, 2]$. Thus if the current state is an outlying pixel value, large jumps can be made. The choice of the median is important; choosing the mean could lead to undesired smoothing at region boundaries.

The solution suggested in Hurn and Jennison (1996), where the parameter β is increased to enforce a higher degree of smoothness, would also partly eliminate the problem, but at the cost of slower convergence.

3.2.2. Sampling the reflectance field. As customary when sampling from Gaussian

distributions, we choose a Gaussian distribution centered at the current state as a proposal distribution for $r_{i,j}$. Following Roberts *et al.* (1997), the standard deviation of the proposal distribution is scaled so that the acceptance ratio equation (17) is about 0.25. This means that we accept about one out of four proposed moves.

3.3 Diagnosing convergence

When designing algorithms, assessing the convergence of the Markov chain is an important issue. Quantitative convergence rates are in most cases hard to find, but there exists a number of convergence diagnostics (Brooks and Roberts, 1998; Roberts, 1995; Mengersen *et al.*, 1998) which can be used as a guide to the convergence of the chain. In this paper we have assessed the rate of convergence by monitoring the functional

$$f(\sigma^2) = \beta \sum_{i,j \sim k,l} w_{ij,kl} \phi \frac{\ln \sigma_{i,j}^2}{\sigma_{k,l}^2} \quad (18)$$

using as burn-in the approximate number of iterations needed until all appear to have converged. To further verify that the algorithm produces consistent results, we have started sampling with over-dispersed initial states. Figure 2 shows the result from such an experiment. We start two different runs in initial σ^2 -fields with constant levels of 10 and 0.1, respectively. Figures 2(a) and 2(b) show the trace of the runs at two different locations, while 2(c) shows traces for the functional f . Clearly, this is a

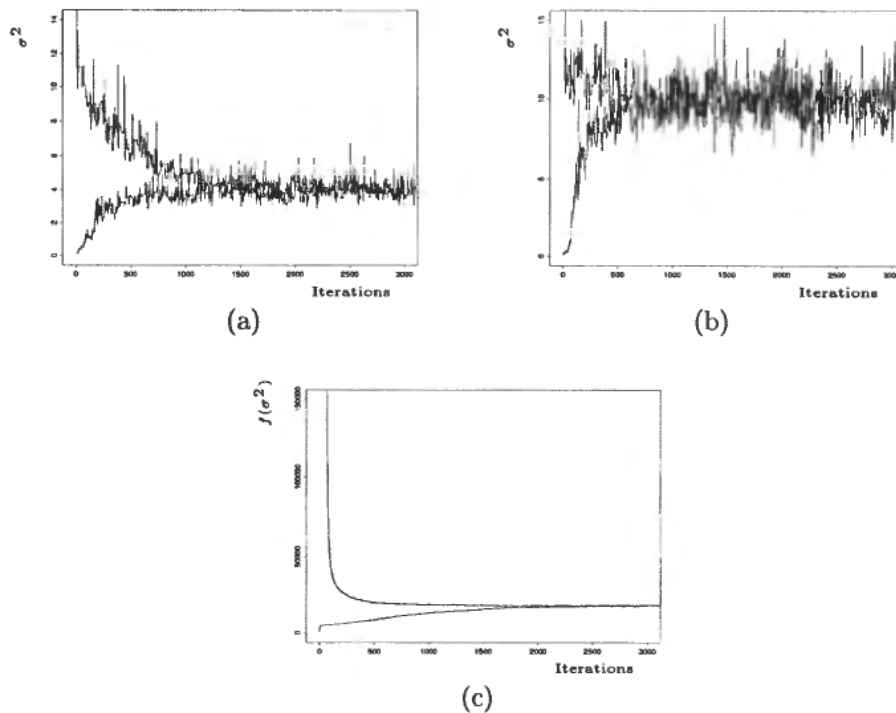


Figure 2. Traceplots of two Markov chains started in different initial states. (a) and (b) show plots of the variance at two different sites, while (c) shows a traceplot of the functional f .

subjective method, but the plots do at least indicate that the estimates produced by the algorithm are independent of the initial values used.

3.4. Parameter selection

The free parameters are β , δ , ε , γ in the prior, the noise variance τ^2 and the point spread function h . As explained in Section II-B we use $\gamma = 1$ in all the experiments. For a discussion of the effects of changing γ , see e.g. Geman and McClure (1987), Geman and Reynolds (1992). The parameter ε in the inverse Gamma distribution is also regarded as fixed; we use the value 0.00001 in the experiments shown, but changes of several orders in magnitude does not seem to have any significant effect on the results. In a considerably simpler setting than ours, Geman and Reynolds (1992) and Hurn and Jennison (1996) find conditions on β and δ so that the estimators satisfy certain desirable properties. In our case this approach does not seem feasible, and so β and δ is chosen by hand. As pointed out in Geman and Reynolds (1992) the value of δ should be chosen approximately the size of the discontinuities considered important, and should thus be set in each specific application. We use $\delta = 2$ in our experiments, but the results seems stable over a wide range of deltas, as was shown quantitatively in Hurn and Jennison (1996). The last prior parameter, β , is chosen by trial and error, and is set to $\beta = 4$. The results do to some extent depend on the value of this parameter, as does the speed of convergence. However, it is hard to do this kind of modeling without having to choose at least one parameter. In particular, one has to chose a parameter balancing the degree of fidelity to the data to the smoothing imposed by the prior distribution.

Following Hokland and Kelly (1996) we use a sinemodulated circular Gaussian point spread function (see Langø *et al.* (1999)). The radial and lateral standard deviations were estimated from the power spectrum of the observed images, while the noise variance is estimated from the power spectrum outside the acoustic bandwidth of the observed images.

4. Experiments

The model was tested on a simulated region image, and on a real ultrasound image; see Langø *et al.* (1999) for more experiments. We implemented the algorithm in LabVIEW, and the tests were run on a Power Macintosh 8600/25 computer, using approximately 50 minutes to process each image. A ten-fold increase in speed can be expected by implementing the routines in C. Furthermore, we claim by no means optimality of the algorithm, as our emphasis was on producing best possible restorations. The processing was done in polar coordinates, but the images were log-compressed and scan-converted into Cartesian coordinates prior to display.

4.1. Simulated images

To produce the simulated images we make a region image (Langø *et al.*, 1999) and use it as a variance field for the diffuse scatterers $r_{i,j}$, which are then sampled. The data y is made by convolving the ideal image with a parametric point spread function and adding white Gaussian noise.

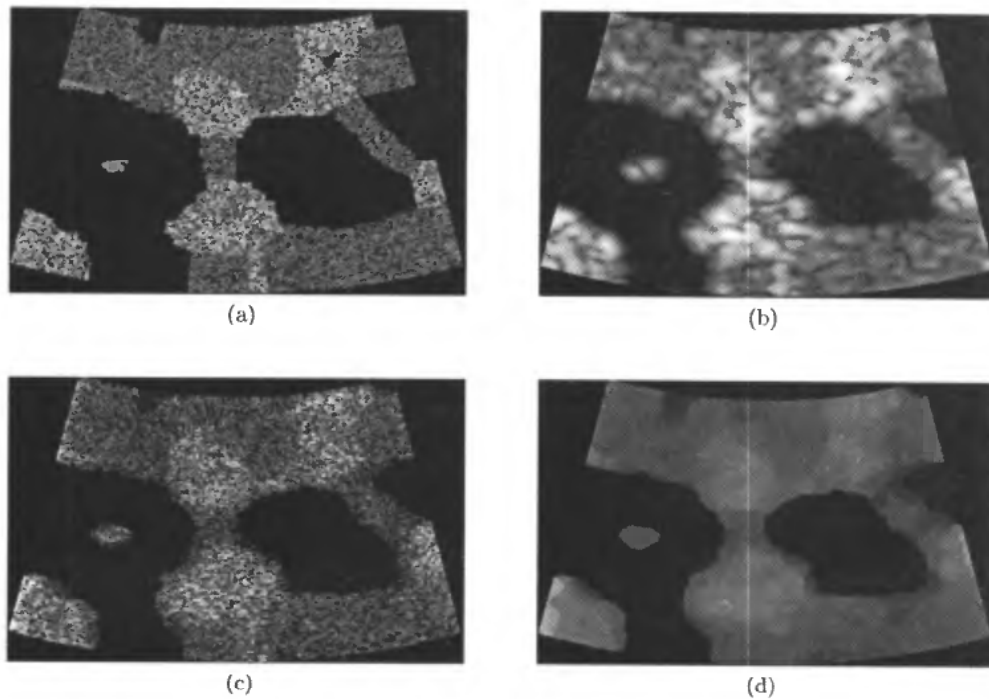


Figure 3. Three region simulation of ultrasound image. (a) Simulated 128×128 image of diffuse scatterers r . (b) The data y produced by blurring and adding noise. (c) Estimate of reflectance field r . (d) Estimate of the variance field σ^2 .

Figure 3a shows a 128×128 simulated image of diffuse scatterers with three region types having variances 0, 9, and 27. Figure 3b shows the simulated data produced by blurring and adding white Gaussian noise with variance 0.5. The restoration in Figure 3c is visually very close to the true image in 3a, although there is some smoothing due to the choice of estimator.

4.2. Real image

The ultrasound image in Figure 4 was acquired using a Vingmed Sound CFM 750 scanner with a special 16 MHz A/D-converter and a 5 MHz 15 mm annular array probe. 64 beams of 128 samples were recorded within one focal zone, and the image represent an area of approximately 1.5–0.7 cm. Figure 4a shows a homogeneous medium of diffuse scatterers submerged in water. In 4a, the lower area representing the diffuse scatterers shows much speckle, which can also be seen in Figure 4d. The image in Fig. 4d was estimated using the Wiener filter $H^*(u, v)/|H(u, v)|^2 + k$ in the frequency domain, with the regularization factor k tuned for optimal visual result. In our realization of r (Figure 4b) the speckle is removed, and the image shows two relatively homogeneous regions with a smooth interface. We used a sample from the posterior distribution as a realization, since the model states that $r_{i,j}|\sigma_{i,j}^2 \sim N(0, \sigma_{i,j}^2), \forall (i,j) \in \mathcal{I}$, and the posterior mean therefore is not meaningful as an estimator. Even though the lower region of the image should be homogeneous, the point spread function has introduced artifacts that our model cannot remove. Figure 4c shows the estimated variance field, an image which in fact summarizes all relevant

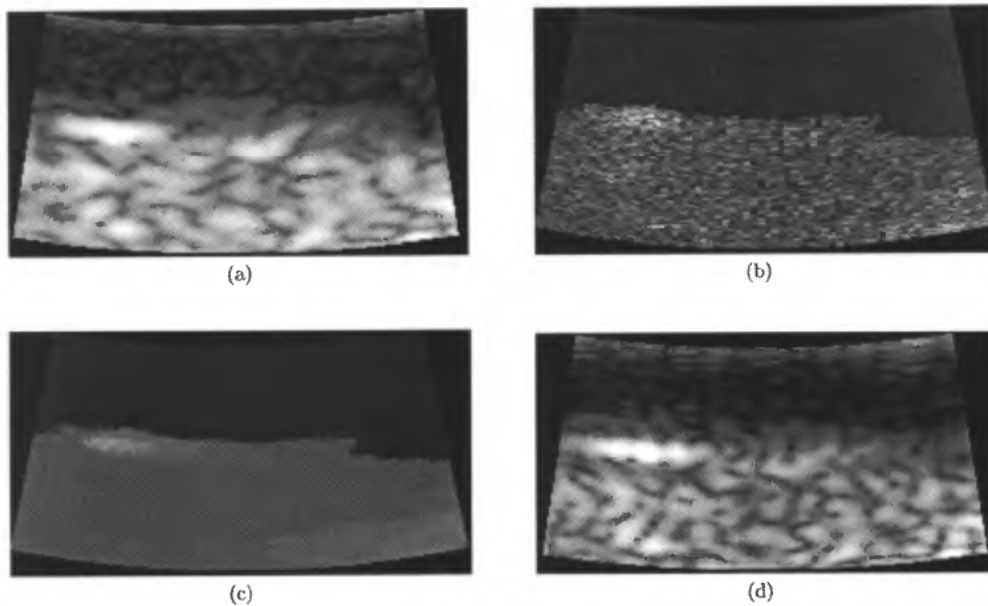


Figure 4. Ultrasound image of a homogeneous medium submerged in water. (a) Observed 64×128 image y (b) A sample from the posterior distribution for r given y and σ^2 . (c) Estimate of the variance field. (d) Wiener filtering of y .

information in a precise fashion. The homogeneity of the two regions, and the border between them are clearly seen, without disturbing noise and image artifacts.

5. Discussion

This paper describes a model for the scattering distribution of homogeneous tissue, and introduces smoothness constraints through the use of the model in Geman and Reynolds (1992) for the underlying scattering characteristics. The model is used in a new algorithm for restoration of ultrasound images.

We believe that the results demonstrate the potential usefulness of Markov random field models in restoration of medical ultrasound images. The restoration of the simulated images produced results that are more representative of the true object than the observed ultrasound images. Also, in the case of the real image, the results obtained are significantly better than the one produced by Wiener filtering. In both cases the speckle patterns are efficiently removed, while at the same keeping important details and not introducing artificial structures.

Our prior model favors homogeneous areas with sharp transitions between regions, which is a reasonable characteristic for a model representing ultrasound images since tissues are mostly homogeneous with sharp transitions between different tissue types. This characteristic of the prior distribution, which can be understood by studying the plot in Figure 1, is a great advantage to the restoration procedure. A disadvantage is that we may lose smaller structures. However, these structures are to a certain degree, depending on their size and the size of the degrading point spread function size, lost anyway in the blurring process.

The removal of speckle is efficient with our method, and we believe this results in visually better images. Speckle removal is important in 3D visualization systems,

where the data in a 3D volume is mapped to a 2D image through a process called rendering. In order to make the 2D image understandable, it is important that only significant structure information participates in the rendering process, and that irrelevant information (i.e. noise and speckle) is excluded. It is important to notice that if a certain speckle pattern is interpreted (by the model/algorithms) as a real structure when it in fact is a homogeneous area, this can lead to misinterpretation by the end observer. This is a crucial point if the clinician wishes to use this restoration method to study and analyze details from a volume scan. In other words, if this method proves useful for real ultrasound data, a surgeon, e.g. during removal of a tumor in the brain, could perform a 3D scan and then use the method to restore the reflection image of a smaller part of the volume to investigate the remains of the tumor. It would then be crucial that the method does not produce false structures due to the speckle pattern.

Further problems have to be solved. As mentioned in Section I our model has limitations, and it is important to consider how to deal with specular structure and quasiperiodic components. In Hokland and Kelly (1996) the specular structures were incorporated in the Markov field model, without significantly improving the results. Because of the continuous variance field, we believe that our model has the ability to at least partly account for the effects of specular structures. The presence of quasiperiodic scatterers will cause problems, but these occur only in some applications. Secondly, it is well known that point spread functions in ultrasound imaging do not have Gaussian shapes, and that they vary both radially and laterally (Jensen and Leeman, 1994; Ødegård, 1995). As we show in Langø *et al.* (1999), poor estimates of the point spread function effects the quality of the restorations.

Thus effort should be put into the estimation of the point spread function in order to get truly good restorations, especially in larger images. Some attempts on estimation of 2D point spread functions exists (Taxt, 1995; Jensen and Leeman (1994), but this is a field that needs further study. From a practical point of view, the efficiency of the algorithm should be improved. Most medical applications require real-time processing or at least on-line response times, and thus further optimization of the algorithm should be investigated. Note that when used on a sequence of images, the proceeding estimates could be used as initialization for the current frame, hopefully leading to a considerable speedup. Also, when doing real-time display of sequences, samples from the posterior distribution could be used instead of the sample means equation (16).

Acknowledgments

This work was partly supported by grants from the Research Council of Norway.

Appendix

1. Alternative model formulation

We derive the posterior distribution $p(\sigma^2|y, h)$, showing that our model can indeed be formulated in terms of σ^2 only. Using equation (3) we need the likelihood, which, using equation (1) and equation (2), is found by integrating out r :

$$p(y|\sigma^2) = \int p(y|r)p(r|\sigma^2)dr \quad (19)$$

For clarity of exposition we define matrices H and Λ as follows. Λ is an $n \times n$ diagonal matrix with entries $\sigma_{1,1}^2, \sigma_{1,2}^2, \dots, \sigma_{n_r, n_l}^2$ on the diagonal, and H is a $n \times n$ band matrix representing the point spread function h . Letting $\Sigma_s = \tau^2 I$ we get

$$\begin{aligned}
 p(y|\sigma^2) &= \int p(y|r)p(r|\sigma^2)dr \\
 &\propto \int \exp\left\{-\frac{1}{2}((y-Hr)^T \Sigma_s^{-1}(y-Hr) - r^T \Sigma^{-1})\right\}dr \\
 &= \exp\left\{-\frac{1}{2}y^T \Sigma_s^{-1}y\right\} \int \exp\{-r^T(H^T \Sigma_s^{-1}H + \Sigma^{-1})r \\
 &\quad + 2r^T H^T \Sigma_s^{-1}y\}dr \\
 &= \exp\left\{-\frac{1}{2}y^T \Sigma_s^{-1}y\right. \\
 &\quad \left.+ \frac{1}{2}y^T \Sigma_s^{-1}H(H^T \Sigma_s^{-1}H + \Sigma^{-1})^{-1}H^T \Sigma_s^{-1}y\right\} \\
 &= \exp\left\{-\frac{1}{2}y^T(\Sigma_s^{-1} - \Sigma_s^{-1}H(H^T \Sigma_s^{-1}H + \Sigma^{-1})^{-1}H^T \Sigma_s^{-1})y\right\} \\
 &= \exp\left\{-\frac{1}{2}y^T(\Sigma_s + H\Sigma H^T)^{-1}y\right\}
 \end{aligned}$$

where the last line comes from the identity

$$\begin{aligned}
 A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \\
 = (A + BCD)^{-1}
 \end{aligned}$$

see e.g. Mardia *et al.* (1979), p. 459.

The posterior density is then found by combining the prior density equation (6) with the density of the likelihood model equation (8):

$$\begin{aligned}
 p(\sigma^2|y) &\propto p(y|\sigma^2)p(\sigma^2) \\
 &\propto \frac{1}{|H\Sigma H^T + \tau^2 I|^{1/2}} \exp\left\{-\frac{1}{2}y^T(H\Sigma H^T + \tau^2 I)^{-1}y - \Phi(\sigma^2) - \Psi(\sigma^2)\right\}
 \end{aligned} \tag{20}$$

However, note that the matrix $Q = (H\Sigma H^T)^{-1}$ in general is full; this can be seen heuristically by expanding Q as a power series. Letting $q_{ij,kl}$ be the elements of Q , it is a well known fact (Besag, 1974; Besag and Kooperberg, 1995) that

$$E(Y_{i,j}|y_{-(i,j)}, \sigma^2) = - \sum_{k,l \neq i,j} \frac{q_{ij,kl}}{q_{ij,ij}} y_{k,l} \tag{21}$$

implying that the neighborhood of site (i, j) consists of all sites (k, l) such that $q_{ij,kl} \neq 0$. Hence since Q is full the local Markov property fails to hold since the neighborhood of each site is simply the whole image. This means that conditional on the data y , any two sites in the variance field σ^2 are directly dependent, which is undesirable for at least two reasons. First, updating is time-consuming because for each site that is updated, the full density has to be computed in equation (17). Secondly, because of the global dependency, the chain will move slowly in the sample space, resulting in slow convergence of the Markov chain. Thus this formulation of the model, while valid and instructive, leads to computational problems. However, this deficiency can be amended by considering the augmented field (σ^2, r) , as shown in section II-C.

References

- BESAG, J. (1974) Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B*, vol. 36, pp. 192–236, with discussion.
- BESAG, J. and GREEN, P. (1993) Spatial statistics and Bayesian computation, *Journal of the Royal Statistical Society, Series B*, vol. 16, pp. 395–407.
- BESAG, J., GREEN, P., HIGDON, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion), *Statistical Science*, vol. 10, no. 1, pp. 3–66.
- BESAG, J. and KOOPERBERG, C. (1995) On conditional and intrinsic autoregressions, *Biometrika*, vol. 82, no. 4, pp. 733–746, December.
- BROOKS, S. P. and ROBERTS, G. O. (1998) Assessing convergence of Markov chain Monte Carlo algorithms, *Statistics and Computing*, vol. 8, no. 4, pp. 319–335.
- CHARBONNIER, P., BLANC-FERAUD, L., AUBERT, G. and BARLAUD, M. (1997) Deterministic edge-preserving regularization in computed imaging, *IEEE Transaction on Image Processing*, vol. 6, no. 2, pp. 298–311, February.
- COHEN, F. S., GEORGIU, G. and HALPERN, E. J. (1997) WOLD decomposition of the backscatter echo in ultrasound images of soft tissue organs, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.*
- GEORGIU, G. and COHEN, F. S. (1998) Statistical characterization of diffuse scattering in ultrasound images, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.*, vol. 45, no. 3, pp. 54–64.
- GEMAN, S. and GEMAN, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741.
- GEMAN, S. and MCCLURE, D. E. (1987) Statistical methods for tomographic image reconstruction, in *Proc. 46th Sess. Int. Stat. Inst. Bulletin ISI*, vol. 52.
- GEMAN, D. and REYNOLDS, G. (1992) Constrained restoration and the recovery of discontinuities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 3.
- GEMAN, D., REYNOLDS, G. and YANG, C. (1993) Stochastic algorithms for restricted image spaces and experiments in deblurring, in *Markov Random Fields: Theory and Application*, Chellappa, R. and Jain, A., Eds. Academic Press, New York.
- GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J. (1996) Markov Chain Monte Carlo in Practice, Chapman & Hall.
- GOODMAN, J. W. (1975) Statistical properties of laser speckle patterns, in *Laser Speckle and Related Phenomena*, J. C. Dainty, Ed. Springer Verlag, Berlin.
- GREEN, P. J. (1995) Reversible jump MCMC computation and Bayesian model determination, *Biometrika*, vol. 82, no. 4, pp. 711–732.
- HASTINGS, W. K. (1970) Monte Carlo simulation methods using Markov chains and their applications, *Biometrika*, vol. 57, pp. 97–109.
- HIGDON, D. M. (1998) Auxiliary variable methods for Markov chain Monte Carlo with applications, *Journal of the American Statistical Association*, vol. 93, no. 442, pp. 585–595, June.
- HOKLAND, J. H. and KELLY, P. A. (1996) Markov models of specular and diffuse scattering in

- restoration of medical ultrasound images, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 43, no. 4, pp. 660–669, July.
- HURN, M. and JENNISON, C. (1996) An extension of Geman and Reynolds' approach to constrained restoration and the recovery of discontinuities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 657–662, June.
- IRACA, D., LANDINI, L. and VERRAZZANI, L. (1989) Power spectrum equalization for ultrasonic image restoration, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.*, vol. 36, no. 2, pp. 216–222, March.
- JAIN, A. K. (1989) *Fundamentals of Digital Image Processing*, Prentice-Hall.
- JENSEN, J. A. and LEEMAN, S. (1994) Nonparametric estimation of ultrasound pulses, *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 10, pp. 929–936.
- LANGØ, T., LIE, T., HUSBY, O. and HOKLAND, J. (2001) Bayesian 2D deconvolution: Effect of using spatially invariant ultrasound point spread functions, *IEEE Trans Ultrason Ferroelect Freq Control*, vol. 48, no. 1, pp. 131–141.
- MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979) *Multivariate Analysis*, Academic Press, 1979.
- MENGERSEN, K. L., ROBERT, C. P. and GUIHENNEUC-JOUYAU, C. (1998) MCMC convergence diagnostics: a 'reviewww', Tech. Rep., CREST, INSEE, Paris, May.
- ØDEGÅRD, L. A. (1995) *Phase aberration correction in medical ultrasound imaging*, Ph.D. thesis, The Norwegian Institute of Technology.
- ROBERT, C. P. (1995) Convergence control methods for MCMC algorithms, *Statistical Science*, vol. 10, no. 3, pp. 231–253.
- ROBERTS, G. O., GELMAN, A. and GILKS, W. R. (1997) Weak convergence and optimal scaling of random walk metropolis algorithms, *Ann. Appl. Prob.*, vol. 7, pp. 110–120.
- TAXT, T. (1995) Restoration of medical ultrasound images using 2-dimensional homomorphic deconvolution, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.*, vol. 42, no. 4, pp. 543–554, June.
- TAXT, T. and FROLOVA, G. (1999) Noise robust one-dimensional blind deconvolution of medical ultrasound images, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.*, vol. 46, no. 2, pp. 291–299, March.
- WAGNER, R. F., INSANA, M. F. and BROWN, D. G. (1987) Statistical properties of radio-frequency and envelope-detected signals with application to medical ultrasound, *J. Opt. Soc. Am. A*, vol. 4, no. 5, pp. 910–922.
- WINKLER, G. (1995) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin.
- ZONG, X., LAINE, A. F. and GEISER, E. A. (1998) Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing, *IEEE Transaction on Medical Imaging*, vol. 17, no. 4, pp. 532–540, August.