NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

High-level Models in Ultrasound Imaging

by

Oddvar Husby

PREPRINT
STATISTICS NO. 18/2001
ISSN: 0804-9173

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S18-2001.ps
Oddvar Husby has homepage: http://www.math.ntnu.no/~okh
E-mail: okh@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

High-Level Models in Ultrasound Imaging

Oddvar Husby
Department of Mathematical Sciences
Norwegian University of Science and Technology

Contents

Chapter 1

Introduction

Bayesian analysis of images has received considerable attention over the last 20 years, following ground-breaking work on pixel based Markov random field models (Geman and Geman, 1984; Besag, 1986) and object models (Grenander, Chow and Keenan, 1991; Grenander and Miller, 1994). In the advent of these new developments, and as the theory for Markov chain Monte Carlo (MCMC) sampling has developed (Besag, Green, Higdon and Mengersen, 1995; Green, 1995; Rue, 2001), image models have become increasingly more complex, focusing on interpretation tasks such as recognition, classification, and detection of abnormalities (Grenander and Miller, 1998; Rue and Hurn, 1999). As models and applications have grown in scope and complexity, there has been interest in fully utilising the Bayesian paradigm's ability to provide more information than just point estimates. For the end user quantitative measurements of uncertainty are often of vital interest, and can be provided for instance by means of an interval estimate. In relation to this many authors advocate the use of fully Bayesian approaches to integrate out the effect of hyper-parameters (Higdon, Bowsher, Johnson, Turkington, Gilland and Jaszczak, 1997; Weir, 1997; Hurn, 1998). There has also been some interest in developing better point estimators more tailored for image interpretation than the usual MAP or MPM estimators (Baddeley, 1992; Rue and Syversveen, 1998).

The development of realistic data models has not received the same amount of interest as the prior models. However, besides the important fact that a realistic likelihood makes the analysis more believable, modelling the image formation process allows us to extract more information from the data. Thus data modelling may increase the scope of applications, as well as reducing uncertainty in the estimates. Some recent work on data modelling can be found in emission tomography (Green, 1990; Weir, 1997), microscopy (Higdon and Yamamoto, 2000), and impedance imaging (Nicholls and Fox, 1998; Andersen, Brooks and Hansen, 2000). In this report we consider ultrasound imaging, and present a new model for contour detection in ultrasound images. Ultrasound is an widely used imaging modality, mainly because of low cost, portable equipment and real-time imaging capability. Increasingly, medical ultrasound images are used in complicated applications such as surgery (Langø, 2000) and evaluation of cardiac diseases (Mulet-Parada and Noble, 2000). Common for these and other applications is the need for detecting anatomical boundaries, eg. for determining the position of a surgical instrument, measuring response to some stimuli, or estimating quantities such as ejection fraction and wall motion. There exists a vast literature on contour detection in medical images, see Duncan and Ayache (2000) and Pham, Cu and Prince (2000) for recent reviews; including a few Bayesian approaches using line processes (Kao, Pan, Hiller and Chen, 1998) and deformable templates (de Figueiredo and Leitao, 1992; Hansen, Møller and Tøgersen, 2000). Compared to most other image modalities, contour detection in ultrasound images is particularly difficult because of the low resolution caused by noise, blur, edge reflections, and image artifacts called speckle; still most of the current approaches use image data as they appear on the ultrasound scanner after some pre-processing, using a simple additive Gaussian noise model as a likelihood. Such a model does not take into account *how* the speckle patterns are formed, and thus do not use all information provided by the data. However, a physical model for ultrasound imaging does exists, based on the raw data – or radio frequency signal – received by the scanner (Goodman, 1975; Wagner, Insana and Brown, 1987). Figure 1.1 (f) shows an example of a radio frequency ultrasound image. Recently the model for radio frequency images has been proposed for restoration of ultrasound images (Hokland and Kelly, 1996; Husby, Lie, Langø, Hokland and Rue, 2001; Langø, Lie, Husby and Hokland, 2001), and the main contribution of this report is to use this physically sound likelihood together with a deformable template model for contour detection in ultrasound images. The derived hierarchical model is complex, hence great care needs to be taken in designing a MCMC sampling algorithm for doing inference. In related applications such as agricultural field experiments (Besag and Higdon, 1999) and spatial epidemiology (Knorr-Held and Rue, 2002) block updating has proved to be very efficient, and we explore different block sampling algorithms for our model. It turns out that the choice of sampler is important for the reliability of the MCMC simulation.

To be more specific, Figure 1.1 shows a collection of real ultrasound images. Figure 1.1 (a) displays a part of the right ventricle of the human heart, while Figure 1.1 (b) and (c) show images of an abdominal aorta aneurism, i.e. a blood-filled dilatation of the aorta. The roundish object in the middle of the aorta is a vessel prosthesis. Finally, panels (d)-(f) displays cross-sectional images of the carotid artery, taken from the neck of a single individual. Common to the images is the characteristic speckle pattern that makes it difficult for the untrained eye to spot the important anatomical features, and this calls for image restoration techniques to improve contrast and resolution. We have previously discussed restoration using our new model for radio frequency ultrasound images and relying on standard single site MCMC algorithms for sampling and inference (Husby et al., 2001; Langø et al., 2001). To preserve discontinuities in the images, our model uses non-convex, edge-preserving prior potentials (Geman and Reynolds, 1992; Charbonnier, Blanc-Feraud, Aubert and Barlaud, 1997) which are known to have poor sampling properties. We did indeed experience slow convergence and inadequate mixing of our MCMC sampler, and the first aim of this report is to explore whether block sampling can improve the sampling properties of edge-preserving models. We start in Chapter 3 by deriving our model for radio frequency ultrasound images, and continue in Chapter 4 by reviewing edge-preserving prior models and discussing some of their important properties, in particular a dual model formulation introduced by Geman and Yang (1995). The dual model depends both on the parameters of interest and a set of auxiliary parameters, and is Gaussian conditional on the auxiliary parameters. We show how to construct such a dual model, and discuss how we can do block updating of the unknown parameters using Rue's algorithm for sampling Gaussian Markov random fields (Rue, 2001). In Chapter 7 we show examples comparing the block sampling algorithm with standard single site sampling.

The second theme of this report is interpretation of ultrasound images. To focus the discussion we present two potential applications involving contour detection. Figure 1.1 (b) and (c) shows an abdominal aorta aneurism, which is a potentially lethal condition, treated by inserting a vessel prosthesis. This is a delicate operation, and in the subsequent year the patient must undergo frequent follow-ups to check if the condition is stable. One particular concern is that the aneurism deflates and obstructs the flow of blood through the vessel prosthesis, and one therefore produces a sequence of cross-sectional images along the whole aneurism, and use the images to get an impression of the shape of the aneurism. The visualisation can be eased by using deformable templates to estimate the contour in each image, and then use eg. a smoothing spline for constructing a 3D model of the aneurism. A second example concerns the vessel images in Figure 1.1 (d)-(f). The carotid artery is susceptible to plaque deposits, a disease called atherosclerosis. This condition is usually diagnosed by angiography, but the method does not work for all patients. An alternative is to use ultrasound, since the vessel walls appear more clearly in that particular modality. Furthermore it is known that healthy arteries will dilate in response to infusion of acetylcholine, and thus comparison of the cross-sectional vessel areas before and after infusion may help in diagnosing atherosclerosis. For this method to be of use, uncertainty estimates of the areas must be provided, and we therefore aim at obtaining interval estimates of the cross-sectional areas.

These applications call for the use of deformable templates in analysis of ultrasound images, and for the need for obtaining both point estimates and interval estimates of contours. Another motivation for using template models is the strong reflections occurring at tissue boundaries in ultrasound images, see eg. Figure 1.1 (b) and (d). These reflections can be much stronger than the other modes of reflection, but are difficult to model without using explicit models for the tissue boundaries. We will demonstrate how we can use templates to model reflections at boundaries, and show experiments comparing models with and without boundary reflection. We discuss deformable template models in Chapter 5, and inference in object models in Chapter 6. We are particularly concerned with obtaining good point estimates of closed curves, and thus discuss distances between such curves. A hierarchical Bayesian model for contour detection is presented in Chapter 8, together with experiments using both a simulated ultrasound image and real images of blood vessels. Finally, we conclude with a discussion in Chapter 9.

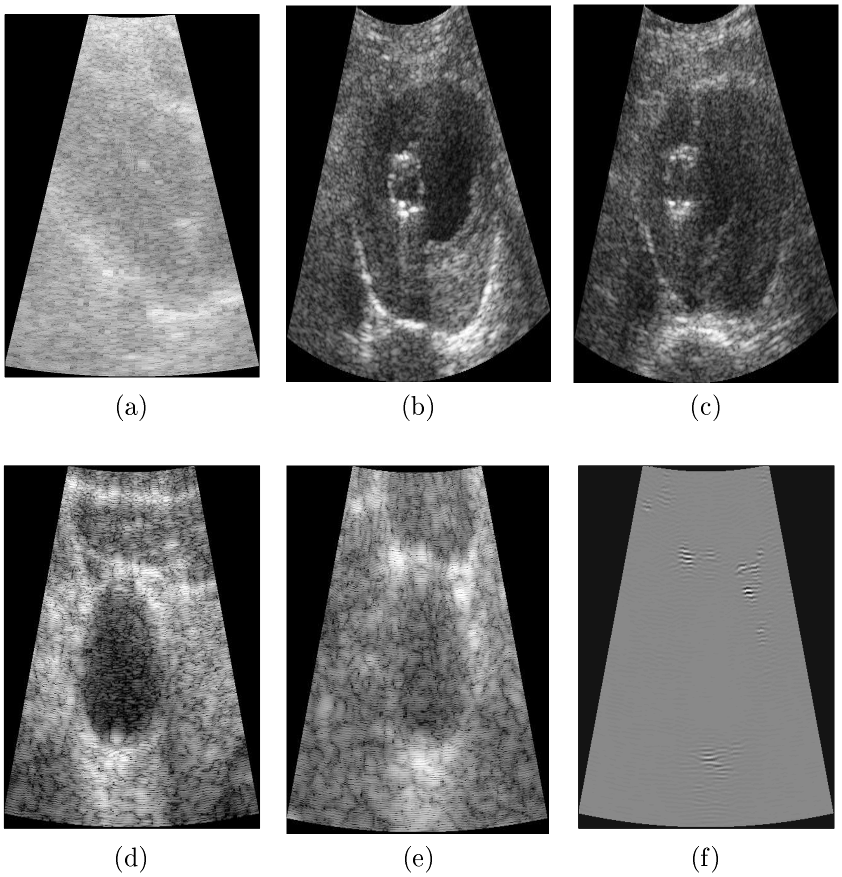


Figure 1.1: Real ultrasound images. Panel (a) shows a log-compressed radio frequency image of the right ventricle of the heart, while panels (b)–(c) show B-scan ultrasound images of an aorta aneurism with a vessel prosthesis in the middle. Panels (d)–(f) display radio frequency ultrasound images of cross sections through the carotid artery. The images in (d)–(e) are log-compressed to enhance contrast.

Chapter 2

Preliminaries

2.1 Bayesian Image Analysis

In image analysis the basic problem is to recover an unknown, "ideal" image $x = \{x_s\,;\, s \in S\}$ from an observation $y = \{y_s\,;\, s \in T\}$ related to x through a forward imaging transformation accounting for noise, scattering, digitisation, etc. We will in the following assume $T = S \subset \mathbb{Z}^2$. Usually the problem is ill-posed, as the information provided by the data is not sufficient to establish x. One way of dealing with this is to introduce a priori knowledge or beliefs on x, and to search for an estimate $\hat{x}$ that (i) satisfies the prior assumptions on x, and (ii) is in "agreement" with the data y. These constraints are determined by the prior distribution $\pi(x)$ on the space of ideal images, and by the likelihood distribution with density on the space of observations; these are combined to yield the posterior distribution

$$\pi(x \mid y) \propto \pi(y \mid x)\pi(x), \tag{2.1}$$

from which we seek an estimator $\hat{x}$ based on some decision-theoretic rule (see Section 6.1). The likelihood distribution is problem dependent, and is based on physical knowledge of the imaging system. A likelihood model for ultrasound images is one of the main topics of this report, and is discussed in Chapter 3. Prior models are discussed in Chapter 4 and Chapter 5. Both distributions are usually and most conveniently expressed in terms of Markov random field models, which we describe briefly. See eg. Besag (1974) and Winkler (1995) for details and further references.

2.2 Markov Random Fields

Consider a finite array $S \subset \mathbb{Z}^2$ with n sites. Let $x = \{x_s\,;\, s \in S\}$, with each x_s defined on a separable space $\mathbb{E}$ with σ-algebra $\mathcal{E}$. A strictly positive probability distribution π on $\mathbb{E}^n$ is called a random field. To obtain Markov random fields we associate a neighbourhood $\partial s \subset S$ to each node $s \in S$. The collection

$\partial = \{\partial s : s \in S\}$ is called a neighbourhood system if it satisfies the conditions (i) $s \notin \partial s$, and (ii) $t \in \partial s \Leftrightarrow s \in \partial t$. We write $s \sim t$ if sites s and t are neighbours.

DEFINITION 1
A random field π is a Markov random field w.r.t. the neighbourhood system ∂ if for all events $A \in \mathcal{E}^n$ and subsets $C \in S$,

$$\pi(x_C \in A \,|\, x_{S\setminus C}) = \pi(x_C \in A \,|\, x_{\partial C}).$$

A nonempty subset $C \subset S$ is called a clique if C contains a single point or if all elements are pairwise neighbours. We let $\mathcal{C}$ be the set of all cliques. The concept of cliques gives us an elegant and efficient way to represent the joint distribution $\pi(x)$ for a Markov field:

THEOREM 1 (HAMMERSLEY-CLIFFORD,(BESAG, 1974))
A random field π is a Markov field w.r.t. the neighbourhood system ∂ if and only if we can write

$$\pi(x) \propto \exp\left(-\sum_{C\in\mathcal{C}} V_C(x)\right), \tag{2.2}$$

where the $V_C(\cdot)$ are arbitrary functions satisfying $\int \exp(-\sum V_C(x))dx < \infty$.

The potential functions $\{V_C\,;\, C \in \mathcal{C}\}$ determines the properties of the Markov field, and the choice of such functions will be discussed in Chapters 4 and 5.

2.2.1 Gaussian Markov random fields

An important Markov random field model is the Gaussian Markov random field, which we will use extensively throughout. We therefore give a brief introduction, and refer to eg. Guyon (1995), Besag and Kooperberg (1995), and Rue (2001) for further details. A Markov random field x is Gaussian if and only if its finite dimensional distributions x_T, $T \subset S$ are Gaussian. Defining the potential function

$$V_C(x) = \begin{cases} \frac{1}{2}q_{ss}x_s^2 - q_{ss}\left(\mu_s + \sum_{r\in\partial s}\frac{q_{sr}}{q_{ss}}\mu_r\right)x_s, & C=\{s\} \\ q_{sr}x_sx_r, & C=\{r,s\}, r\neq s, \end{cases}$$

one can show that the conditional distribution for each element x_s is Gaussian with characteristics

$$\mathrm{E}\left(x_s \,|\, x_{-s}\right) = \mu_s - \sum_{r\in\partial s}\frac{q_{sr}}{q_{ss}}x_r \tag{2.3}$$

$$\mathrm{Var}\left(x_s \,|\, x_{-s}\right) = 1/q_{ss}. \tag{2.4}$$

We often write the distribution on vector form as

$$\pi(x) \propto \exp\left(-\frac{1}{2}x^T Qx + b^T x\right), \tag{2.5}$$

where the inverse covariance, or precision, matrix Q has entries $Q_{sr} = q_{sr}$ and b determines the mean μ through the relation $\mu = Q^{-1}b$.

Rue (2001) discusses sampling of Gaussian Markov random fields using Cholesky decomposition, ie. setting $Q = LL^T$ and solving $L^T x = z$. Noting that Q usually is sparse due to the local Markov structure, and that the bandwidth usually can be made smaller by permuting the columns of Q, he shows that for a $n_1 \times n_2$ $(n_1 \leq n_2)$ lattice with a $(2n_h + 1) \times (2n_h + 1)$ neighbourhood, the Cholesky decomposition costs $n_2 n_1^3 n_h^2$ floating point operations (flops), while solving $L^T x = z$ costs $2n_2 n_1^2 n_h$ flops. Note that the cost is linear in the largest dimension, so that sampling rectangular arrays is relatively more efficient than sampling square arrays. This is convenient in our ultrasound applications, as ultrasound data usually are measured on rectangular lattices.

2.2.2 Approximating Gaussian fields by Gaussian Markov random fields

One important aspect of Gaussian Markov random field models is the specification of the precision matrix Q. One wants the GMRF to fit with prior knowledge about the spatial behaviour of the unknown parameter. For instance, if the correlation structure of the unknown parameter can be represented by a Gaussian field π with a given correlation function ρ, how can we specify a GMRF $\widetilde{\pi}$ to approximate the Gaussian field? Rue and Tjelmeland (2001) solves this problem by minimising the metric

$$D(\pi, \widetilde{\pi}) = \sum_{r,s \in S \setminus \{0,0\}} (\rho_{rs} - \widetilde{\rho}_{rs}) \frac{1}{2\pi d(r,s)}, \tag{2.6}$$

where $d(r, s)$ is the Euclidean distance between (r, s) and $(0, 0)$ on the torus, and $\widetilde{\rho}$ is the correlation function of the GMRF. They show that Gaussian Markov random fields with 5×5-neighbourhoods can provide very good approximations to Gaussian fields with exponential, Gaussian and spherical correlation functions; for instance, Table 2.1 shows the coefficients of a GMRF approximating a Gaussian field having an exponential correlation function with range 20. Note that the coefficients are not very intuitive; this indicates that one should be careful when specifying the precision matrix of a GMRF model.

Table 2.1: Coefficients of the GMRF fitted to a Gaussian field having an exponential correlation function with range 20. The table shows the parameters of the lower right of the 5×5 neighbourhood.

.330	.285	$-.150$
.285	.015	.080
$-.150$	.080	$-.060$

CHAPTER 3

Ultrasound Images

In this section we describe ultrasound images: Their use in medicine and the physics behind the image formation. We derive our model for radio frequency ultrasound images, and show how we can model the strong reflections at tissue boundaries.

3.1 INTRODUCTION

Ultrasound is widely used in medical imaging, mainly because of its low cost, ease of use and its real-time imaging capability. The principle of ultrasound imaging is simple: A pulse of ultra high frequency sound is sent into the body, and the backscattered signal is measured after a time delay corresponding to the desired depth. The sound waves travel at different speeds in different types of tissue, and when the pulse hits a boundary between tissues having different acoustic impedances, it is partially reflected and partially transmitted. In addition there is reflection within homogeneous material due to randomly distributed (diffuse) or regularly spaced (quasi-periodic) scatterers. Thus variations in acoustic impedance is the basis for identifying regions of interest in the imaged tissue.

However, the quality of ultrasound images is often low, due to noise and blur introduced in the imaging process, and speckle, a characteristic granular structure resulting from the interference of the echoes from the randomly distributed scatterers in the tissue. Speckle may cause artificial structures in the image, thus reducing the diagnostic value.

Reducing speckle to enhance ultrasound image quality has been an area of active research over the last decade. Most work has been done on filtering (Iwai and Asakura, 1996; Zong, Laine and Geiser, 1998), and deconvolution (Taxt, 1995; Jensen and Leeman, 1994) techniques. These approaches have in common that they do not utilise all knowledge of the physics behind ultrasonic image formation. In addition, most work on the B-mode or intensity images that are obtained by transforming the backscattered radio frequency signal.

The intensity images can be stored with less sampling and memory than the original signal (Wagner et al., 1987), but contain less information and are more difficult to model.

There has been little work using Bayesian methods and Markov random field models for restoration of ultrasound images. To our knowledge, Hokland and Kelly (1996) presents the first attempt, building a discrete Markov random field model using a well-known model for radio frequency signals. Husby et al. (2001) and Langø et al. (2001) extends the model to the continuous case, obtaining a more stable and computationally tractable model. Kao et al. (1998) use the same model to develop a discrete MRF model for restoration of B-scan images.

Perhaps more than restoration, tasks such as estimation of areas and volumes, counting and classifying objects, and measuring distances between shapes, are the force of Bayesian methods. There has been some work in that direction, estimating areas of cross-sections of blood vessels (Hansen et al., 2000), and estimating boundaries between tissue types in sheep (Glasbey, 1998). In the above works the imaging models were simple, but as such tasks become more complicated, there is a need for good imaging models to obtain a desirable accuracy of the estimates. Thus further work on developing good statistical models for ultrasound images is needed.

This chapter is organised as follows: First we review the model for diffuse scattering in radio frequency images, and then briefly discuss some other scattering types, in particular the specular reflection occurring at interfaces between tissues.

3.2 A Model for Ultrasonic Image Formation

In this section we derive the model for ultrasonic image formation. The model is based on results from Goodman (1975), Wagner et al. (1987), Lizzi, Greenbaum, Felappa and Elbaum (1983), and Kao et al. (1998); for a thorough introduction to ultrasound imagery, see eg. Angelsen (2000).

The source of information in ultrasonic imaging is the radio frequency signal: The recorded acoustic echo from focused high-frequency beams transmitted into the imaged medium from a transducer. The radio frequency signal has a random character coming from two different sources: the coherent interference of the backscattered echo from a collection of scatterers, and diffraction in the imaging system (Goodman, 1975). Trying to model this process in its full generality is a very ambitious task, and the model used below ignore some physical factors such as absorption at the scatterers, multiple scattering, etc (Kao et al., 1998).

The geometry of the problem is shown in Figure 3.1; we assume a planar transducer lying in the xy-plane, and a short acoustic pulse being transmitted in the positive z-direction into the body. Kao et al. (1998) argue that the received echo from depth z_0 at a transducer location (x_0, y_0) is

$$\zeta\left(x_0, y_0, z_0\right) \approx \int_V \frac{\kappa^2}{z^2} T(x,y,z) \exp\left(i\frac{\kappa}{z}\left(x^2+y^2\right)\right) \exp\left(i2\kappa z\right) h(x_0-x, y_0-y, z_0-z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \tag{3.1}$$

where κ is the wave number, h is the point spread function of the imaging system,

$$T = \frac{1}{\rho_0 c_0}\frac{\partial \rho c}{\partial z}$$

is the fractional variation of the acoustic impedance ρc, and V is the imaged volume. Note that we assume a stationary point spread function, and that the derivation rely on the assumption that the density ρ and wave propagation speed c are assumed to satisfy $\rho(x,y,z) = \rho_0 + \Delta\rho(x,y,z)$ and $c(x,y,z) = c_0 + \Delta c(x,y,z)$, where $|\Delta\rho(x,y,z)| \ll \rho_0$ and $|\Delta c(x,y,z)| \ll c_0$.

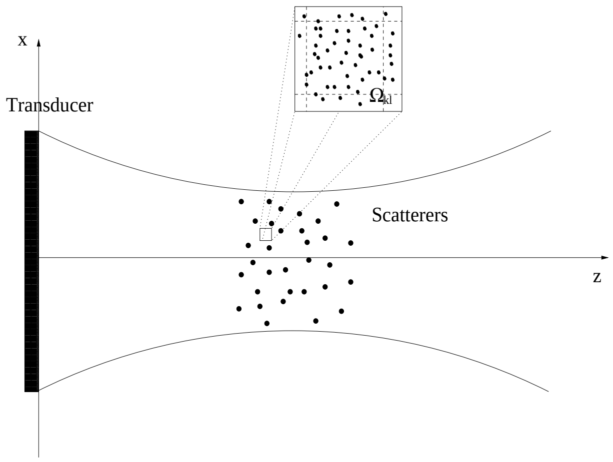


Figure 3.1: Geometry of the imaging system. The enclosing curves represents the imaged area.

In reality the signal is recorded on a discrete grid, and so we fix a y_0 and divide the corresponding xz-plane into square bins Ω_{kl} corresponding to pixels (k,l) of the recorded image. Using Eq. (3.1) and assuming separability of the point spread function, the received signal $\xi_{kl} = z_l\, \zeta(x_k, y_0, z_l)$ at pixel (k, l) is

$$\xi_{kl} \approx z_l \sum_{k',l'} \int_{\Omega k'l'} \int_{-\infty}^{\infty} \frac{\kappa^2}{z^2} T\,(x,y,z) \exp\left(i\frac{\kappa}{z}\left(x^2+y^2\right)\right) \exp\left(i2\kappa z\right)$$
$$\times\, h_x(x-x_k)h_y(y-y_0)h_z(z-z_l)\, \mathrm{d}y\, \mathrm{d}x\, \mathrm{d}z. \quad (3.2)$$

Note that ξ_{kl} is defined as $z_l\, \zeta$ since the decay of ζ with z is compensated by a time gain control (Kao et al., 1998).

The properties of this signal depends on the scatterers or reflectors in the imaged tissue. The simplest case occurs when there is a large number M of discrete, uniformly distributed scatterers in the volume V. We can then write the fractional variation as

$$T(x,y,z) = \sum_{m=1}^{M} \tilde{t}_m \delta(x-x_m, y-y_m, z-z_m),$$

where (x_m, y_m, z_m) is the position of scatterer m, and δ is Dirac's delta function. Furthermore, by assuming that the cells Ω_{kl} are small relative to the point spread function h, we can use the approximations $h_z(z-z_l) \approx h_z(z_{l'} - z_l)$ and $h_z(x-x_k) \approx h_x(x_{k'} - x_k)$, where $(x_{k'}, z_{l'})$ is the centre of cell $\Omega_{k'l'}$. We also assume $h_z(z-z_l)/z^2 \approx h_z(z-z_l)/z_l^2$, getting

$$\begin{aligned}\xi_{kl} &\approx \sum_{k'l'} h_x(x_{k'}-x_k)h_z(z_{l'}-z_l)\frac{\kappa^2}{z_l} \\ &\quad \times \left(\sum_{m\,:\,(x_m,z_m)\in\Omega_{k'l'}} \tilde{t}_m\, h_y(y_m-y_0)\, \exp\left(i2\kappa z_m\left(1+\frac{x_m^2+y_m^2}{2z_m^2}\right)\right)\right) \\ &= \sum_{k'l'} h_{k'l'}^{(k,l)} \left(\sum_{m\,:\,(x_m,z_m)\in\Omega_{k'l'}} t_m\, \exp(i\phi_m)\right),\end{aligned} \quad (3.3)$$

where $t_m = \frac{\kappa^2}{z_l}\, h_y(y_m - y_0)\tilde{t}_m$ and $\phi_m = 2\kappa z_m\left(1 + \frac{x_m^2+y_m^2}{2z_m^2}\right)$.

Now denote the bracketed term in Eq. (3.3) by $\xi^*_{k'l'}$: the true, unobserved radio frequency signal at position (k', l'). If we assume that the phases ϕ_m are independent and uniformly distributed on $[0, 2\pi)$, and independent of the amplitudes t_m, we easily see that for all (k, l)

$$\begin{aligned}
\mathrm{E}\,\Re\xi^*_{kl} &= \sum_m \mathrm{E}\,t_m \mathrm{E}\,\cos\phi_m = 0 \\
\mathrm{E}\,\Im\xi^*_{kl} &= \sum_m \mathrm{E}\,t_m \mathrm{E}\,\sin\phi_m = 0 \\
\mathrm{E}\,(\Re\xi^*_{kl})^2 &= \sum_m \sum_{m'} \mathrm{E}\,t_m t_{m'} \mathrm{E}\,\cos\phi_m \cos\phi_{m'} = \frac{1}{2}\sum_m \mathrm{E}\,t_m^2 \\
\mathrm{E}\,(\Im\xi^*_{kl})^2 &= \sum_m \sum_{m'} \mathrm{E}\,t_m t_{m'} \mathrm{E}\,\sin\phi_m \sin\phi_{m'} = \frac{1}{2}\sum_m \mathrm{E}\,t_m^2 \\
\mathrm{E}\,\Re\xi^*_{kl}\Im\xi^*_{kl} &= \sum_m \sum_{m'} \mathrm{E}\,t_m t_{m'} \mathrm{E}\,\cos\phi_m \sin\phi_{m'} = 0,
\end{aligned}$$

where $\Re\xi^*_{kl}$ and $\Im\xi^*_{kl}$ denote the real and imaginary parts of ξ^*_{kl}, respectively. Furthermore, as the t_m's and ϕ_m's are independent, the central limit theorem gives that $\Re\xi^*_{kl}$ and $\Im\xi^*_{kl}$ are asymptotically Gaussian with zero mean and variance

$$\sigma^2_{kl} = \lim_{M_{kl}\to\infty} \frac{1}{2M_{kl}} \sum_{m=1}^{M_{kl}} \mathrm{E}\,t_m^2. \tag{3.4}$$

Concerning the correlation structure of the process ξ^*, we see that

$$\begin{aligned}
\mathrm{E}\,\xi^*_{kl}\overline{\xi^*_{k'l'}} &= \sum_m \sum_{m'} t_m t_{m'} \exp\left(i\left(\phi_m - \phi_{m'}\right)\right) \\
&= \delta(k-k', l-l') \sum_m \mathrm{E}\,t_m^2 \\
&\approx \sigma^2_{kl}\delta(k-k', l-l'),
\end{aligned}$$

and thus ξ^* is a circular complex Gaussian Markov Random field with no spatial interactions. This is referred to as the *diffuse scattering* case, and is important because diffuse scattering is the main component of most ultrasonic signals, and because it is a simple case for which there exists a tractable mathematical model. The model is believed to be very accurate for body liquids (Insana, Wagner, Garra, Brown and Shawker, 1986), and works well for most soft tissue. Below we summarise the conditions for the diffuse scattering model:

Conditions for the diffuse scattering model

A. *All amplitudes and phases are independent, ie. all areas are uncorrelated, and the strength of the echo from a given scatterer is independent of its phase.*

B. *The phases are uniformly distributed on* $[0, 2\pi)$*, ie. the structure of the tissue is rough compared to the wavelength of the incoming beam.*

C. *The imaged tissue consists of randomly distributed scatterers.*

D. *The resolution cell of the imaging system is small relative to the size of the point spread function.*

E. *The spatial variation in scatter density is small within the support of the resolution cell.*

F. *There is a large number of scatterers per resolution cell.*

3.3 Specular Reflection

Other well known types of scattering are quasi-periodic scattering and specular reflection. Quasi-periodic scattering occur in tissues where scatterers are organised in regular structures, such as in the liver and the collagenous fibres surrounding muscle bundles (Hokland, 1995). In the presence of a coherent scatter component either the real or imaginary part of the radio frequency signal ξ^* will have a non-zero mean. We will not consider quasi-periodic scattering further in this report.

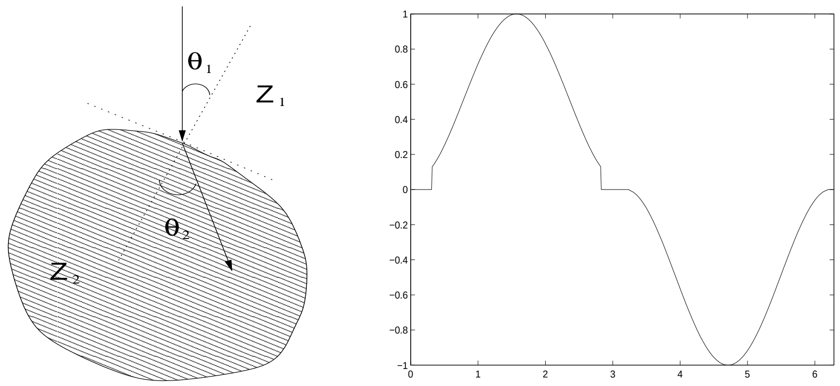
Specular reflection occur at the interface between two tissue types with sufficiently large differences in acoustic impedance (Christensen, 1988). The effect is strongest when the incoming pulse is approximately normal to the tissue interface; in that case the echo can then be much stronger than the diffuse component. The specular reflection is given by the reflection coefficient (Christensen, 1988)

$$R = \frac{\frac{Z_2}{\cos\theta_2} - \frac{Z_1}{\cos\theta_1}}{\frac{Z_2}{\cos\theta_2} + \frac{Z_1}{\cos\theta_1}}, \tag{3.5}$$

where Z_1 and Z_2 are the acoustic impedances on the incident and transmitted side of the interface, respectively; and θ_1 and θ_2 are the angles between the interface and the incoming and transmitted beams, see Figure 3.3 (a). The incident and transmitted beam angles θ_1 and θ_2 are related by Snell's law $\sin\theta_1/\sin\theta_2 = c_1/c_2$, where c_1 and c_2 are the beam velocities. The coefficient in Eq. (3.5) measures the total reflection, but the amount of the reflection returning to the ultrasound transducer will depend on the angle of incidence, whence we follow Hokland and Kelly (1996) and model the specular reflection returning to the transducer as

$$r_i = \frac{\frac{Z_2}{\cos\theta_2} - \frac{Z_1}{\cos\theta_1}}{\frac{Z_2}{\cos\theta_2} + \frac{Z_1}{\cos\theta_1}} \cos^\gamma\theta_1, \tag{3.6}$$

for a site i on the tissue interface. Figure 3.3 (b) shows the specular reflection coefficient R measured counter-clockwise along a circle, starting on the middle right of the circle. The impedances are $Z_1 = 162.0\,\mathrm{kg/(s\cdot cm^2)}$ and $Z_2 = 166.6\,\mathrm{kg/(s\cdot cm^2)}$ corresponding to muscle and blood (Christensen, 1988). Note that when the beam velocity is higher in the transmitted medium than in the incident medium, there exists critical angles for which no energy is transmitted.



(a) (b)

Figure 3.2: Specular reflection. (a) Diagram showing the path of the ultrasound pulse through a boundary between tissues having acoustic impedances Z_1 and Z_2. (b) The specular reflection coefficient R measured counter-clockwise from the middle right of a circle. The impedances inside and outside the circle are $Z_2 = 166.6$ kg/(s $\cdot$ cm^2) and $Z_1 = 162.0$ kg/(s $\cdot$ cm^2), respectively.

Note that it is difficult to model the specular reflection without having an explicit model for the edges, and this is a strong motivation for using deformable templates for modelling tissue interfaces in ultrasound images.

Chapter 4

Edge-Preserving Restoration

In this chapter we discuss prior models for edge-preserving restoration of images. Such models play a key role in our Bayesian model for restoration of radio-frequency ultrasound images, and will be used later in Chapter 7. We will in particular be concerned with constructing algorithms for efficient sampling of such models; this is discussed in Section 4.3.

4.1 Introduction

We will in this chapter consider restoration of images that are piecewise smooth, ie. consisting of smooth regions separated by edges. This problem was introduced in Geman and Geman (1984), and is thoroughly discussed in Geman and Reynolds (1992) and Geman and Yang (1995), on which we base our exposition.

We will consider prior densities of the form $\pi(x) \propto \exp(-H(x))$, where

$$H(x) = \beta \sum_{C \in \mathcal{C}} \omega_C \phi\left(D_C(x)\right), \tag{4.1}$$

ϕ and D_C are real functions, the ω_Cs are positive weights, and β is a positive scaling parameter. Following Hurn and Jennison (1996), we define the neighborhood of a site $s \in S$ to be its eight nearest sites, and define first and second order cliques consisting of pairs and triplets of pixels. The $M = 14$ resulting cliques are shown in Figure 4.1. Using eight neigbors increases the model complexity compared to the four neighbors usually chosen, but more importantly it eliminates the bias towards horizontal and vertical edges. We impose roughness penalties by defining the D_C's to be discrete difference operators corresponding to approximations of first- and second-order derivatives, ie.

$$D_s^{(m)}x = (x_s - x_t)/\delta \qquad m = 1, \dots, 4 \tag{4.2}$$
$$D_s^{(m)}x = (x_t - 2x_s + x_u)/\delta \qquad m = 5, \dots, 8 \tag{4.3}$$
$$D_s^{(m)}x = (x_s - x_t - x_v + x_u)/\delta \qquad m = 9, \dots, 14 \tag{4.4}$$

The weights are chosen to accomodate the differences in distance between diagnoal and vertical or horizontal sites; for instance, Hurn and Jennison (1996) choose $\omega_1 = \omega_2 = 1$, $\omega_3 = \omega_4 = 1/\sqrt{2}$, $\omega_5 = \omega_6 = \omega_9 = 1$, $\omega_7 = \omega_8 = \omega_{10} = 1/2$, and $\omega_{11} = \dots = \omega_{14} = 1/\sqrt{2}$.

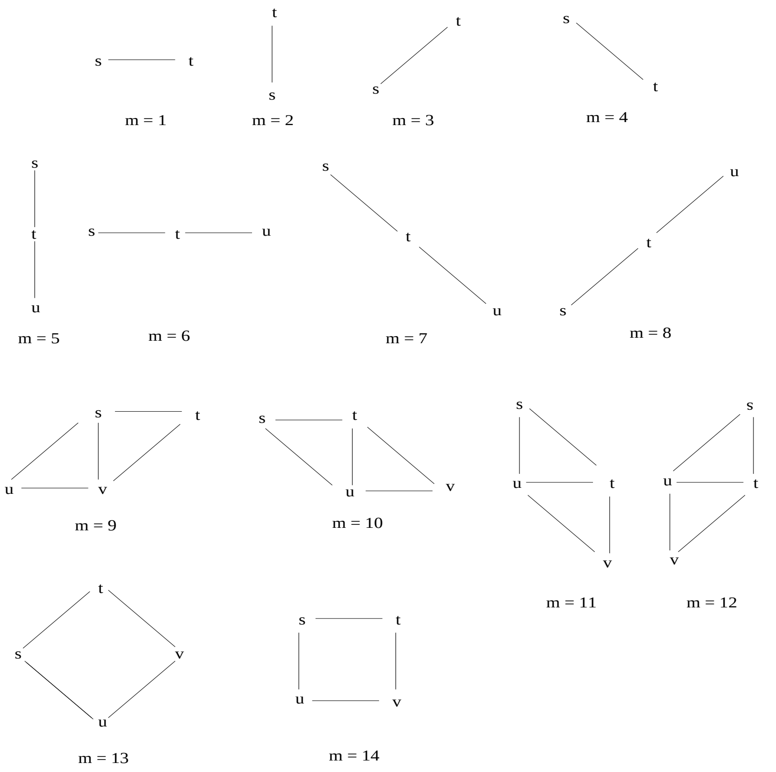


Figure 4.1: Cliques for the prior density Eq. (4.1).

Using the above definitions we write the prior energy as

$$\Phi(x) = \beta \sum_{m=1}^{M} \omega_m \sum_{s \in S} \phi\left(D_s^{(m)}x\right). \tag{4.5}$$

By chosing an appropriate potential function $\phi(\cdot)$ we are able to punish large local variations in the first and second derivatives, thus obtaining fields that are locally constant or planar. The choice of the potential function has implications for the properties of the solution, and will be the topic of the next section.

4.2 Edge-Preserving Functionals

The potential function $\phi(\cdot)$ in the prior density (4.5) imposes smoothnes conditions on the discrete derivatives $D^{(m)}x$. It is natural to assume ϕ to be symmetric, so that positive and negative gradients of equal magnitude are punished equally; we will also assume it to be continuous and nondecreasing on $\mathbb{R}^+$ and to have value zero at the origin. Note that the last condition is not really neccessary since any offset would be incorporated in the normalizing constant.

One particular and important choice is the quadratic potential $\phi(u) = u^2$ which leads to a Gaussian prior; and, combined with a Gaussian likelihood, an unimodal aposteriori distribution. This of course simplifies the optimisation and sampling problems. Gaussian models can be efficiently sampled using sparse matrix methods (Rue, 2001), or FFTs in the special case of toroidal boundary conditions (Wood and Chan, 1994). However, this choice is not suited for estimation of piecewise constant or planar fields, as the rapid growth as $u \to \infty$ severly punishes the intensity jumps that occur across edges. In addition, the slow variation around origo might cause excessive smoothing and interpolation.

The choice of potential functions for edge-preserving restoration is widely discussed in the litterature. One possibility is to explicitly model the edges using discrete line processes (Geman and Geman, 1984) or deformable templates (Grenander and Miller, 1994). We will in this section consider *implicit* line models only; the use of deformable templates in the modelling of edges and discintinuities is discussed later, in Chapter 5. Concerning implicit models, one school, starting with Geman and McClure (1987), consider non-convex potential functions having a horizontal asymptote, while others use convex functions (Green, 1990; Bouman and Sauer, 1993; Schultz and Stevenson, 1995). Charbonnier et al. (1997) gives a unifying definition of edge-preserving functionals and shows which conditions need to be fulfilled.

We will largely follow Geman and Yang (1995) and consider potential functions in the class

$$\mathcal{E} = \left\{ \phi(\cdot) \in \mathrm{C}^{(0)}(\mathbb{R}) \,|\, \phi(0) = 0, \phi(u) = \phi(-u), \lim_{u\to\infty} \phi(u) < \infty, \frac{d\phi}{du} \geq 0, u \in \mathbb{R}^+ \right\}. \tag{4.6}$$

The arguments for this class of model are largely based on heuristics, but it is clear that the finite limit and slow growth for large u ensures that intensity jumps over edges are not too severely punished. The following example, taken from Blake and Zisserman (1987), shows that using potential functions in $\mathcal{E}$ has links to both line processes and robust inference.

Example 4.2.1 Let u be a Markov Random field with neighborhood relation $\sim$, and define the line process

$$l_{ij} = \begin{cases} 1 & \exists \text{ edge between } u_i \text{ and } u_j \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, define the prior energy functional

$$\Phi(u,l) = \sum_{i \sim j} \left((u_i - u_j)^2 - 1 \right) (1 - l_{ij})$$

smoothing within the disjoint regions defined by the line process l. Then Blake and Zisserman (1987) observed that

$$\begin{aligned} \inf_l \Phi(u,l) &= \sum_{i \sim j} \left((u_i - u_j)^2 - 1 \right) \mathbb{I}_{[(u_i - u_j)^2 < 1]} \\ &= \sum_{i \sim j} \left((u_i - u_j)^2 - 1 \right)^- = \Phi^*(u), \end{aligned}$$

where $x^- = \min(x, 0)$. Thus $\inf_u \inf_l \Phi(u,l) = \inf_u \Phi^*(u)$, and so in terms of model behaviour there is no need to model the edges explicitly; one can instead use the truncated quadratic Φ^*.

In addition to the behaviour of $\phi(\cdot)$ as u grows large, the behaviour around the origin is important. Geman and Reynolds (1992) advocate functions that are strictly concave, such as

$$\phi(u) = \frac{-1}{1 + |u|}, \qquad \phi(u) = \frac{|u|}{1 + |u|}, \tag{4.7}$$

and base their argumentation on considerations of coordinate-wise minima. Let x^* be a coordinate-wise minimum of the posterior energy, and consider a small pertubation $x_i^* + tu$ toward the data y_i. This will lead to a decrease of order tu in the likelihood component, but an increase in the prior energy since $\phi'(0+) > 0$. By appropriately choosing the scaling parameter β the combined effect will very likely be an increase of the aposteriori likelihood. This is in contrast to the case where $\phi'(0) = 0$, when there will be interpolation between neighbouring sites.

Geman and Yang (1995) study the behaviour of concave functions ϕ applied to a one-dimensional signal $x = (x_0, \dots, x_n)$. In particular, they show that the choice of clique functions affects the properties of the minimizer of the aposteriori energy (4.5), as shown in the following result.

Proposition 1 (Geman and Yang (1995))
Let $\phi(u) \in C^{(2)}(\mathbb{R})$ be symmetric with $\frac{d^2\phi}{du^2} < 0$ on $(0, \infty)$, and define

$$H(x) = \sum_{i=0}^{n-1} \omega_1 \phi\left(\frac{x_{i+1} - x_i}{\delta_1}\right) + \sum_{i=1}^{n-1} \omega_2 \phi\left(\frac{x_{i+1} - 2x_i + x_{i-1}}{\delta_2}\right), \tag{4.8}$$

where δ_1 and δ_2 are positive constants. Then we have the following.

(i) If $\omega_1 > 0$ and $\omega_2 = 0$, the global minimizer of H subject to $x_0 = 0$ and $x_n = c > 0$ is piecewise constant with exactly one jump.

(ii) If $\omega_1 = 0$ and $\omega_2 > 0$, the global minimizer of H subject to $x_0 = x_1 = 0$ and $x_{n-1} = x_n = c$ has three linear segments, with the least possible fluctuation in slopes.

(iii) If $\omega_1 > 0$ and $\omega_2 > 0$, the global minimizer of H subject to $x_0 = x_1 = 0$ and $x_{n-1} = x_n = c$ is piecewise linear with exactly three segments.

Proof. Represent x as a continous and picewise linear graph over $[0, n]$.

(i) Obviously, x can have at most $n - 1$ segments. Let $m \leq n - 1$ be the number of segments, then $H(x) = \omega_1 \sum_{j=1}^{m} \phi(c_j/\delta_1)$ where $\sum c_j = c$. But since ϕ is concave we have

$$\phi(c/\delta_1) < \phi(tc/\delta_1) + \phi((1-t)c/\delta_1), \quad \forall\, 0 < t < 1,$$

and thus the energy is uniquely minimized by choosing one jump.

(ii) Let the $m \leq n - 1$ linear segments have slopes $a_1, \ldots, a_m$. Then

$$\omega_2^{-1} H(x) = \phi\left(\frac{a_2}{\delta_2}\right) + \phi\left(\frac{a_{m-1}}{\delta_2}\right) + \sum_{j=2}^{m-2} \phi\left(\frac{a_{j+1} - a_j}{\delta_2}\right),$$

since $a_1 = a_m = 0$. Again, because of the concavity of ϕ, H is minimized by setting $a_2 = \ldots a_{m-1} = c/(n-1)$, see Figure 4.2.

(iii) See Geman and Yang (1993).

□

Further insight might be gained from considering the asymptotic behaviour of the model. Let $\phi(\cdot)$ be the Geman-McClure functional (Table 4.1) and assume $\omega_1 = 1$ and $e_2 = 0$ in Eq. (4.8). Then Rosati (2000) shows that under some regularity consitions, $H(x)$ converges to the functional

$$E(x) = \int (x'(t))^2\, dt + \delta\, \mathrm{card}\,(S_x),$$

where S_x is the set of singularities (ie. jumps) for x. In two dimensions the last term generalizes to $\delta\Delta_1(S_x)$, where $\Delta_1(\cdot)$ is a one dimensional Hausdorff type

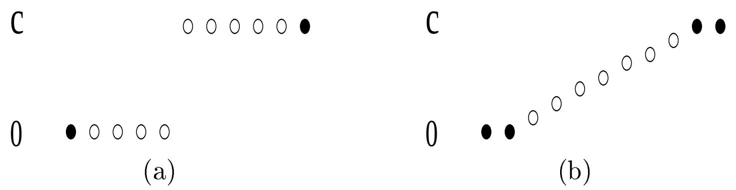


Figure 4.2: Minimal energy configurations for a 1D signal. (a) First order model. (b) Second order model.

measure called the cab-driver measure (Chambolle, 1995). The resulting model is a modified form of the well known Mumford-Shah functional (Mumford and Shah, 1989).

Note that most functions in the class $\mathcal{E}$ are convex on some interval $(-t, t)$ around origo, but as noted in Geman and Yang (1995) this is only a matter of scaling. If $\phi(u)$ is convex on $(-t, t)$, then $\phi(u/\epsilon)$ is convex on $(-c\epsilon, c\epsilon)$, and thus the above results are valid for all $\phi \in \mathcal{E}$ by chosing appropriate parameters δ_1 and δ_2.

Charbonnier et al. (1997) argues that certain non-quadratic convex potential functions will perform equally well, at least when considering optimization or MAP-estimation. (As an aside, note that it is generally difficult to say something sensible about the choice of ϕ when we consider estimators such as the posterior mean, since we then need to consider the distribution as a whole rather than just the mode). The conditions for an edge-preserving potential function are as follows:

Conditions for edge-preserving functionals
(Charbonnier et al., 1997)

A. $\phi(u) \geq 0 \quad \forall u$ with $\phi(0) = 0$.

B. $\phi(u) = \phi(-u)$.

C. ϕ continuosly differentiable.

D. $\phi'(u) \geq 0 \quad \forall u \geq 0$.

E. $\frac{\phi'(u)}{2u}$ continuous and strictly decreasing on $[0, \infty)$.

F. $\lim_{u\to\infty} \frac{\phi'(u)}{2u} = 0$.

G. $\lim_{u\to 0^+} \frac{\phi'(u)}{2u} = M < \infty$.

Condition G is not satisfied by the strictly concave potential functions in Eq. (4.7), but most functions used in the litterature satisfy the above condition; see Table 4.1 for some examples.

Table 4.1: Some edge preserving functionals.

Potential function	Reference
$\min(1, u^2)$	Blake and Zisserman (1987)
$\frac{u^2}{1+u^2}$	Geman and McClure (1987)
$\log\cosh u$	Green (1990)
$\log(1+u^2)$	Hebert and Leahy (1989)
$2\sqrt{1+u^2}-2$	Charbonnier (1994)

4.3 Sampling and Optimization

One of the disadvantages of non-convex prior functionals is that they introduce additional computational problems. Deterministic minimization of a non-convex function is a hard problem, and also stochastic algorithms such as the Metropolis-Hastings sampler tend to be inefficient, especially when the likelihood introduces strong spatial interactions. One promising way of dealing with this problem was introduced by Geman and Reynolds (1992), who shows that for a large family of potential functions ϕ there exist a dual model that in some sense has the same distributional properties (ie. the same mode) as the original model, but that has better computational properties and hopefully better sampling properties. Their main idea is to model edges explicitly by exploring correspondences similar to the one in Example 4.2.1.

Most of the work in this field concerns MAP-estimation or determinstic optimization, and we will start by reviewing some of the important results in this area. At the end of the section we give some results in a more "statistical" setting.

4.3.1 Half-quadratic optimization

The concept of half-quadratic optimization was introduced in Geman and Reynolds (1992), and the idea is to introduce a dual prior functional $\Phi^*(x, b)$ that depends both on x and an auxillary Markov field b. The world "half-quadratic" stems from the fact that $\Phi^*(x, b)$ is constructed so that it is quadratic in x. This is desirable in at least two different respects; the prior energy has an unique minimizer x^*, and the distribution of x given b is Gaussian. As mentioned above there exist several algorithms for effiecient sampling of Gaussian Markov random fields; in particular is it possible to do global updates, a feature that might be important in spatially structured models, see eg. Knorr-Held and Rue

(2002) and Rue (2001).

In order that optimization of the dual model shall be equivalent to optimization in the original one, Φ^* is constructed such that

$$\Phi(x) = \inf_b \Phi^*(x, b), \tag{4.9}$$

implying that $x^* = \arg\inf_{x,b} \Phi^*$ is the global minimizer of both Φ and Φ^*. The auxillary variables in b can be interpreted as edge variables much in the same way as the line process in Example 4.2.1; but unlike the edge variables introduced in Geman and Geman (1984), there are no interactions between them, and they are not constrained to lie in the interval $[0, 1]$.

Given the above conditions, the dual model Φ^* can be constructed using the Legendre transform, as in Geman and Yang (1995).

DEFINITION 2
Let f, g be functions on $\mathbb{R}$. The Legendre transform *of f is*

$$f^*(v) = \sup_{u\in\mathbb{R}} (uv - f(u)).$$

The pair (f, g) is a Legendre pair *if both functions are convex, $f = g^*$ and $g = f^*$.*

Using these definitions we have the following result:

PROPOSITION 2 (GEMAN AND YANG (1995))
Let the functional Φ be as in Eq. (4.5) and define

$$\Phi^*(x, b) = \beta \sum_{m=1}^{M} \omega_m \sum_{s\in S} \left(\frac{1}{2} \left(D_s^{(m)} x - b_s^{(m)} \right)^2 + \psi\left(b_s^{(m)}\right) \right), \tag{4.10}$$

where $b = (b^{(1)}, \ldots, b^{(M)}) \in \mathbb{R}^{nM}$ and ψ is a real-valued function.

If

$$\frac{u^2}{2} - \phi(u) \quad \text{and} \quad \frac{v^2}{2} + \psi(v)$$

is a Legendre pair, then

$$\Phi(x) = \inf_b \Phi^*(x, b).$$

Proof. Assume that $\phi(\cdot)$ and $\psi(\cdot)$ are given, and define $f_\phi(u) = u^2/2 - \phi(u)$ and $g_\psi(v) = v^2/2 + \psi(v)$. If (f_ϕ, g_ψ) is a Legendre pair it is easy to show that

$$\phi(u) = \inf_v \left(\frac{(u-v)^2}{2} + \psi(v) \right),$$

from which it follows that

$$\begin{aligned}\inf_b \Phi^*(x,b) &= \inf_b \beta \sum_{m=1}^{M} \omega_m \sum_{s\in S} \left(\frac{1}{2}\left(D_s^{(m)}x - b_s^{(m)}\right)^2 + \psi\left(b_s^{(m)}\right)\right) \\ &= \beta \sum_{m=1}^{M} \omega_m \sum_{s\in S} \inf_{b_s^{(m)}} \left(\frac{1}{2}\left(D_s^{(m)}x - b_s^{(m)}\right)^2 + \psi\left(b_s^{(m)}\right)\right) \\ &= \beta \sum_{m=1}^{M} \omega_m \sum_{s\in S} \phi\left(D_s^{(m)}x\right) = \Phi(x).\end{aligned}$$

□

Barone (1999) study deterministic minimization of the above model. He shows that under mild conditions it is only neccesary to minimize with respect to the auxillary variable b; given the minimizer $\hat{b}$ there exists an unique $\hat{x} = x(\hat{b})$ so that $(\hat{x}, \hat{b})$ is a global minimizer of the aposteriori energy. However, when ψ is non-convex, the minimizer $\hat{b}$ is not neccesarily easy to find.

4.3.2 Bayesian methods

In the engineering litteratur focus has often been on MAP-estimation. However, from a statistical perspective it is often desirable to know more about the posterior distribution, and from that point of view other estimators, such as the posterior mean are more sensible to use. In the following we show how the above framework can be modified for estimation of the posterior mean.

Again, given a non-quadratic prior functional $\Phi(x)$ we want to find a dual functional $\Phi^*(x,b)$ which is quadratic in x and so that x has the right marginal distribution. Let

$$\pi(x) = \frac{\exp\left(-\Phi(x)\right)}{\int \exp\left(-\Phi(x)\right) dx} \quad \text{and} \quad \pi^*(x,b) = \frac{\exp\left(-\Phi^*(x,b)\right)}{\iint \exp\left(-\Phi^*(x,b)\right) dx\, db},$$

where Φ and Φ^* are as defined in Eq. (4.5) and Eq. (4.10). Then we have the following result

PROPOSITION 3 (GEMAN AND YANG (1995))
Let $\psi(\cdot)$ satisfy

$$c\,\phi(u) = -\ln \int \exp\left(-\frac{c}{2}(u-v)^2 - c\,\psi(v)\right) dv \tag{4.11}$$

for some $c > 0$. Then

$$\pi(x) = \int \pi^*(x,b)\, db.$$

Proof. A straightforward calculation shows that

$$\begin{aligned}
\int \pi^*(x,b)\,db &= \int \exp\left(-\beta \sum_m \omega_m \sum_s \left(\frac{1}{2}\left(D_s^{(m)}x - b_s^{(m)}\right)^2 + \psi(b_s^{(m)})\right)\right) db \\
&= \prod_{m=1}^{M} \prod_{s\in S} \int \exp\left(-\beta\omega_m \left(\frac{1}{2}\left(D_s^{(m)}x - b_s^{(m)}\right)^2 + \psi(b_s^{(m)})\right)\right) db_s^{(m)} \\
&= \prod_{m=1}^{M} \prod_{s\in S} \exp\left(-\beta\omega_m \phi\left(D_s^{(m)}x\right)\right) = \pi(x).
\end{aligned}$$

□

Consequently $\mathrm{E}_{\pi^*}(x \mid y, b) = \mathrm{E}_{\pi}(x \mid y)$, and we can utilize the nice sampling properties of Φ^* to estimate the mean of x under π. As above, this is especially desirable if the likelihood is Gaussian or near Gaussian, since we then can do block sampling. Some experiments using this method are shown below, in Chapter 7.

The function ψ can be found numerically given ϕ, but in most cases it is easier to go the other way around. The following results shows that if ψ is in the family $\mathcal{E}$ of edge-preserving functions, then ϕ is "almost" in $\mathcal{E}$.

Proposition 4
Let $\psi \in \mathcal{E}$ and ϕ be related by Eq. (4.11). Then there exists a constant $k \in \mathbb{R}$ so that $\phi + k \in \mathcal{E}$.

Proof. We need to show that $\phi(u)$ is even, non-decreasing on $[0,\infty)$, and finite as $u \to \infty$. The first part is obvious, so we start by showing that $\phi'(u) > 0$ on $\mathbb{R}^+$. Let $\lim_{v\to\infty}\psi(v) = M$ and define

$$I(u) = \int_{\mathbb{R}} e^{-\frac{1}{2}(u-v)^2 - \psi(v)} dv.$$

Then $\sqrt{2\pi}e^{-M} \le I \le \sqrt{2\pi}$, and for $u > 0$

$$\frac{dI}{du} = \int_{\mathbb{R}} (v-u) e^{-\frac{1}{2}(v-u)^2 - \psi(v)} dv \le \sqrt{2\pi}\left(e^{-M} - 1\right) u < 0.$$

Thus for $u > 0$, $\phi'(u) = -\frac{1}{I(u)}\frac{dI}{du} > 0$.

Furthermore, since $\phi(u)$ is non-decreasing and bounded, $\phi(0) \ge -\ln\sqrt{2\pi}$ and $\lim_{u\to\infty}\phi(u) \le M - \ln\sqrt{2\pi}$. Hence ϕ has a finite limit and there exists an $k \le \ln\sqrt{2\pi}$ such that $\phi + k \in \mathcal{E}$. □

It can in fact be shown that ϕ satisfies all the conditions (i) – (vii) in Section 4.2. As an example Figure 4.3 shows ϕ computed when $\psi(v) = |v|/(1 + |v|)$.

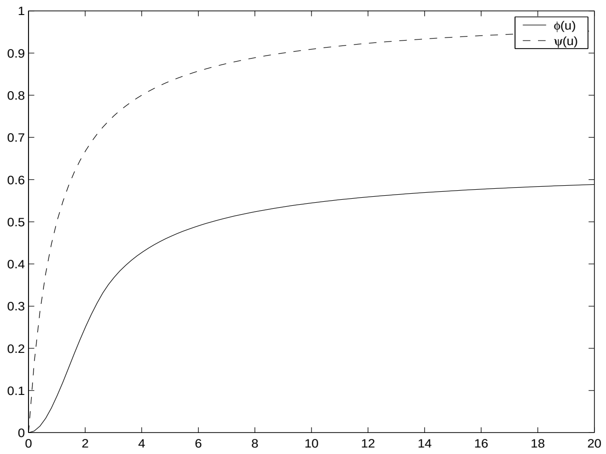


Figure 4.3: Functions $\phi(v)$ and $\psi(v) = |v|/(1 + |v|)$ related by Eq. (4.11).

CHAPTER 5

High-Level Prior Models for Object Recognition

In this chapter we discuss high-level models for object recognition in images, with a special focus on the deformable template model we will use for contour detection in Chapter 8.

5.1 Introduction

High-level models makes definitive statements about the image under study in the sense that it is the imaged objects, or their shapes, which are directly modelled. The idea of using compact, parametric models of object shape was pioneered by Ulf Grenander in the late 60s (Grenander, 1967; Grenander, 1969; Grenander, 1981), but it was not until the early 90s advances in computer hardware and stochastic simulation algorithms made the methods applicable in practice (Grenander et al., 1991; Grenander and Miller, 1994). This has opened up for a whole range of applications: By augmenting the imperfect image data with the prior information provided by the high-level model, it becomes possible to locate and characterise the objects present in the image.

The information contained in a high-level model can range from local – eg. imposing smoothness and connectivity constraints – to global: representing the variation in a specific class of shapes. The nature of the model will depend on the application of interest, but in general we are often interested in models that have a generic parameterisation, and are expressive enough to represent a large range of potential shapes, while at the same time constraining the problem to lie within the bounds defined by our prior knowledge. Especially appealing are the so-called prototype-, or deformable template-, models. Here prior knowledge is represented through a reference object I_0. One then assumes that all objects of interest lies in the orbit $\mathcal{G}I_0$ of I_0 under some transformation group $\mathcal{G}$. This group is chosen so as to represent the natural variability in object shape, and is equipped with a probability distribution describing the relative frequency of shapes. Alternatively, a probability distribution could be be assigned to the parameters of the prototype directly, in which case it is called a geometric template model (Jain, Zhong and Dubuisson-Jolly, 1998).

It should be noted that high-level models to a much larger extent than pixel-based models allow structured understanding of the images. For instance does the estimated transformation from template to observed object make classification and detection of pathologies highly possible. An important concept in this respect is the definition of distances between objects in terms of minimal group actions (Grenander, 1993, chap. 12). A further characteristic of high-level models is their capability of including uncertainty over the number of objects in view, requiring more complex sampling algorithms than traditional fixed-dimension approaches.

High-level models have been used for object recognition in many applications, for instance for recognising hands (Grenander et al., 1991), galaxies (Ripley and Sutherland, 1990), pellets (Baddeley and Van Lieshout, 1993), potatoes (Grenander and Manbeck, 1993), cells (Grenander and Miller, 1994; Rue and Hurn, 1999), mushrooms (Mardia, Qian, Shah and de Souza, 1997), fish (de Souza, Kent and Mardia, 1999), roads (Stoica, Descombes and Zerubia, 2000), and blood vessels (Hansen et al., 2000). Other applications include such diverse tasks as target recognition (Miller, Srivastava and Grenander, 1995), object tracking (Blake, Bascle, Isard and MacCormick, 1998; Kervrann and Heitz, 1999), database retrieval (Zhong and Jain, 1997), and disease testing (Grenander and Miller, 1998), to site just a few.

In the rest of this chapter we will give a survey of some of the high-level models in use. The survey is by no means complete; we will focus on work in a statistical setting, excluding much of the computer vision literature. We will also restrict ourselves to 2D applications. Please refer to McInerey and Terzopoulos (1996) and Grenander and Miller (1998) for other reviews of high-level models in biomedical applications.

5.2 Active Contour Models

Active contour models were introduced in Kass, Witkin and Terzeopoulos (1988), who considered curves controlled by an internal energy enforcing smoothness and connectedness, and an external energy attracting the curve to desired image features. Active contour models are mostly considered in a non-statistical setting, but we include a brief introduction since they are widely used, especially in applications where there is no underlying shape information.

Taking a variational approach, Kass et al. (1988) define the internal energy for a curve $v(s) \in \mathbb{C}$ as

$$E(v) = \frac{1}{2} \int_0^1 \left(\alpha(s) \left| \frac{\partial v}{\partial s} \right|^2 + \beta(s) \left| \frac{\partial^2 v}{\partial s^2} \right|^2 \right) ds, \tag{5.1}$$

where the functions α and β control the degree of elasticity and stiffness, respectively. After an appropriate initialisation, the energy is minimised using some optimisation methods, eg. gradient descent. However, because the model contains no global information, the need for a good initialisation is important, lest the contour get stuck in a local minimum. There have been numerous attempts at improving the active contour models, notably Cohen (1991) introduced a "balloon" force that can either inflate or deflate the contour. This will make it easier to escape global minima caused by noise and image artifacts. We refer to Blake and Isard (1998) for a comprehensive introduction to active contour models.

5.3 Parametric Models

Parametric models are used when there exists prior information about the objects under study. In contrast to the local and rather weak information carried by the active contour models, the parametric models encode global information that in some sense constrains the shape of the modelled objects. For instance, one can assume shapes to lie within the class of closed and simple curves, and then impose further structure by penalising deviations from some mean shape, eg. a circle.

We divide the models into four classes: Geometrical models, polygonal models, continuous models, and landmark models. This is not the only possible subdivision, and we do not claim that the different approaches are fundamentally different, nor that all models fit into one of these four groups. However, as a means of organising the material, the subdivision is practical.

5.3.1 Geometrical models

Geometrical template models are defined through parametric curves such as circles, ellipses, and parabolas. Variability is introduced by assigning probability distributions to the parameters of the curves. For instance, Ripley and Sutherland (1990) and Baddeley and Van Lieshout (1993) use circles to model galaxies and pellets, respectively. Yuille, Hallinan and Cohen (1992) and Phillips and Smith (1994) have used geometrical templates for locating facial features. Their model for an eye consists of two parabolas tracing out the top and bottom eyelids, and a circle representing the iris. The parameters are the centre and radius of the circle, together with the parameters of the parabolas.

The model of Phillips and Smith (1994) is designed in a hierarchical fashion with global shape features constraining local features. The distributions for the local variables are given conditional on the global variables in the following hierarchy:

- Boundary for the head
- Boundary for the face, inside the head boundary
- Spatial arrangement of facial features, within the face boundary
- Scaling of facial features, given their positions

- Shapes of the facial features, given location and size.

The use of geometrical templates is restricted to applications where shapes are well defined and can be represented consistently by a low number of curves. However, for well structures objects such as the face, models as the one above give rise to very compact and precise representations.

5.3.2 Polygonal models

One of the most studied applications of high-level models is the detection of featureless planar objects in images, see e.g. (Grenander et al., 1991; Grenander and Manbeck, 1993; Grenander and Miller, 1994; Qian, Titterington and Chapman, 1996; Rue and Syversveen, 1998; Rue and Hurn, 1999). To fix ideas, consider a solid object in the plane, and suppose that its shape can be approximated by an m-sided simple polygon with vertices $v_0^0, \dots, v_{m-1}^0$ placed counterclockwise around the outline. This polygon is the *template*, representing the shape of a typical object. The set of all object shapes is generated by applying stochastic transformations to the template, and thus the most important modelling issue is to specify the transformation group and the probability distribution on this group.

Most work has concentrated on applying transformations to the edges rather than to the vertices directly, but see Kent, Mardia and Walder (1996) for review of different ways of modelling templates. The edges of the outline are

$$e_k^0 = v_{k+1}^0 - v_k^0, \quad k = 0, \dots, m-1, \tag{5.2}$$

where the indices are interpreted modulo m. Following the original work (Grenander et al., 1991), the usual approach is to consider the edge transformation

$$e_k = e_k^0 + r_k \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix} e_k^0 = \begin{pmatrix} 1 + t_{0,k} & t_{1,k} \\ -t_{1,k} & 1 + t_{0,k} \end{pmatrix} e_k^0, \tag{5.3}$$

where the variables r_k and θ_k control the scale and angle in the change of edge e_k. The parametrisation in terms of t_0 and t_1 is used for the following reason: If r_k is proportional to a Rayleigh random variable and θ_k is uniformly distributed on $[0, 2\pi)$, then $t_{0,k}$ and $t_{1,k}$ are independent and identically distributed zero mean Gaussian random variables. Note that in order to ensure that the deformed polygon is closed, we have to impose the constraint

$$\sum_{k=0}^{m-1} e_k = 0. \tag{5.4}$$

This will introduce a dependency between the vectors $t_0 = (t_{0,0}, \dots, t_{0,m-1})$ and $t_1 = (t_{1,0}, \dots, t_{1,m-1})$.

Vertex transformation models are often considered when the objects are star-shaped, for instance Mardia et al. (1997) considers vertices $v_k^0 = \mu^0 +$

$r_k^0(\cos\theta_k^0, \sin\theta_k^0)^T$ that are shifted, scaled, and rotated by the transformation

$$v_k^0 \mapsto v_k = \mu^0 + \mu + \rho_k r_k^0 \begin{pmatrix} \cos(\theta_k^0 + \theta_k) \\ \sin(\theta_k^0 + \theta_k) \end{pmatrix}. \tag{5.5}$$

There is a linear relationship between the edge and vertex transformation vectors, so it is easy to go back and forth between the representations. Hobolth, Kent and Dryden (1999) derives the relation for the special case when the template is a circular polygon, ie. having vertices $v_k^0 = (\cos(2\pi k/m), \sin(2\pi k/m))$. Letting $\rho_k \cos\theta_k = 1 + u_{0,k}$ and $\rho_k \sin\theta_k = -u_{1,k}$, they arrived at

$$2t_{0,k} = u_{0,k+1} + u_{0,k} + \frac{1}{\tan\pi/m}(u_{1,k+1} - u_{1,k}) \tag{5.6}$$

$$2t_{1,k} = \frac{1}{\tan\pi/m}(u_{0,k+1} - u_{0,k}) - u_{1,k+1} - u_{1,k}, \tag{5.7}$$

for $k = 0, \ldots, m-1$.

Rue and Syversveen (1998) and Hansen et al. (2000) consider additive models for vertices and edges, respectively; this type of models is also discussed in Kent et al. (1996). Additive models are simpler and easier to parametrise then the multiplicative models described above, but are somewhat less flexible. A nice feature of the additive edge model in Hansen et al. (2000) is that the closure condition Eq. (5.2) is automatically satisfied by letting the edges follow a certain improper cyclic Gaussian distribution, see details below.

Note that representing the object through the edge vector e removes any location information. This translation invariance is sensible when the locations of objects are unknown. In order to locate the template we place the first vertex v_0 at some location $c \in \mathbb{R}^2$ usually assumed to be uniformly distributed over $\mathbb{R}^2$.

Probability distributions Grenander et al. (1991) let the unconstrained t_0 and t_1 be first order cyclic Gaussian Markov random fields, i.e. they let $t^T = (t_0^T, t_1^T)$ be Gaussian with inverse covariance matrix $Q_t = I_2 \otimes Q$, where Q is a Toeplitz circulant matrix with entries

$$Q_{k,k} = \beta, \qquad Q_{k,k+1 \bmod m} = Q_{k,k-1 \bmod m} = \delta. \tag{5.8}$$

It is well known, see eg. Lauritzen (1996), that under the above model the conditional distribution for $t_{0,k}$, say, is Gaussian with mean and variance

$$\mathrm{E}(t_{0,k} \mid t_{0,-i}) = -\frac{\delta}{\beta}(t_{0,k-1} + t_{0,k+1}), \qquad \mathrm{Var}(t_{0,k} \mid t_{0,-i}) = \frac{1}{\beta}. \tag{5.9}$$

Thus $1/\beta$ can be seen as controlling the deviation from the identity transformation, while $-\delta/\beta$ controls the smoothness of the configuration. The conditions $\beta > 0$ and $\beta + 2|\delta| > 0$ are sufficient to ensure positive definiteness of Q, and hence of Q_t. Note that the circulant representation means that the model is invariant under cyclic permutations of the edges; this is a natural choice when the objects have no apparent landmarks.

Imposing the closure constraint $\sum_k e_k = 0$ reduces the dimension to $2m-2$, and destroys the simple Markov structure of the model. Strictly speaking we also have to introduce an additional constraint, namely that the deformed template be simple. The conditioning only affects the normalising constant, and thus Rue and Hurn (1999) argues that the constraint is not a problem if the model is sampled using Markov chain Monte Carlo methods. One just rejects all non-simple proposals, and uses the unconditional density in the acceptance ratio for the others.

Kent et al. (1996) discuss a similar structure for additive edge models, letting $e \sim \mathrm{N}(e^0, Q_t)$, where $e^T = (e_0^T, \dots, e_{m-1}^T)$, and e^0 is defined similarly. Again conditioning on closure destroys the Markov structure, but this could be alleviated by expressing the model in terms of the vertices v. Letting the first vertex v_0 have an uniform distribution over $\mathbb{R}^2$, Kent et al. (1996) show that the conditional model for e corresponds to an improper second order cyclic Gaussian Markov random field model for v. This distribution has precision matrix $I_2 \otimes P$, where P has entries

$$P_{kl} = \begin{cases} 2\beta - 2\delta & l = k \\ -\beta + 2\delta & l = k-1, k+1 \bmod m \\ -\delta & l = k-2, k+2 \bmod m \end{cases}$$

Hansen et al. (2000) notes that the closure constraint can be automatically satisfied by considering a special case of the additive edge model. Using the same circular template as above, they introduce the $m \times m$ discrete Fourier transform matrix $F = \left\{m^{-1/2} \exp(2\pi i kl/m)\right\}_{0 \le k,l \le m-1}$, and diagonalise Q as $Q = F^* \Lambda F$, where $F^* = \bar{F}^T$ and Λ is a diagonal matrix containing the eigenvalues

$$\lambda_k = \beta + 2\delta \cos 2\pi i k/m, \quad k = 0, \dots, m-1.$$

Denoting the columns of F by $f_0, \dots, f_{m-1}$, and noting that $\sum_k e_k = 0 \Leftrightarrow e^T f_0 = 0$, we see that the eigenvectors $f_1, \dots, f_{m-1}$ span the subspace where closure is satisfied. Thus by setting the first eigenvalue to zero, ie. setting $\delta = -\beta/2$, closedness is automatically guaranteed.

The model in Eq. (5.8) is perhaps too simplistic in some applications, since it restricts the normal and tangential components of the edge transformation vector to be independent and identically distributed. Kent, Dryden and Anderson (2000) use a more general model, with block circulant covariance matrix

$$K = \begin{pmatrix} K(0) & K(1) & \dots & K(n-1) \\ K(n-1) & K(0) & \dots & K(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ K(1) & K(2) & \dots & K(0) \end{pmatrix},$$

having blocks

$$K(k) = \mathrm{E}\left[\begin{pmatrix} t_{0,i} \\ t_{1,i} \end{pmatrix} \begin{pmatrix} t_{0,i+k} & t_{1,i+k} \end{pmatrix}\right] = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix}. \tag{5.10}$$

Again the circulant structure implies cyclical invariance. Kent et al. (2000) discusses the statistical properties of this model in detail, and explores the eigenstructure of the covariance matrix. In its full generality the model has $2m-2$ free parameters, and is thus rather complicated. However, Kent et al. (2000) discusses simplifications such as assuming variability in the normal component only, or using parametric functions for the eigenvalues. An approach somewath similar in spirit is found in Hobolth and Jensen (2000), who model contonuous deformations as the limit of a second order Gaussian Markov random field. Finally, Hobolth, Pedersen and Jensen (2000) and Hobolth, Pedersen and Jensen (2001) consider continuous models using truncated Fourier series to describe the radius vector function of start-shaped objects. Such continuous models are perhaps not well suited for object recognition in blurred and noisy images, but are very useful for eg. describing and classifying objects.

5.3.3 Model parameterisation

As mentioned above, the precision matrix Q is usually chosen to be a band matrix with bandwidth corresponding to the order of the Markov field. It is often of interest to choose a model which behaviour is approximately independent of the number of edges, for instance when considering templates with a variable number of edges, as in Rue and Hurn (1999). Grenander (1993, chap. 11) considers the limit of the cyclic Markov random field model as the the number of edges grows and the coupling between the edges grow stronger, and shows that for a certain parameterisation the limit is a zero mean cyclic stationary Gaussian process. For the second order model the parameterisation chosen by Grenander (1993) is

$$Q_{kl} = \begin{cases} \frac{\kappa}{m} + 6\eta m^3 & l = k \\ -4\eta m^3 & l = k-1, k+1 \bmod m \\ \eta m^3 & l = k-2, k+2 \bmod m \end{cases} \tag{5.11}$$

where κ, η are positive parameters. For simplicity we use the notation $Q = \mathrm{circl}(\kappa/n + 6\eta m^3, -4\eta m^3, \eta m^3)$. Hobolth and Jensen (2000) derives the variance

τ^2 and correlation $\rho(h)$ function of the limiting process; these are given by

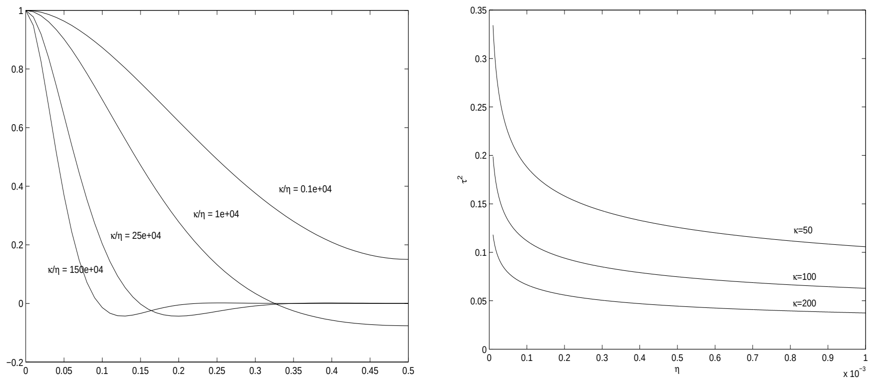
$$\tau^2 = \frac{\psi}{2\kappa} \frac{\psi_1\psi_2 + \psi_3\psi_4}{(\psi_2\psi_3)^2 + (\psi_1\psi_4)^2}. \tag{5.12}$$

$$\rho(h) = \frac{(\psi_2\psi_3 + \psi_1\psi_4)\,\psi_1(h)\psi_3(h) + (\psi_2\psi_3 - \psi_1\psi_4)\,\psi_2(h)\psi_4(h)}{\psi_1\psi_2 + \psi_3\psi_4}, \tag{5.13}$$

for $0 \leq h \leq 1$, and with the notation

$$\begin{aligned}
\psi_1(h) &= \cos\psi(h - 1/2), & \psi_2(h) &= \sin\psi(h - 1/2) \\
\psi_3(h) &= \cosh\psi(h - 1/2) & \psi_4(h) &= \sinh\psi(h - 1/2) \\
4\psi^4 &= \frac{\kappa}{\eta} & \psi_i &= \psi_i(1), \qquad i = 1, \ldots, 4.
\end{aligned}$$

Note that the correlation function only depends on the ratio κ/η, whereas the variance decreases for increasing κ. Figure 5.1 shows plot of the correlation and variance for various values of κ/η and κ.



(a) (b)

Figure 5.1: (a) Limiting correlation function for the second order Markov random field model in Eq. (5.11) for different values of κ/η. (b) Limiting variance for the same model and for different values of κ, plotted as a function of η.

To get a further feel for the parameters we have drawn samples from the model, see Figure 5.2. As discussed above they indicate that κ/η measures the smoothness, while κ measure deviances from the original configuration.

Concerning the choice of parameters, Kent et al. (2000) and Hobolth and Jensen (2000) considers maximum likelihood estimation of the deformation parameters based on training samples of data. Hurn, Steinsland and Rue (2001) goes a step further and assumes that the training data are generated as noisy

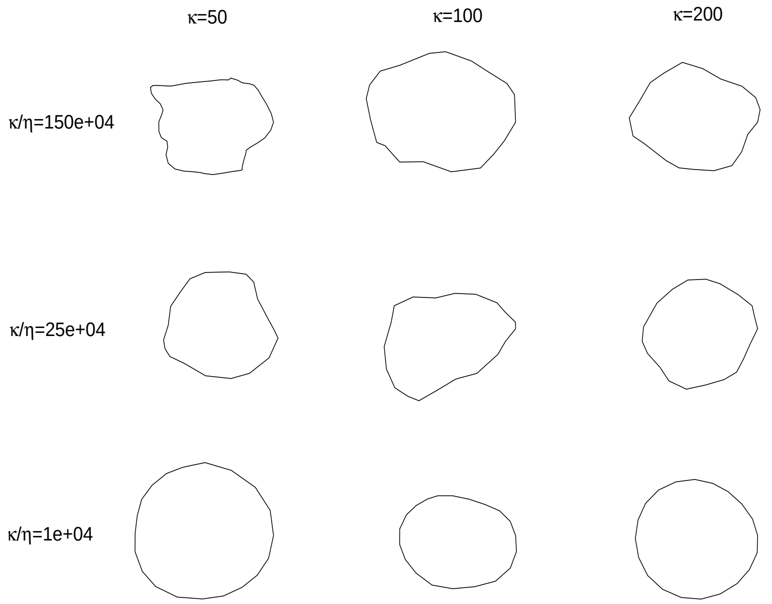


Figure 5.2: Samples from the edge transformation template model with precision matrix (5.11), and with different values of the parameters κ and η.

observations of deformable templates. However, in many cases the best approach is to integrate out the effect of the parameters by doing a fully Bayesian analysis as in Hurn (1998) and Hansen et al. (2000). We will follow this line and specify hyperpriors for the parameters of the template model. The analysis above gives hint on how to specify the hyperpriors for the parameters, but we have carried out a preliminary experiment to investigate the range of the parameters, and see if there is a need for different parameters for the normal and tangential components of the edge transformation vector.

We have used a template with $m = 20$ edges to locate the object in a simple image containing one object. We used a simple likelihood model assuming independent, additive Gaussian noise, see Section 8.2.4 below for details, and let the parameters κ and η of the precision matrix Q (Eq. (5.11)) be uniformly distributed on $[0, 400]$ and $[0, 10]$, respectively. We tried both using different parameters for the tangential and normal components, and having common parameters. The results are summarised below in Table 5.1, and show that η is fairly concentrated on values of the order 10^{-3}, while κ is spread out over a large range, but with mean approximately 100. For the model with four parameters there seems to be a significant difference between the parameters of the normal and tangential components. However, visually there was little difference between the samples from the models with two and four parameters, and we will in the following use the simplest model with just two parameters.

Table 5.1: Posterior mean estimates of the model parameters κ and η in Eq. (5.11), using uniform prior distributions. In the left column we have used different distributions for the normal and tangential components of the edge transformation model; in the right column the distribution is the same.

	Mean	Sdev.		Mean	Sdev.
κ_{n}	214.5	131.2	κ	92.7	61.1
κ_{t}	87.0	86.1			
η_{n}	.00956	.0126	η	.00385	.00814
η_{t}	.00313	.0056			

5.4 Continuous Models

In many situations the outline-based models are too simplistic to be able to capture the details and internal structure of the modelled objects. This is for instance the case in many biomedical applications, where one need models that can capture the detailed structure of the human anatomy. Objects are therefore more easily represented as images (Bajcsy and Kovacic, 1989; Amit, Grenander and Piccioni, 1991; Grenander and Miller, 1994; Christensen, Rabbitt and Miller, 1996; Thompson and Toga, 1996; Gee, 1999), or as non-parametric curves and surfaces (Bookstein and Green, 1992; Joshi, Miller, Christensen, Banerjee, Coogan and Grenander, 1995; Davatzikos, 1996; Younes, 1998; Bakircioglu, Grenander, Khaneja and Miller, 1998), sometimes with additional information in the form of anatomical landmarks (Bookstein, 1989; Mardia, Kent, Goodall and Little, 1996); see e.g. Grenander and Miller (1998) for further references. Again, one does not model the objects themselves, but rather the transformations acting on a template object. For instance, if the object is an image, transformations could be taken from the collection of smooth coordinate transforms of the image domain. Important modelling issues would then be the choice of boundary conditions, and the choice of prior on the set of transformations. Note that image matching is another application of the same methodology, see e.g. Glasbey and Mardia (1998) and, more recently, Glasbey and Mardia (2001) for a review, and Sampson and Guttorp (1992) and Perrin and Senoussi (1999) for matching in different contexts.

One of the fist works in a statistical setting was Amit et al. (1991), who consider reconstruction of X-ray images of hands. Images are defined as mappings $I : \Omega \to \mathcal{T}$ from some fixed background space Ω to a range space $\mathcal{T}$. The space of mappings is denoted $\mathcal{I}$, and a particular image $I_0 \in \mathcal{I}$ is chosen as the template. Variability in image appearance is modelled thorough the group $\mathcal{H}$ of diffeomorphic transformations $h : \Omega \ni x \mapsto h(x) = x - u(x) \in \Omega$. The mappings should be $1-1$, onto and differentiable with a differentiable inverse so as to preserve certain topological properties of the images, e.g. that connected sets remain connected and critical points are mapped to critical points. Sometimes it is natural to let the template be the orbit $I_{\text{temp}} = \mathbb{E}(3) I_0$ under the special Euclidean group $\mathbb{E}(3) = \mathbb{R}^3 \ltimes \mathrm{SO}(3)$. Then the maps in $\mathcal{H}$ are defined modulo the rigid motions.

Amit et al. (1991) assume the displacement field $u(x)$ to be Gaussian distributed with covariance kernel equal to the Greens function of a squared differential operator

$$Lu = \alpha \Delta u, \tag{5.14}$$

where Δ is the Laplacian and α some constant. Boundary conditions for Eq. (5.14) is chosen so that the boundary is mapped onto itself. To be able to do inference the infinite-dimensional random variable u is approximated by the truncated orthonormal expansion

$$u_h(x) = \sum_{k=0}^{h} \xi_k^T \Phi_k(x) / \lambda_k, \tag{5.15}$$

where λ_k are the eigenvalues and Φ_k the matrix of eigenvectors for the (discretized) linear operator L, and the coefficients ξ_k are i.i.d. $\mathrm{N}(0, I_2)$. Amit and Piccioni (1991) show that the distribution of u_h converges weakly to the distribution of u on the space of continuous functions on Ω. Assuming the target image to be observed with the addition of independent and identically distributed

Gaussian noise, the posterior distribution for the coefficients becomes

$$\pi(\xi) \propto \exp\left\{-\frac{1}{2}\sum_{k=0}^{h}|\xi_k|^2 - \frac{1}{2\sigma^2}\sum_{l}\left(I(x_l) - I_0(x_l + u_h(x_l))\right)^2\right\}, \tag{5.16}$$

where the x_l are the nodes of a grid Ω_g. Note that Amit et al. (1991) use a Fourier basis in Eq. (5.15), thus getting a posterior where the terms all have global support. To be able to do local updates in areas with high variability, Amit (1994) uses a wavelet basis in the expansion (5.15).

In the above framework Amit et al. (1991) implicitly assume that the deformations are small so that u can be approximately assumed to be in a Hilbert space $\mathcal{U}$ with norm $||u||^2_{\mathcal{U}} = ||Lu||^2$. In reality the group $\mathcal{H}$ of diffeomorphism is a curved manifold, and assuming a vector space structure leads to inconsistencies. In particular, as pointed out by Dupuis, Grenander and Miller (1998), the quadratic penalties $||Lu||^2$ in amongst others Amit et al. (1991), Bajcsy and Kovacic (1989) and Miller, Christensen, Amit and Grenander (1993) implies restoring forces that are proportional to the displaced distance. Thus large deformations are severely penalised. Whether this is undesirable or not depends on the application, but the lack of symmetry – since the inverse h^{-1} is not necessarily a diffeomorphism – means that the minimiser of Eq. (5.16) cannot be considered a distance on $\mathcal{I}$. In situations where no obvious choice of template and images need to be interchangeable, symmetry is natural.

To construct large-deformation maps Christensen et al. (1996) and Dupuis et al. (1998) let $h(x)$ be the output $h(x,1)$ of a flow $h(x,t)$ generated by a velocity field $v(x,t)$:

$$\frac{\partial h(x,t)}{\partial t} = -\Big(\left(\nabla_x^T h\right)(x,t)\Big)v(x,t), \quad h(x,0) = x. \tag{5.17}$$

In this case the deformation stress is not accumulated, hence the term viscous deformations (Christensen et al., 1996). A similar approach can be found in Trouvé (1998) and Piccioni, Scarlatti and Trouvé (1998). Since v is tangent to $\mathcal{H}$ we can assume it to be in a Hilbert space $\mathcal{V}$ with norm $||\cdot||_{\mathcal{V}}$. As in Amit et al. (1991) the norm is induced by a linear differential operator L. In most image matching applications the operator is of the form $L = a\nabla^2 + b\nabla\nabla\cdot + c$; chosen so as the solution obeys certain laws from continuum mechanics for deformable bodies, see Christensen (1994), Christensen et al. (1996), and Grenander and Miller (1998). The posterior energy for v becomes

$$\frac{1}{2}\int_0^1\int_\Omega |Lv(x,t)|^2 \mathrm{d}x\mathrm{d}t + \frac{1}{2\sigma^2}\int_\Omega |I_0 \circ h(x,1) - I(x)|^2\,\mathrm{d}x, \tag{5.18}$$

from which an estimate v^* can be found. The optimal match is then given by the integral equation

$$h^*(x,1) = x - \int_0^1 \Big(\left(\nabla_x^T h^* \right)(x,t) \Big) v^*(x,t) \mathrm{d}t. \tag{5.19}$$

These equations can be solved using expansions similar to Eq. (5.15); see Grenander and Miller (1998) for details. Note that taking the minimum of the prior energy subject to $I_0 \circ h(0) = I_0$ and $I_0 \circ h(1) = I$ defines a distance on $\mathcal{I}$, see e.g. (Younes, 1998; Younes, 1999; Trouvé, 1998). This enables us to compare images quantitatively, which is relevant for inference (see Chapter 6).

Sometimes is is natural to constrain the solution to match given landmarks. This can be treated in the same manner, replacing the likelihood in Eq. (5.18) with

$$\pi(y \mid h) \propto \exp\left\{ -\frac{1}{2} \sum_{l=1}^{L} \left(y_l - h(x_l,1)\right)^T \Sigma_l^{-1} \left(y_l - h(x_l,1)\right) \right\}, \tag{5.20}$$

where taking the limit $\Sigma \to 0$ results in exact matching (Joshi, 1997). In the small-deformation setting landmark matching has been done using thin-plate splines (Bookstein, 1989) and kriging (Kent and Mardia, 1994; Mardia et al., 1996). This is closely related to the shape theory (Kendall, 1984; Dryden and Mardia, 1998) which considers statistical analysis of n-tuples of points, including distributions and distances. Younes (1998) shows an interesting relation between distances in Kendall's shape space and geodesic distances defined on infinite dimensional transformation groups.

5.5 Landmark Models

In most applications an useful model needs to be specific: it must only represent certain legal shapes. At the same time it must allow for natural variability. One attractive way of building such models is by learning from representative training samples. In such a procedure, both mean shape and variability within the object class is learned. For instance, Cootes, Hill, Taylor and Haslam (1994) and Cootes, Taylor, Cooper and Graham (1995) computes a mean shape after aligning the training sets into some suitable coordinate system, and then use principal component analysis to extract the important modes of variation. Their models are called Point Distribution models; see below for further details.

In this approach the template is specified by a set of m landmarks collected along the outline. For two-dimensional applications one often uses complex arithmetic, and we denote the vector of landmarks by $v \in \mathbb{C}^m$. For this to be useful, each landmark must represent a distinguishable feature present in all samples of the shape; thus amorphous objects such as cells and (often) blood vessels do not fit well into this methodology. For such objects the polygonal models in Section 5.3.2 are most suitable.

Landmarks models usually consider the *shape* of such m-point configurations, ie. the geometrical information that remains when location, scale, and rotational effects are filtered out (Dryden and Mardia, 1998). There are two common ways of doing this (Kent, 1994). The first representation, called Kendall coordinates Kendall (1984), are constructed as follows. Let H be a $(m-1) \times m$ row orthonormal matrix with row sums equal to zero, $H1_m = 0$. Setting $z = Hx/\|Hx\|$ removes location and scale effects. Note that z lies on the unit sphere $\mathbb{C}S^{m-2}$. The final standardisation is achieved by representing z by the equivalence class $[z] = \{\exp(i\theta)z\,;\, \forall\,\theta \in [0, 2\pi)\}$. The second representation uses coordinates z obtained by translating, scaling, and rotating the landmarks such that the two first landmarks are placed at the points $(-1/2, 0)$ and $(1/2, 0)$. This representation was suggested by Bookstein (1986).

One way of proceeding is by specifying a probability distribution for the shape coordinates z. Dryden and Mardia (1991) derive the distribution for z given that the landmarks are circular complex Gaussian, $v \sim \mathbb{C}\mathrm{N}(\mu, \Sigma)$, but this distribution is very complicated. A more feasible approach is to use the complex Bingham distribution (Kent, 1994), which basically is constructed by conditioning a zero mean complex normal vector to have unit norm. Even simpler is the complex Watson distribution

$$\pi(z) \propto \exp\left(\xi |z^* \mu|\right),$$

which is obtained by assuming that the landmarks are independent and isotropic with equal variances (Mardia and Dryden, 1999). The parameter $\xi \geq 0$ is called the concentration parameter. Both distributions exhibit complex symmetry, and have thus some of the same limitations as the cyclical Gaussian Markov random field models in Section 5.3.2. We refer to Dryden and Mardia (1998) for a thorough discussion of shape analysis and shape distributions.

In stead of specifying a distribution, Cootes et al. (1994) use principal component analysis for modelling shape variation. They essentially use coordinates v in the tangent space of the unit sphere (Kent, 1994), and assume

$$v = \bar{v} + \Phi b, \tag{5.21}$$

where $\bar{v}$ is the mean of K training samples, and Φ contains the M eigenvectors corresponding to the M largest eigenvalues of the sample covariance matrix

$$S = \frac{1}{K-1} \sum_{k=1}^{K} \left(v_k - \bar{v}\right)^T \left(v_k - \bar{v}\right).$$

The vector b contains the K shape parameters that together with the 4 global parameters (corresponding to global translation, scaling, and rotation) control the deformable model. In general only $5-7$ modes are sufficient to account for most of the variations in a shape class (Cootes et al., 1994), and thus the method is very effective in reducing the model dimension. Since variation is divided into orthogonal components, the parameters are often easily interpretable.

This model is often called the Point Distribution model (PDM), and has proved to be very effective in modelling variable shapes, see eg. Glasbey (1998), Kervrann and Heitz (1999), and de Souza et al. (1999). One drawback is perhaps that the model deforms in a way that reflects the variation inherent in the training sample, and is thus not very robust with respect to "unexpected" deformations. Several non-linear extensions exists that are more flexible. Sozou, Cootes, Taylor and Mauro (1995) use polynomial regression, while Heap and Hogg (1996) use polar coordinates for sub-parts of the model. Finally, Cootes and Taylor (1999) use a mixture of Gaussian distributions to model the shape distribution.

5.6 Models for an Unknown Number of Objects

Deformable template models such as such as polygon models and point distribution models fit nicely into a framework for analysing images with an unknown number of objects. Baddeley and Van Lieshout (1993) present a marked point process model where each object x_k consist of a point specifying its location, and a mark containing information about shape and object type. The points lie in a window $\Omega \subset \mathbb{R}^2$, and the marks in some mark space $\mathcal{M}$ determined by the particular template model. An object configuration is a finite unordered set $x = \{x_1, \dots, x_m\}$ of objects $x_k \in \Omega \times \mathcal{M}$, and a random object process is an element of the set $\cup_{m=1}^{\infty} \{x_1, \dots, x_m\} \cup \emptyset$. Baddeley and Van Lieshout (1993) use a Poisson object process on $\Omega \times \mathcal{M}$ as their reference process, and define the density of a configuration $\{x_1, \dots, x_m\}$ as a pairwise interaction model

$$\pi(x, m) \propto \beta^m \prod_{x_k \sim x_l} h(x_k, x_l) \tag{5.22}$$

with respect to the Poisson object process. Here $\sim$ is some symmetric and reflexive relation. Baddeley and Van Lieshout (1993) discusses different interaction models, for instance the area interaction model

$$h(x_k, x_l) = \gamma^{|B(x_k,r) \cap B(x_l,r)|},$$

where $B(x, r)$ is the ball with radius r centred at x. A similar model for line processes is discussed in Stoica et al. (2000), while Rue and Syversveen (1998) and Rue and Hurn (1999) use hard-core object processes in applications where objects are not allowed to overlap.

Models of this type can be sampled using the reversible jump MCMC algorithm (Green, 1995), but convergence can be very slow unless great care is taken in designing the required split and merge moves, see Rue and Hurn (1999) for a discussion. Grenander and Miller (1994) uses a different approach, called jump-diffusion dynamics. They construct a single posterior distribution over the union of all configurations, and then sample from the distribution using a Markov process that at random times jumps between configurations of different size. In between jumps the process follows a Langevin diffusion on the relevant fixed-dimension subspace.

Chapter 6

Inference in High-Level Models

6.1 Introduction

In this section we briefly discuss inference in image analysis, with a special focus on problems in object recognition. In practical Bayesian inference, estimators for the quantity of interest, θ say, must be based on samples from the posterior distribution $\pi(\theta \mid y)$ for θ given the observed data y. Using a decision-theoretic approach, the estimators should be based on a loss function $L(\theta, \hat{\theta})$ representing the loss incurred by estimating θ with $\hat{\theta}$. The optimal Bayes estimator (OBE) is then selected as the minimiser of the posterior loss,

$$\theta^* = \arg\min_{\hat{\theta}} \mathrm{E}_{\theta \mid y} L(\theta, \hat{\theta}) = \arg\min_{\hat{\theta}} \int L(\theta, \hat{\theta}) \pi(\theta \mid y) \, d\theta. \tag{6.1}$$

Common choices of loss functions are $L_{\mathrm{MAP}} = \mathbb{I}_{[\theta \neq \hat{\theta}]}$ and $L_{\mathrm{PM}} = \sum_i (\theta_i - \hat{\theta}_i)^2$ with corresponding OBE's the posterior mode and mean, respectively. These estimators are easy to compute, but in general no explicit formulas exists, and the OBE must be based on a direct minimisation of the posterior loss (6.1).

For all their advantages, the above mentioned loss functions are too simplistic for many imaging problems since they are based on element-by-element errors, ignoring the global structure of the quantities. Consider for instance the case of object recognition: the posterior mean is obviously not well suited for estimating quantities such as object shape, location, and type. Not much work has been done on exploring other loss functions, but Rue and Syversveen (1998) and Rue and Hurn (1999) has successfully applied loss functions based on Baddeley's delta metric (Baddeley, 1992) to object recognition problems, and have demonstrated that they in many respects perform better than standard loss functions. We review this approach in Section 6.2. Baddeley's delta metric compares images of objects, giving a natural embedding of the object in the coordinate system of the image. In other situations it might be natural to consider image and object separately, and we briefly review some approaches to defining distances between grey-level images (Section 6.4) and planar curves (Section 6.3).

6.2 Baddeley's Delta Metric in Object Recognition

Baddeley's delta metric was introduced in Baddeley (1992) (see also Rue (1999)), and has been used for image classification (Frigessi and Rue, 1997) and object recognition (Rue and Syversveen, 1998; Rue and Hurn, 1999). The material in this section is based on these references.

The delta metric was introduced for comparing discrete images with a small number of colours, and was motivated by the need for a measure of discrepancy (between two images) that agrees with the human visual system. Let x_1 and x_2 be two binary images defined on grid $S \subset \mathbb{Z}^2$, and let X_1 and X_2 be the corresponding foreground sites, $X_k = \{s \in S \,:\, x_{k,s} = 1\}$. Let ρ be a metric on S. The distance from a site $s \in S$ to a set $T \subset S$ is defined as

$$d(s,T) = \min_{t \in T} \rho\,(s,t)\,. \tag{6.2}$$

Let w be a strictly increasing concave function defined on $[0,\infty]$ and satisfying $w(0) = 0$. Then the delta metric between x_1 and x_2 is

$$\Delta_w^p(x_1,x_2) = \left(\frac{1}{|S|}\sum_{s\in S}\left|w\big(d\,(s,X_1)\big) - w\,\big(d\,(s,X_2)\big)\right|^p\right)^{1/p}, \quad 1 \le p < \infty. \tag{6.3}$$

We see that the metric is the L^p average of the differences between the distances $w(d(\cdot,\cdot))$ from each pixel to the nearest foreground pixel. One usually assumes $w(t) = \min(t,c)$ so that the contribution from pixel i will only depend on the sites in x_1 and x_2 contained in the ball $B(s,c)$. One then uses the notation Δ_c^p.

The delta metric has two other metrics as its limits. If we use $w(t) = t$ and let $p \to \infty$, we obtain the Hausdorff metric $\max_{s\in S}|d(s,X_1) - d(s,X_2)|$. Secondly, if c is smaller than the smallest distance between two distinct sites, $(\Delta_c^2/c)^2$ is the average marginal misclassification,

$$\left(\Delta_c^2(x_1,x_2)/c\right)^2 = \frac{1}{|S|}\sum_{s\in S}\mathbb{I}_{[x_{1,s}\neq x_{2,s}]}.$$

Moreover, we note that the metric is more sensitive to add-on components than to errors in the interior of an object, see Rue and Syversveen (1998) for an example. We refer to Baddeley (1992) for further properties and examples.

Rue and Syversveen (1998) and Rue and Hurn (1999) applies the metric for recognising and identifying an unknown number of objects of different type.

Recall from Section 5.6 that an object configuration x is defined as an unordered collection $\{x_1, \ldots, x_n\}$ of objects $x_i \in \Omega \times \mathcal{M}$, where $\Omega \subset \mathbb{R}^2$ and $\mathcal{M}$ is the mark space. We assume that Ω is bounded and covers the lattice S on which the data is observed. For each object x_k, let $X_k \subset \mathbb{R}^2$ be the set enclosed by its contour, and let $X = \cup_k X_k$. Furthermore, let ρ be the Euclidean metric on $\mathbb{R}^2$, and let $d(\xi, x) = \inf\{\rho(\xi, \eta) : \eta \in X\}$. Then Rue and Syversveen (1998) use the squared delta metric as a loss function. This loss function implicitly gives the loss for the number of objects and their outline, but not for other object properties such as grey-level or type. However, Rue and Hurn (1999) notes that because of the variable dimension of the model, it is not possible to simply add terms that specify the loss if we estimate one object type by another. Instead they associate the loss with the underlying space, adding a loss if point ξ is covered by wrong object, ie.

$$L_{\text{type}}(x, x') = \int_\Omega \mathbb{I}_{\text{type}}(\xi, x, x')\, d\xi,$$

where $\mathbb{I}_{\text{type}}(\xi, x, x')$ takes the value one if location ξ is covered by an object of wrong type, and zero otherwise. The total loss function is then

$$L(x, x') = \int_\Omega \left(w\left(d(\xi, x)\right) - w\left(d(\xi, x')\right)\right)^2 d\xi + \gamma \int_\Omega \mathbb{I}_{\text{type}}(\xi, x, x')\, d\xi,$$

for some $\gamma > 0$.

The optimal estimator in Eq. (6.1) cannot be found explicitly, but Rue (1995) has described a MCMC algorithm for computing the estimator of a wide class of loss functions. He notes that the expected loss can be written as (up to an additive constant)

$$\begin{aligned} \mathrm{E}_{x \mid y} L(x, x') = \int_\Omega w\left(d(\xi, x')\right) \Big(w\left(d(\xi, x')\right) - 2\mathrm{E}_{x \mid y} w\left(d(\xi, x')\right)\Big)\, d\xi \\ + \gamma \int_\Omega \mathrm{Pr}_{x \mid y} \mathbb{I}_{\text{type}}(\xi, x, x')\, d\xi. \quad (6.4) \end{aligned}$$

The algorithm then proceeds as follows. Estimate the posterior expectation $\mathrm{E}_{x \mid y} w\left(d(\xi, x')\right)$ by using the empirical average of samples form the posterior distribution $\pi(x \mid y)$. The resulting estimate $\widehat{\mathrm{E}_{x \mid y} L(x, x')}$ is then a function of x' only, having the optimal estimator as its minimiser. This minimiser is found by interpreting

$$\pi_\Delta(x') \propto \exp(-\widehat{\mathrm{E}_{x \mid y} L(x, x')})$$

as a density and using simulated annealing (Robert and Casella, 1999). Rue and Syversveen (1998) give details on implementation, and reports good performance even with fast cooling schemes.

In conclusion, Baddeley's delta metric has proved to be well suited as a loss function for object recognition. The corresponding estimators shows improved performance compared to traditional estimators (Rue and Syversveen, 1998), and the additional computational effort is moderate.

6.3 Metrics for Closed Curves

In this section we discuss a few other approaches to estimating shapes of objects. The framework in the last section is very powerful in the sense that it makes it possible to the estimate the number of objects in an image, as well as their shape, location and type. However, in some situations simpler approaches will do, and in this section we consider the following problem. Suppose we have n samples $x_1, \ldots, x_n$ of closed planar curves from a posterior distribution $\pi(x \mid y)$. We want to produce a single curve representing the "average" of the sampled curves. Moreover we might want to compare this average curve or the individual curves to a known "true" curve. Note that these two tasks are related, for if we have define a distance function d on the set of closed curves, the average curve could be obtained as

$$\hat{\mu} = \arg\inf_{\mu} \sum_{i=1}^{n} d^2\left(x_i, \mu\right).$$

We shall review two approaches for solving the first task. In Section 6.3.1 we review how Grenander's "least action principle" (Grenander, 1993, Chap. 12) can be used for defining distances between curves. Secondly, in Section 6.3.2 we review another approach based on the theory for random closed sets (Baddeley and Molchanov, 1998).

6.3.1 Metrics for closed curves

We start with a comment on representation. In many applications there is no natural intrinsic representation of the outlines. Thus, even if an outline in practice must be represented by a finite number of vertices, constructing a distance by matching vertices is only natural when the vertices corresponds to identifiable points on the outline, eg. some anatomical feature or extremal point. If this is the case, one often uses Procrustes methods, in which distances are based on matching the vertices using the Euclidean similarity transform. For instance, if $x_1, x_2 \in \mathbb{C}^n$ are two planar point configurations, the full Procrustes distance between the configurations is given by

$$\begin{aligned} d_F(x_1, x_2) &= \inf_{\theta \in [0, 2\pi), \beta \in \mathbb{R}, \alpha \in \mathbb{C}} \left\| \frac{x_1}{\|x_1\|} - \frac{x_2}{\|x_2\|} \beta e^{i\theta} - a \right\| \\ &= \left(1 - \frac{x_1^* x_2 x_2^* x_1}{x_1^* x_1 x_2^* x_2}\right)^{1/2}, \end{aligned} \tag{6.5}$$

where $x^* = \bar{x}^T$. We shall not discuss such methods further, and refer to Dryden and Mardia (1998) for a comprehensive treatment of Procrustes methods and shape analysis.

Now we briefly review how one can define distances in terms of the least action principle, and thereafter discuss a particular distance derived for plane curves (Younes, 1998). The following discussion is sketchy; for more details, please refer to eg. Grenander (1993, Chap. 12) and Younes (1998). Assume that we are given a set $\mathcal{C}$ of objects (eg. curves), and a group G of deformations acting transitively on $\mathcal{C}$. That is, for all $c_1, c_2 \in \mathcal{C}$ there exists an $g \in G$ so that $gc_1 = c_2$. The idea of the least action principle is to let the distance on $\mathcal{C}$ by induced by a distance on G. Loosely speaking, for $c_1, c_2 \in \mathcal{C}$, we find the closest g_1, g_2 such that $g_1^{-1}c_1 = g_2^{-1}c_2$ and use the distance between g_1 and g_2. By using such a construction one hopes to fulfil a very natural requirement: If $g \in G$ is in some sense a "small" deformation, then the distance between c and gc should also be small.

To be more specific, we start by defining a distance.

Definition 3
A distance over a set E is a mapping $d : E \times E \to [0, \infty]$ such that for all $x_1, x_2, x_3 \in E$

$$\begin{array}{ll} (i) & d(x_1, x_2) = 0 \Leftrightarrow x_1 = x_2 \\ (ii) & d(x_1, x_2) = d(x_2, x_1) \\ (iii) & d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3). \end{array}$$

If (i) is replaced by $d(x, x) = 0, \forall x$, we call d a pseudo-distance.

Let d_G be a distance on G. We will require this distance to be right-invariant with respect to the action of G on itself, that is, for all $g_1, g_2, h \in G$, $d_G(g_1h, g_2h)$ $= d_G(g_1, g_2)$. Then we have the following result, which is a simplification of a result in Grenander (1993, Chap. 12),

Proposition 5
Let d_G be a distance on G that is invariant by the right action of G. The function $d : \mathcal{C} \times \mathcal{C} \to [0, \infty]$ defined by

$$d(c_1, c_2) = \inf \left\{ d_G(g_1, g_2) \,:\, g_1^{-1}c_1 = g_2^{-1}c_2 \right\} \tag{6.6}$$

is a pseudo-distance on $\mathcal{C}$.

Proof. Conditions (i) and (ii) are obvious. To prove the triangle inequality, note that as a consequence of the right-invariance and the triangle inequality for d_G, we have for all $g_1, g_2, g_3, g_4 \in G$:

$$\begin{aligned} d_G(g_1, g_2) + d_G(g_3, g_4) &= d_G(g_1, g_2) + d(g_2, g_4 g_3^{-1} g_2) \\ &\geq d_G(h_1, h_3) \end{aligned}$$

with $h_1 = g_1$ and $h_3 = g_4 g_3^{-1} g_2$. Now fix $c_1, c_2, c_3 \in \mathcal{C}$ and choose g_1, g_2, g_3, g_4 such that $g_1^{-1} c_1 = g_2^{-1} c_2$ and $g_3^{-1} c_2 = g_4^{-1} c_3$. Then

$$h_3^{-1} c_3 = g_2^{-1} g_3 g_4^{-1} c_3 = g_2^{-1} g_3 g_3^{-1} c_2 = h_1^{-1} c_1.$$

Thus

$$\begin{aligned}
&d(c_1, c_2) + d(c_2, c_3) \\
&\qquad = \inf\left\{d_G(g_1, g_2) \,:\, g_1^{-1} c_1 = g_2^{-1} c_2\right\} + \inf\left\{d_G(g_3, g_4) \,:\, g_3^{-1} c_2 = g_4^{-1} c_3\right\} \\
&\qquad \geq \inf\left\{d(h_1, h_3) \,:\, h_1^{-1} c_1 = h_3^{-1} c_3\right\} \\
&\qquad = d(c_1, c_3).
\end{aligned}$$

□

The problem is now to compute a suitable distance on G. One usually assumes that G is a Lie group so that one can construct a metric in the standard way: If $g : [0,1] \to G$ is a path on G we can define the speed $g'(t)$ as the tangent vector to G at $g(t)$. Choosing an inner product $\langle \cdot, \cdot \rangle_g$ on the tangent space, the length of the path g is given by

$$E(g) = \int_0^1 \|g'(t)\|_{g(t)}^2 \, dt,$$

and the corresponding Riemannian distance between two elements g_1 and g_2 is defined by

$$d_G(g_1, g_2) = \inf\left\{\sqrt{E(g)},\, g(0) = g_1,\, g(1) = g_2\right\}, \tag{6.7}$$

ie. the length of the shortest path connecting g_1 and g_2. We refer to Younes (1998) for details on how to determine a right-invariant inner product on a general group G, and instead give a simple finite dimensional example.

EXAMPLE 6.3.1 We consider again two n-dimensional point configurations x_1,x_2 $\in \mathbb{R}^{2n}$, being for instance the vertices of two polygons. We assume the point sets to be scaled to have unit size and centroid in origin. The polygons are deformed using global rotations in the plane, whence $G = \mathrm{SO}(2)$ with elements

$$g = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in [0, 2\pi).$$

We can define a path from the identity I_2 to g, say, by using the matrix exponential

$$g(t) = \exp(tX) = \exp\left(t \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}\right),$$

and this path can in fact be shown to be the shortest path (see eg. Boothby

(1986, chap. 9). Furthermore it is well known (see eg. example 8.6, p. 353 Boothby (1986)) that $\langle X_1, X_2\rangle = \mathrm{trace}(X_1 X_2^T)$ defines a bi-invariant metric on SO(2), and thus

$$\begin{aligned} d_G^2(g_1,g_2) &= \int_0^1 \|g'(t)\|^2\, dt = \int_0^1 \mathrm{trace}\left(g'(t)g'(t)^T\right) dt \\ &= \mathrm{trace}\left(XX^T\right), \end{aligned}$$

where X satisfies $g_2 = g_1 \exp X$. Consequently $d_G(g_1,g_2) = \sqrt{2}\theta$ for some $\theta \in (-\pi,\pi]$. Since x_1 and x_2 are both of unit size, the minimum rotation angle between them must be given by $\arccos(x_1^T x_2)$, so by using Eq. (6.6) we get

$$d(x_1,x_2) = \sqrt{2}\arccos\left(x_1^T x_2\right).$$

While this framework reduces to simple calculations in the finite-dimensional case, for infinite-dimensional groups the calculations often becomes intractable. However, by considering a certain deformation group, Younes (1998) derives a closed form expression for the distance between continuous plane curves.

Younes (1998) represents a curve $C = \{x(t), y(t), t \in [0,1]\}$ through the function $e^{i\theta(t)} = \zeta : [0,1] \to \mathbb{C}S^1$ satisfying $x'(t) = \Re\zeta$ and $y'(t) = \Im\zeta$. Deformation of the curve is defined through the (infinite dimensional) group action

$$(\phi, r).(l, \zeta) = (r\zeta \circ \phi), \tag{6.8}$$

where ϕ is a diffeomorphism on $[0,1]$, and r is a function $r : [0,1] \to \mathbb{C}S^1$. One can think of $r(t)$ as a torsion of C at t, and $\phi'(t)$ as a tangential stretching. Using the above framework he derives the following translation and scale invariant distance between two plane curves C_1 and C_2:

$$d_G(C_1,C_2) = 2\inf_{\phi} \arccos \int_0^1 \sqrt{\phi'(t)} \left|\cos\frac{\theta_2(t)\circ\phi(t) - \theta_1(t)}{2}\right| dt, \tag{6.9}$$

where $g = (\phi, r)$. Note that in this derivation it is assumed that the starting point of the curves are known. For open curves there are only two choices, but for a closed curve the starting point may be any point in the curve. The distance (6.9) should therefore be minimised over all possible starting points. This complicates the computations, see Younes (1998) for details on this and how to implement the distance using polygonal approximations.

6.3.2 Averaging closed curves using the distance transform

In this section we suggest a simpler method for obtaining the average of a collection of closed curves. The method is appealing in the sense that it is independent of the particular representation of the curves, so eg. templates of different resolution could be used in computing the average. This invariance is obtained by viewing the curve as the boundary of a connected random closed set, and then using theory for random closed sets for computing the expectation. We start by reviewing some facts about random closed sets, and then describe how Baddeley and Molchanov (1998) construct expectations of random closed sets using representations of sets by distance functions. Finally, we show how this could be straightforwardly extended to obtaining expectations of random closed curves.

Preliminaries We start with some definitions related to random closed sets and distances.

DEFINITION 4
A Random closed set is random element in the space $\mathcal{F}$ of all closed subsets of a metric space $(\mathbb{E}, \rho)$, measurable in the sense that $\{X \cap K \neq \emptyset\}$ is a random event for all compact K.

The distance from a point to a set $F \subset \mathbb{E}$ can be defined using a distance function:

DEFINITION 5
Let $\mathcal{F}' = \mathcal{F} \setminus \emptyset$ be the space of all nonempty closed sets. A function $d : \mathbb{E} \times \mathcal{F}' \to \mathbb{R}$ is said to be a distance function if it is lower semi-continuous with respect to its first argument, measurable with respect to the second, and satisfies the following conditions

(D1) If $F_1 \subset F_2$, then $d(x, F_1) \geq d(x, F_2)$ for all $x \in \mathbb{E}$.

(D2) $F = \{x : d(x, F) \leq 0\}$.

In addition, it is often desirable to impose the following conditions.

(D3) $d(x, F) \leq d(x, \{y\}) + d(y, F)$ for all $x, y \in \mathbb{E}$ and $F \in \mathcal{F}$.

(D4) $d(x, F) = d(x, \{y\})$ for some $y \in F$.

(D5) $d(x, \{y\}) = d(y, \{x\}, \forall x, y \in \mathbb{E}$.

(D6) $d(x, F_1 \cup F_2) = \min\left(d(x, F_1), d(x, F_2)\right)$ for all $x \in \mathbb{E}$ and $F_1, F_2 \in \mathcal{F}'$.

One simple example is the metric distance function $d(x, F)$ that we have previously used. We recall that it is equal to the shortest distance between $x \in \mathbb{E}$ and $F \in \mathcal{F}$ in the metric ρ,

$$d(x, F) = \rho(x, F) = \inf\left\{\rho(x, y) : y \in F\right\}, \quad x \in \mathbb{E}. \tag{6.10}$$

In $\mathbb{R}^n$ with it usual metric this is the Euclidean distance. In the following we will use the related signed distance function

$$d(x, F) = \begin{cases} \rho(x, F), & x \notin F, \\ -\rho(x, F^C), & x \in F. \end{cases}. \tag{6.11}$$

Note that both the metric and the signed distance functions satisfy the conditions D3-D6.

Baddeley and Molchanov (1998) use the distance function d to define distances between sets. The embedding $F \mapsto d(\cdot, F)$ takes the family $\mathcal{F}'$ of non-empty closed sets into the space $\mathbb{F} = \{d(\cdot, F) : F \in \mathcal{F}'\}$ of distance functions. This space is equipped with a pseudo-distance d_F which can be used for defining distances between sets. For instance, when d is the metric distance function, the Hausdorff distance between compact sets K and L is

$$\begin{aligned} \rho_H(K, L) &= \sup_{x \in \mathbb{E}} |d(x, K) - d(x, L)| \\ &= d_F\left(d(\cdot, K), d(\cdot, L)\right), \end{aligned} \tag{6.12}$$

where d_F is the uniform metric. Alternatively, one can define distances using L^p-metrics and obtain the delta metric described above (Section 6.2) and in Baddeley (1992).

Distance averages of random closed sets The main idea in Baddeley and Molchanov (1998) is to employ the above embedding and work with the distance function $d(\cdot, X)$ instead of the random set X. This is convenient because it is possible to find the mean value of $d(x, X), x \in \mathbb{E}$. Moreover, if $d(x, X)$ is integrable for all $x \in \mathbb{E}$, we can define the mean distance function

$$\bar{d}(x) = \mathrm{E}d(x, X). \tag{6.13}$$

Note that this mean may not itself be a distance function. Baddeley and Molchanov (1998) now proceeds by defining the mean of the random closed set X as a certain level set of the mean distance function: Fix a closed set W and let d_{FW} be the restriction of d_F to W. Furthermore define an increasing family of sets

$$X(\varepsilon) = \left\{x \in W : \bar{d}(x) \leq \varepsilon\right\}, \varepsilon \in \mathbb{R}. \tag{6.14}$$

It is easy to show that $X(\varepsilon)$ is a closed set.

Proposition 6 (Baddeley and Molchanov (1998))
$X(\varepsilon)$ is a closed set for all $\varepsilon \in \mathbb{R}$.

Proof. Since $d(\cdot, X)$ is lower semi-continuous we get from Fatou's lemma:

$$\begin{aligned} \bar{d}(x) = \mathrm{E}d(x, X) &\leq \mathrm{E} \liminf_{y \to x} d(y, X) \\ &\leq \liminf_{y \to x} \mathrm{E}d(y, X) = \liminf_{y \to x} \bar{d}(y). \end{aligned}$$

Hence $\bar{d}(\cdot)$ is lower semi-continuous and $X(\varepsilon)$ is closed. □

Then Baddeley and Molchanov (1998) use a suitable distance d_F in the function space $\mathbb{F}$ to find the optimal threshold:

Definition 6 (Baddeley and Molchanov (1998))
Let $\bar{\varepsilon}$ be the minimum point of the d_F-distance between the distance function of $X(\varepsilon)$ and the mean distance function of X,

$$\overline{\varepsilon} = \arg\inf_{\varepsilon} d_F \left(d(x, X(\varepsilon)), \bar{d}(x) \right). \tag{6.15}$$

Then the set

$$\bar{X} = X(\bar{\varepsilon}) \tag{6.16}$$

is called the distance average of X.

The distance average computed using the metric distance function has been successfully applied for averaging images (Baddeley and Molchanov, 1998) and feature maps (Lewis, Owens and Baddeley, 1999). The method seems to be well adept in preserving the geometrical and topological properties of the input objects, and insensitive to reasonably small misregistrations. In that respect Lewis et al. (1999) provides the following result

Proposition 7 (Lewis et al. (1999))
Let x_1 and x_2 be two binary images on the same lattice S, and assume there exists a mapping $\omega : S \to S$ such that $x_2 = x_1 \circ \omega$. Let

$$h = \max_{i \in S} d(i, \omega(i))$$

be the maximum distance over which any pixel is displaced by ω. Then the distance transforms of x_1 and x_2 differ by at most h:

$$\max_{i \in S} |d(i, X_i) - d(i, X_2)| \leq h.$$

Still, the objects must not undergo too large rotations and translations for the distance average to be meaningful, but this is of course a concern for all averaging techniques.

In practice we compute the empirical distance average using n independent samples $X_1, \dots, X_n$ form the random set X:

$$d_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} d(x, X_i),$$

and then proceeds as above. Baddeley and Molchanov (1998) shows that the empirical distance average is a consistent estimator. Concerning implementation, the distance transform of a binary image can be computed in linear time using a discrete approximation of the Euclidean distance. We have used the chamfer(5,7,11) transform (Borgefors, 1986) in our experiments.

Distance averages of random closed curves The above framework can obviously be extended to compute the distance average of closed curves in $\mathbb{R}^2$. A closed curve drawn from some distribution can be seen as the boundary ∂X of a connected random closed set X. Note that ∂X is also a random closed set, but we will from now on identify a curve with its interior set X.

We will use the signed distance function, since it uniquely identifies the boundary of a set X as the subset whose distance to X is zero, $\partial X = \{x \in X : d(x, X) = 0\}$. For a thorough discussion on the use of the signed distance function in shape analysis, see Delfour and Zolesio (1994).

The following examples illustrates some interesting facts about the signed distance average.

Example 6.3.2 (Baddeley and Molchanov (1998)) Let ξ be a random variable on $[0, R] \subset \mathbb{R}^+$, and let $X_J = B(0, \xi) \subset \mathbb{R}^2$ be the closed ball with radius ξ and centre at the origin. Then the mean signed distance function is

$$\bar{d}(x) = \mathrm{E}\left(\|x\| - \xi\right) = \|x\| - \mathrm{E}\xi,$$

which is exactly the signed distance function for the the ball $B(0, \mathrm{E}\xi)$. Now by Eq. (6.14) and (D1),

$$X(0) = \{x : d(x, B(0, \mathrm{E}\xi)) \le 0\} = B(0, \mathrm{E}\xi),$$

and thus $\bar{X} = B(0, \mathrm{E}\xi)$.

Note that this is not true if we use the metric distance function. In fact, if one uses a non-negative distance function, the mean distance function is in itself a distance function if and only if X is deterministic.

Example 6.3.3 Let $X = B(\xi, r) \subset \mathbb{R}$, where ξ is a random variable uniformly distributed on $[0, 1]$. Using the signed distance function we get that $d(x, X) = \|x - \xi\| - r$ and

$$\bar{d}(x) = \mathrm{E}\, d(x, X) = \int_0^1 |x - \xi|\, d\xi - r = x^2 - x + \frac{1}{2} - r,$$

whence $X(\varepsilon) = \left[\frac{1}{2} - \frac{1}{2}\sqrt{1 + 4(r + \varepsilon)}, \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4(r + \varepsilon)}\right]$. We let d_F be the Hausdorff metric, so that

$$\begin{aligned}\bar{\varepsilon} = \arg\inf_{\varepsilon} \max_{x} \Bigg(& \sup_{x \le 1/2} \left| x^2 - x + \frac{1}{2} - r + x - \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4(r + \varepsilon)} \right|, \\ & \sup_{x > \frac{1}{2}} \left| x^2 - x + \frac{1}{2} - r - x + \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4(r + \varepsilon)} \right| \Bigg) = 0.\end{aligned}$$

Thus $\bar{X} = B(\frac{1}{2}, \frac{1}{2}\sqrt{1 + 4r})$, a ball with radius different from r.

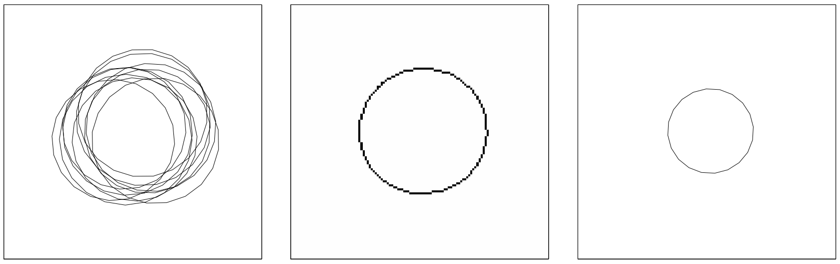
From these examples we can conclude that while $\mathrm{E}\,B(0,\xi) = B(0,\mathrm{E}\,\xi)$, the same is not true for fixed objects with random centre points, eg. $\mathrm{E}\,B(\xi,r) \neq B(\mathrm{E}\,\xi,r)$. In general, if X is a deterministic set and $\xi \in \mathbb{R}^2$ is a random position,

$$\overline{X+\xi} \neq X + \overline{\xi}.$$

Examples We have performed some experiments to get a feel for the performance of the method. The first experiment considers a deformable template model where samples are obtained by performing random translations, scalings and rotations of the circular polygon template. This is obviously a situation where the pointwise mean fails, and the results confirm this. Figure 6.1(a) shows some of the samples, while Figure 6.1(b) and (c) shows the distance average and the pointwise average, respectively. The advantage of constructing the mean independent of the representation is evident.

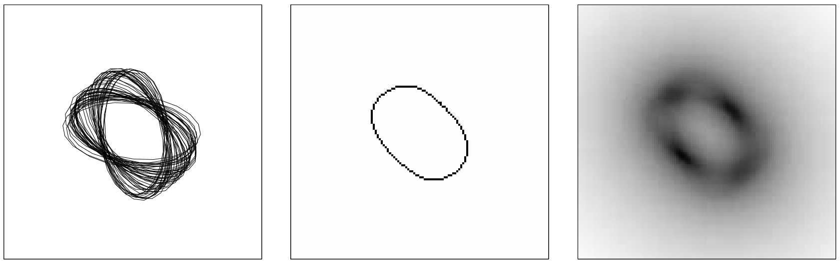
The second experiment illustrates that the method is sensitive to too large variations in shape. The setup is as above, but with an ellipsis as a template, deformed by applying rotations in the range $[0, \pi/2]$. The distance average in Figure 6.2(b) is slanted at an angle being approximately $\pi/4$, but is smaller and more roundish than the samples. To get a feel for the variation in the sample we may plot the pixelvise mean of the absolute valued distance transform for each object, see Figure 6.2(c). This might be of interest for instance in situations where one wants to known how often an objects covers a certain subset of the plane.

Throughout this report we have used the distance average for computing the average shape of samples from our template models, and the method has proven to give a very reasonably summary of the samples.



(a) (b) (c)

Figure 6.1: Average of circular polygons transformed by Euclidean similarities. (a) Samples of the polygons. (b) The distance average of the polygons, and (c) the pointwise average of the polygons.



(a) (b) (c)

Figure 6.2: (a) Samples made by random rotations of a template elliptical polygon. (b) The distance average of the polygons. (c) Pointwise mean of the absolute value of the distance transforms of each polygon.

6.4 Metrics for Grey-Scale Images

Measuring the similarity between two images is important in image analysis, for instance to measure the fit of an image estimate to a true image. It is often convenient that the measure is a metric on the space of images, and we will briefly discuss such metric for grey scale images, which we for now view as mappings $x : S \to [0,1]$ from some image domain $S \subset \mathbb{Z}^2$. Ideally an image metric should correspond to the human notion of "distance" between images, but such images are notoriously difficult to construct. Certainly, the widely used L_p-metrics

$$\|x_1 - x_2\|_p = \left(\frac{1}{|S|} \sum_{s \in S} (x_{1,s} - x_{2,s})^p \right)^{1/p} \tag{6.17}$$

are not well suited for all imaging applications. Their main drawback is that they compare images pixel by pixel, without taking into account global structures. A "good" and useful metric should account for the spatial distributions in the image, while at the same time being practically computable.

Rue (1995) and Rue (1997) uses the local correlation structure of the residuals $e = \|x_1 - x_2\|_2$ to construct image metrics that measure spatial structures. In its simples form the metric is

$$d_R(x_1, x_2) = \sum_{s \in S} L_{(1)}(x_{1,s}, x_{2,s}) + \lambda \sum_{s \in S} \sum_{k \in \Gamma} R^2_{\omega_s}(k),$$

where $L_{(1)}$ is a pixel by pixel distance, ω_s is the 3×3 neighbourhood centred in s, $\Gamma = \{(1,0),(0,1),(1,1),(-1,1)\}$, and

$$R_{\omega_s}(k) = \frac{1}{|\omega_s \cap \omega_{s-k}|} \sum_{l \in \omega_s \cap \omega_{s-k}} (e_l - \overline{e_{\omega_s}})\,(e_{l+k} - \overline{e_{\omega_s}})\,.$$

Friel and Molchanov (1998) take a different approach, using the theory for random sets. They represent images as random sets, and construct an image metric via a probability metric for their corresponding random variables. For simplicity we now treat images as mappings $x : \mathbb{R}^2 \supset \Omega \to [0,1]$. To describe an image through random sets, note that we can represent an image x as a family of sets,

$$X_t = \{\xi \in \Omega \,:\, x(\xi) \geq t\}\,, \quad 0 \leq t \leq 1,$$

and that if $u \sim \mathrm{U}[0,1]$, X_u is a random set. One metric for random sets is the integral metric

$$d_I(x_1,x_2) = d_I(X_{1,u},X_{2,u}) = \sup_{h \in \mathcal{H}} \left| \int_0^\infty \sup_{\xi \in H_t} x_1(\xi)dt - \int_0^\infty \sup_{\xi \in H_t} x_2(\xi)dt \right|,$$

where $\mathcal{H}$ is a family of non-negative functions and $H_t = \{\xi \,:\, h(x) \geq t\}$. For the case of binary images Friel and Molchanov (1998) show that the integral metric is equivalent to the Hausdorff metric Eq. (6.12) if $\mathcal{H} = \{h^\zeta \,:\, \zeta \in \Omega\}$, and $h^\zeta(\xi) = 1 - \|\xi - \zeta\|/c$. Using the same family $\mathcal{H}$ for grey scale images, the metric writes as

$$d_I(x_1,x_2) = \sup_{\zeta \in \Omega} \left| \int_0^c \left(\sup_{\xi \in B(\zeta,t)} x_1(\xi) - \sup_{\xi \in B(\zeta,t)} x_2(\xi) \right) dt \right|.$$

They now proceed by generalising the above distance in much the same way as Baddeley (1992) generalises the Hausdorff metric to construct the delta metric. Thus Friel and Molchanov (1998) defines the double integral (pseudo)-metric as

$$d_{DI}(x_1,x_2) = \left(\int_\Omega \left| \int_0^1 \left(\sup_{\xi \in B(\zeta,t)} x_1(\xi) - \sup_{\xi \in B(\zeta,t)} x_2(\xi) \right) dt \right|^p d\zeta \right)^{1/p}. \quad (6.18)$$

Smaller values of c make the metric more local; note in particular that if $c = 0$ the metric reduces to the L_p-metric, while $d_{DI}(x_1,x_2) \to |\sup x_1 - \sup x_2|$ when $c \to \infty$.

Below inn Chapter 7 we will use the metric distance function for computing distances between grey-level images, both as a means for monitoring convergence of MCMC samplers, and for comparing the fit of image estimates to a true image.

Chapter 7

Block Sampling in Restoration of Images

7.1 Introduction

In this section we will consider block sampling of spatial models with edge-preserving priors; with a particular view towards restoration of ultrasound images. It is well known that single site updating schemes might have poor convergence and mixing properties, especially for models with long-range spatial interactions. However, although block updating schemes long have been used for temporal Markov field models (Früwirth-Schnatter, 1994; Carter and Kohn, 1994), block sampling of spatial Markov fields is more difficult due to the lack of ordering of the sites. In the case of Gaussian Markov random fields (GMRFs), Rue (2001) has recently described an efficient block sampling algorithm based on sparse matrix methods. The algorithm has been shown to work well for a range of spatial models with Gaussian or near-Gaussian distributions (Rue, 2001; Knorr-Held and Rue, 2002). The ability to handle near-Gaussian distributions is especially important, as such distributions often occur in spatial models, eg. in imaging and spatial epidemiology.

The main contribution of this section is the use of block sampling for edge-preserving models of the form Eq. (4.5). Block sampling in this class of models was first used by Geman and Yang (1995), who introduced the dual formulation in Eq. (4.10), and did MAP estimation using Fast Fourier Transforms. Later other algorithms for optimisation have been introduced eg. by Charbonnier et al. (1997) and Barone (1999). However, all these attempts assume stationary Gaussian likelihoods and toroidal boundary conditions. We propose instead to use the GMRF algorithm of Rue (2001), thus allowing for non-stationary Gaussian likelihoods and to some extent non-Gaussian likelihoods. To our knowledge no comparison between single site and block samplers has been performed for edge-preserving models, and so we want to investigate whether the dual model formulation leads to better sampling properties. For even if the dual formulation allows for block sampling, it is not certain that the augmented model $\pi^*(x, b)$ (see Section 4.3) is easier to sample from; it might be inherently more resistant to sampling. We describe the block sampling approach in Section 7.2, together with a short example comparing block and single site samplers.

We then extend the block sampling approach to edge-preserving prior models with non-Gaussian likelihoods and consider restoration of radio frequency ultrasound images. The restoration is based on the imaging model described in Chapter 3. This model was first used for image restoration and segmentation in Hokland and Kelly (1996), and extended in Husby et al. (2001) and Langø et al. (2001). The Bayesian model is formulated in Section 7.3, and results of the experiments are given in Section 7.4.

7.2 Block Sampling in Edge-Preserving Models

We start with two simple examples performing restoration of a piecewise constant toy image and a confocal microscopy image of a human melanoma cancer cell. The purpose of the experiments is to compare block sampling with standard single site sampling for an edge-preserving model.

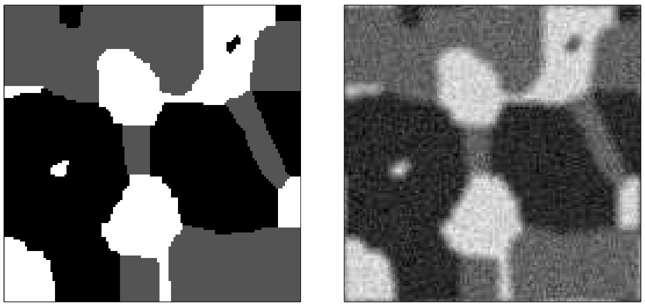
The piecewise planar image is very well described by an edge-preserving prior model with only first order cliques. The image x is shown in Figure 7.1 (a), and is degraded by applying a flat 5×5 blur mask and adding Gaussian noise. The resulting image is shown in Figure 7.1 (b). Using the dual formulation of Section 4.3.1 the conditional density for x and the auxiliary variable b is

$$\pi(x, b \mid y) \propto \exp\left(-\sum_{s \in S} \beta \sum_{m=1}^{M} \omega_m \left(\frac{1}{2}\left(D_s^{(m)} x - b_s^{(m)}\right)^2 + \psi\left(b_s^{(m)}\right)\right) + \frac{1}{\sigma^2}\left(y_s - \sum_k h_k x_{s+k}^2\right)^2\right) \tag{7.1}$$

Let $D^{(m)}$ be matrices representing the difference operators $\{D_s^{(m)}\}$, $D^T = (D^{(1)T}, \ldots, D^{(M)T})$, and let $b^T = (b^{(1)}, \ldots, b^{(M)})$. Furthermore, define $W = \operatorname{diag}(\omega_1, \ldots, \omega_M) \otimes I_n$, and let H be a matrix representing the point spread function h. Then it is easy to see that the full conditional distribution for the true image x has density

$$\pi(x \mid y, b) \propto \exp\left(-\frac{1}{2} x^T \left(\beta D^T W D + \frac{1}{\sigma^2} H^T H\right) x + b^T W D x\right), \tag{7.2}$$

which is a Gaussian Markov random field with inverse covariance matrix $Q = \beta D^T W D + \sigma^{-2} H^T H$ and mean vector $\mu^T = b^T W D\, Q^{-1}$. Assuming toroidal boundary conditions x can be sampled very efficiently using FFTs as detailed in Geman and Yang (1995). In the general situation one can use Cholesky decomposition and sparse matrix methods as in Rue (2001). Also note that if the likelihood is non-Gaussian, the methods in Rue (2001) can be used to construct a Gaussian approximation of the aposteriori distribution; this approximation can be used as a proposal distribution in a Metropolis-Hastings algorithm.

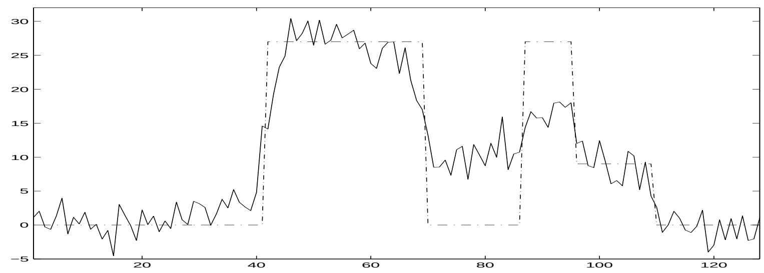


(a) (b)

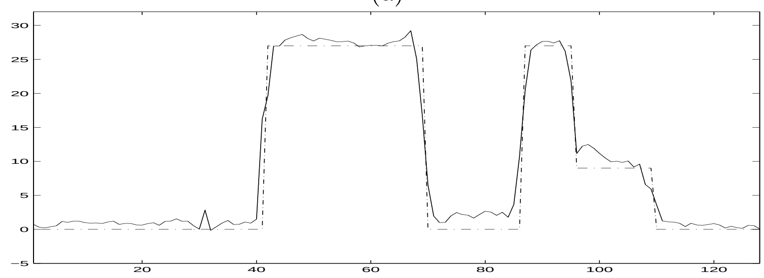
Figure 7.1: Piecewise planar image. (a) original image. (b) The image in (a) degraded by a constant 5×5 blur and Gaussian noise with standard deviation 2.

We sampled from the posterior distribution for x using both the augmented model Eq. (7.1) and the original model Eq. (4.5). In both cases we used $\beta = 50$ and $\delta = 10$, and the samplers were initialised with flat images having value zero. The block sampler converged very quickly; Figure 7.2 (b) shows a cross section through the true image, together with the 10th sample of the block sampler. For comparison we have displayed the data y in Figure 7.2 (a), and the 100th sample of the single site sampler in Figure 7.2 (c). The block sampler seems to be very efficient; a few iterations seems to be enough to get a good restoration. It compares very well with the single site sampler that uses far more iterations to obtain the same results. It should be noted that while in the current implementation both samplers use approximately the same amount of time for each iteration, the single site sampler should be slightly faster in an optimal implementation.

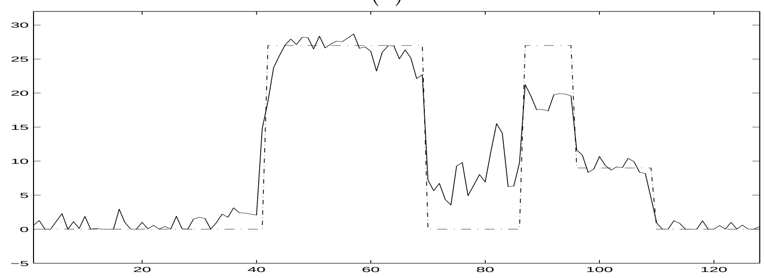
Next we consider the cell images, one of which is shown in Figure 7.3 (a). This problem is slightly more challenging, and it is also of interest to see how well the model can handle images that are not planar. To get quantitative measurements we degrade the true image x by blurring with a Gaussian kernel with standard deviation 3, and adding independent zero mean Gaussian noise with $\sigma = 15$. The resulting image y is shown in Figure 7.3 (b). Figure 7.3 (c) shows a posterior mean estimate of the cell image based on 1000 iterations of the block sampler with parameters $\beta = 200$ and $\delta = 50$, and $\psi(u) = |u|/(1 + |u|)$. Visually the restoration is quite close to the original image, although some smoothing has taken place. The results were similar for a wide range of parameters, but smaller values of the ratio β/δ^2 led to smoother images. Since the truth is known, a qualitative comparison between the restoration and



(a)



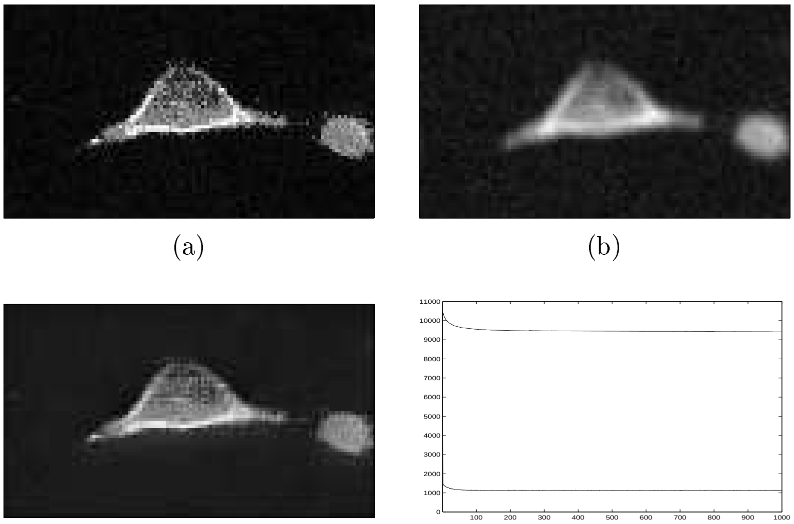
(b)



(c)

Figure 7.2: Piecewise planar image. (a) Cross section through the image. The dashed line is the truth, and the full line is the data. (b) The tenth sample of the block sampler with $\beta = 50$ and $\delta = 10$, overlaid the truth. (c) The 100th sample of the corresponding single site sampler.

the true image can be made. Figure 7.3 (d) plots the double integral distance (Section 6.4) between the Markov chain samples and the true image Figure 7.3 (a), both for the block sampler (lower curve) and a single site sampler run for comparison. A constant image was used as an initial point, and the block sampler seems to reach a stable state within a few iterations. In comparison, the single site sampler seems to converge much slower. In conclusion, the dual model formulation combined with a good block sampling algorithm seems to well suited for recovering discontinuities in piecewise smooth images, and seems to be more efficient than the original model combined with a single site sampler.



(a) (b) (c) (d)

Figure 7.3: Edge preserving restoration. (a) Confocal microscopy image of a human melanoma cancer cell. (b) Data obtained by blurring and adding Gaussian noise. (c) Restoration of the image using the Geman & Yang model and the GMRF algorithm of Rue (2001). (d) Trace plot of the double integral distance between the true image and the Markov chain samples from a single site sampler (top curve) and a block sampler (bottom curve), both with the same initial image.

7.3 A Bayesian Model for Ultrasound Images

7.3.1 Introduction

In this section we derive the Bayesian Markov random field formulation of the ultrasound imaging model in Chapter 3. We represent the imaged region as a grid $S \subset \mathbb{Z}^2$ with $n = n_1 n_2$ nodes, and index the nodes or pixels by lateral index s_1 and radial index s_2, and write $s = (s_1, s_2)$ when appropriate.

Let $y = \{y_s,\, s \in S\}$ be the observed radio frequency image, modelled as resulting from a convolution of the true radio frequency image $x = \{x_s,\, s \in S\}$ with the imaging system point spread function h, with the addition of noise consisting of independent and identically distributed Gaussian random variables. The goal in this section is to estimate x from y. We follow standard practice and assume the pulse function h to be spatially invariant (Christensen, 1988; Angelsen, 2000), in which case we can write

$$y_{s_1,s_2} \,|\, x, h \sim \mathrm{N}\left(\sum_{k_1,k_2} h_{k_1,k_2} x_{s_1+k_1,s_2+k_2}\,, \lambda^{-1} \right), \tag{7.3}$$

where λ is the unknown precision of the noise process. We assign a $\Gamma(a, b)$ prior to λ, ie. a Gamma distribution with mean a/b and variance a/b^2. We assume the pulse function to be a separable Gaussian function with a sine oscillation in the radial direction, ie.

$$h_{k_1,k_2} \propto \exp\left(-\frac{k_1^2}{2\sigma_{ra}^2} - \frac{k_2^2}{2\sigma_{la}^2} \right) \cos\frac{2\pi k_1}{\omega}. \tag{7.4}$$

Empirical studies indicates that this is a good approximation (Ødegård, 1995) which seems to be quite robust with respect to misspecification of the parameters, see Langø et al. (2001). Whether the stationarity assumption is reasonable or not, depends on properties of the imaged tissue; see eg. Ødegård (1995) for a discussion. Certainly having a non-stationary pulse will introduce additional computational costs, and should thus be avoided if possible. Our result did not seem to justify such an increase in computation time.

7.3.2 A model for diffuse scattering

Recall from Section 3.2 that in the model for diffuse scattering, the radio frequency echoes are apriori independent zero mean Gaussian variables with variance dependent on the acoustical properties of the underlying tissue, ie.

$$x_s \,|\, \sigma_s^2 \sim \mathrm{N}\left(0, \sigma_s^2\right), \quad \forall s \in S. \tag{7.5}$$

The independence assumption may not be realistic in all applications, and the prior could be replaced by a distribution with some simple dependency structure. We have not tried this. The variance σ_s^2 is given by the expression in Eq. (3.4), and depends on the scattering intensity of the tissue region imaged at site s. Usually the variance has been assumed to be constant over the whole image, and an important contribution in Hokland and Kelly (1996) and Husby et al. (2001) is to let the variance vary throughout the image, acknowledging that this particular parameter contains information about the anatomy of the imaged tissue. In fact, the radio frequency field x contains no additional information, and can thus be seen as a nuisance parameter. However, as argued in Husby et al. (2001), a model formulation containing x has great computational advantages, since the distribution for the variance field σ^2 given the data has no Markov structure, whereas the distribution for σ^2 given the data y *and* the radio frequency field x has a neighbourhood structure depending on the support of the point spread function h.

To avoid problems with positivity we reparameterize the model and define a log-variance field $\nu = \{\ln \sigma_s : s \in S\}$. The choice of prior model for this field should be justified from physical considerations about the imaged tissue, and we use the following assumptions:

- scattering intensity tend to be approximately constant within regions of homogeneous tissue,
- abrupt changes in scattering intensity may occur at interfaces between different tissue types.

Based on these assumptions it is reasonable to model the log-variance field ν as being piecewise smooth with homogeneous subregions corresponding to the different tissue types in the imaged region. As explained in Section 4.2, edge-preserving functionals are well suited for modelling such fields. Thus we define the prior distribution for ν as

$$\pi(\nu) \propto \exp\left(-\beta \sum_{m=1}^{M} \omega_m \sum_{s \in S} \phi\left(D_s^{(m)} \nu\right)\right), \tag{7.6}$$

where $\phi(\cdot)$ is a functional from the edge preserving class defined in Eq. (4.6), $\omega_1, \ldots, \omega_M$ are positive constants, β is a positive scaling factor, and $D_s^{(m)}$ are the discrete derivatives defined in Eq. (4.2). Unless otherwise stated, we will use the four first order cliques displayed in Figure 4.1, with the corresponding constants $\omega_1 = \omega_2 = 1$, $\omega_3 = \omega_4 = 1/\sqrt{2}$. The scaling parameters β and δ are discussed below, in Section 7.3.4. Finally, we note that other approaches are also feasible, for instance using line processes (Geman and Geman, 1984; Higdon et al., 1997). However, line processes introduce additional complexity, and experiments suggests that they introduce too much smoothing in blurred images (Higdon et al., 1997).

Note that this diffuse scattering model rests on a number of conditions, notably that there is a large number of diffuse scatterers in the regions imaged by each pixel, and that there is a negligible coherent scatter component, see Section 3.2. These conditions do not apply to all tissue types, but there exists statistical tests for identifying the regions of an image for which the conditions apply (Georgiou and Cohen, 1998).

7.3.3 Posterior distribution

Combining the Gaussian likelihood Eq. (7.3), the model for the true radio frequency signal Eq. (7.5), and the prior for the underlying log-variance field Eq. (7.6), we obtain the posterior density

$$\begin{aligned}\pi(x,\nu,\lambda\,|\,y) \propto \lambda^{n/2} \exp\left(-\frac{1}{2}\lambda\sum_{s\in S}\left(y-\sum_k h_k x_{s+k}\right)^2\right)\\ \times \exp\left(-\frac{1}{2}\sum_{s\in S} x_s^2 \exp(-2\nu_s) - \sum_{s\in S}\nu_s\right)\\ \times \exp\left(-\beta\sum_{m=1}^{M}\omega_m\sum_{s\in S}\phi\left(D_s^{(m)}\nu\right)\right)\times\lambda^{a-1}\exp\left(-b\,\lambda\right).\end{aligned} \tag{7.7}$$

In Husby et al. (2001) the precision λ was assumed constant, and the rest of the parameters where updated using single site Metropolis - Hastings samplers. Note that they used a slightly different formulation of the model. The best is perhaps to use random walk samplers, but both x and λ can be updated directly from their full conditionals

$$\begin{aligned} x_s\,|\,x_{-s},y,\nu,\lambda &\sim \mathrm{N}\left(\frac{\lambda\sum_k h_{k-s}\Upsilon_k}{\lambda\sum_k h_k^2+\exp(-2\nu_s)},\ \left[\lambda\sum_k h_k^2+\exp(-2\nu_s)\right]^{-1}\right)\\ \lambda\,|\,y &\sim \Gamma\left(a+\frac{1}{2}n,\ b+\frac{1}{2}\sum_s\left(y_s-\sum_k h_k x_{s+k}\right)^2\right).\end{aligned}$$

Here $\Upsilon_k = y_k - \sum_{l\neq k-s} h_l x_{k-l}$. To summarise, sampling from the full model where done as follows in Husby et al. (2001).

ALGORITHM 1

STEP 1. *Select initial values for x and ν.*

STEP 2. *Select a random ordering of S, and for each $s\in S$, do*

STEP 3. *Propose a new ν_s' from $U[\nu_s/f,\nu_s f]$, $f>0$.*

STEP 4. *Accept the proposed sample with probability*

$$\alpha = \min\left\{1, \frac{\pi(\nu_s'\,|\,x,\nu_{-s})\,\nu_s}{\pi(\nu_s\,|\,x,\nu_{-s})\,\nu_s'}\right\}. \tag{7.8}$$

STEP 5. *Propose a new x_s' from $N(x_s,\sigma_{rw}^2)$*

STEP 6. *Accept the proposed value with probability*

$$\alpha = \min\left\{1, \frac{\pi(x'_s \mid x_{-s}, \nu)}{\pi(x_s \mid x_{-s}, \nu)}\right\}. \tag{7.9}$$

STEP 7. *Repeat steps* $2-6$.

However, Husby et al. (2001) observed quite slow convergence of the single site sampler, and it is thus of interest to investigate whether block sampling will lead to computational improvements. For the radio frequency field x block sampling is easily implemented. For simplicity we assume the parameters β, δ, and λ to be fixed in the following discussion. Let H be a matrix representing the point spread function h, ie. having entries

$$H_{sr} = h_{|s-r|}, \quad 1 \leq s, r \leq n.$$

In practice the point spread function is truncated so that $h_k \equiv 0$ if k is greater than some threshold K. Furthermore, let V be the diagonal matrix $V = \mathrm{diag}(\exp(-2\nu_1), \ldots, \exp(-2\nu_n))$, then it is easy to see that the conditional distribution for x may be written as

$$\pi(x \mid \lambda, \nu) \propto \exp\left(-\frac{1}{2}x^T\left(\lambda H^T H + V\right)x + y^T H x\right), \tag{7.10}$$

which is of the form Eq. (2.5) with $b_s = \sum_k h_k y_{s+k}$ and

$$Q_{sr} = \begin{cases} \sum_k h_k^2 + \exp(-2\nu_s) & s = r \\ \sum_k h_{|s-k|} h_{|r-k|} & |s-r| \leq 2K+1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence we can efficiently sample x in one single block using the GMRF algorithm of Rue (2001) (see Section 2.2.1).

The conditional distribution for the log-variance field ν is non-convex, and therefore block sampling is not straightforward to implement. On the other hand, because of the non-convexity, this part of the model might benefit the most from going from single site to block updating. By using the dual model formulation of Geman and Yang (1995) (Section 4.3.1) it is at least in theory possible to construct a block Metropolis-Hastings scheme as follows. Remember from Section 4.3.2 that we can substitute the prior distribution Eq. (7.6) with a dual distribution

$$\pi^*(\nu, b) \propto \exp\left(-\beta\sum_{m=1}^{M}\omega_m\sum_{s\in S}\left(\frac{1}{2}\left(D_s^{(m)}\nu - b_s^{(m)}\right)^2 + \psi\left(b_s^{(m)}\right)\right)\right), \tag{7.11}$$

where $\{b_s^{(m)}\}$, $m = 1, \ldots, M$ are auxiliary fields, and the functional $\psi(\cdot)$ is related to $\phi(\cdot)$ by Eq. (4.11). The corresponding full conditional distribution for ν is given by

$$\pi^*(\nu\,|x,b) \propto \exp\left(-\sum_{s\in S}\left(\frac{1}{2}\beta\sum_{m=1}^{M}\omega_m\left(\left(D_s^{(m)}\nu\right)^2 - 2b_s^{(m)}D_s^{(m)}\nu\right)\right.\right.$$
$$\left.\left.+\frac{1}{2}x_s^2\exp(-2\nu_s)+\nu_s\right)\right), \quad (7.12)$$

which is quadratic in ν except for the term involving $\exp(-2\nu_s)$. Since there exists efficient methods for block sampling Gaussian Markov random field models, it is tempting to try to construct a Gaussian approximation of the density (7.12). In a neighbourhood around the current value ν_s we can approximate the exponential as

$$\exp(-2\nu_s) \approx A_s + 2B_s\nu_s' + C_s{\nu_s'}^2,$$

where the coefficients depend on ν_s. Then we get an inhomogeneous Gaussian Markov random field that approximates the distribution (7.12) and that can be used as a proposal distribution in a block Metropolis-Hastings sampler:

$$q(\nu'\mid\nu,b,x) = \frac{1}{Z(\nu)}\exp\left(-\sum_{s\in S}\left(\frac{\beta}{2}\sum_{m=1}^{M}\omega_m\left(\left(D_s^{(m)}\nu'\right)^2 - 2b_s^{(m)}D_s^{(m)}\nu'\right)\right.\right.$$
$$\left.\left.+\frac{1}{2}x_s^2C_s{\nu'}_s^2 + \left(x_s^2B_s+1\right)\nu_s'\right)\right). \quad (7.13)$$

This is again a distribution of the form Eq. (2.5) with

$$\begin{aligned} b_s &= \frac{\beta}{\delta}\sum_{m=1}^{M}\omega_m\left(b_s^{(m)} - b_{r(s,m)}^{(m)}\right) - x_s^2B_s - 1 \\ Q_{rs} &= \begin{cases} 2\beta/\delta^2\sum_{m=1}^{M}\omega_m + C_sx_s^2 & r = s \\ \beta/\delta^2\omega_{m(s,r)} & r\sim s \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

where $m(s,r)$ the clique type for sites s and r, and $r(s,m)$ is the unique site belonging to the m-clique of site s. This kind of approximation is used in Rue (2001) and Knorr-Held and Rue (2002), which also discuss different ways of obtaining the approximation Eq. (7.13).

The auxiliary variable b has conditional distribution

$$\pi^*(b\,|\,\nu) \propto \exp\left(-\beta \sum_{m=1}^{M} \omega_m \sum_{s\in S}\left(\frac{1}{2}(b_s^{(m)})^2 - b_s^{(m)} D_s^{(m)}\nu + \psi\left(b_s^{(m)}\right)\right)\right). \tag{7.14}$$

Note that all the components are independent, so that this distribution can easily be sampled using eg. envelope rejection sampling (Robert and Casella, 1999). Alternatively one can use a random walk Metropolis sampler.

Sampling from the full model can now be done as follows.

ALGORITHM 2

STEP *1. Select initial values for x, ν, and b.*

STEP *2. Draw a sample ν' from $q(\nu' \mid \nu, b, x)$.*

STEP *3. Accept the proposed sample with probability*

$$\alpha = \min\left\{1, \frac{\pi^*(\nu' \mid x, b)\, q(\nu \mid \nu', b, x)}{\pi^*(\nu \mid x, b)\, q(\nu' \mid \nu, b, x)}\right\}. \tag{7.15}$$

STEP *4. Generate a sample b from $\pi^*(b \mid \nu)$ using envelope rejection sampling.*

STEP *5. Draw x from the conditional distribution $\pi(x \mid \nu, y)$.*

STEP *6. Repeat steps $2-5$.*

Note that the normalising constant of q does not cancel in the ratio Eq. (7.15); but this constant can be computed efficiently as in Rue (2001). We will return to this model later, in Section 7.4, where it is applied for restoration of simulated and real ultrasound images.

7.3.4 Treatment of hyperparameters

The edge preserving prior model in Eq. (7.6) contains two unknown hyperparameters, β and δ. The value of these parameters greatly affects the properties of the restored images: a large value of β yields smooth images, whereas a small value gives too much noise; likewise would a too large value of δ smooth out edges. Consequently there has been a great deal of interest in estimating these parameters. Geman and Reynolds (1992) and Hurn and Jennison (1996) use heuristics to find an upper bound for β given δ, while Jalobeanu, Blanc-Feraud and Zerubia (1998) use Markov chain Monte Carlo maximum likelihood to estimate the hyperparameters. Perhaps more interesting is the use of fully Bayesian methods to integrate out the effect of the hyperparameters altogether. Notable examples are Higdon et al. (1997) and Weir (1997), the last of which considers a model similar to the one discussed above, but with only one parameter. The problem is to evaluate the normalising constant of the prior Eq. (7.6), and Weir (1997) uses reverse logistic regression (Geyer, 1991). This method is highly computer intensive, and Chen, Shao and Ibrahim (2000) reviews other ways of approximating the normalising constant. A particularly simple and elegant method proceeds by considering the derivative of the normalising constant. If we let $\pi(x) \propto \exp(\beta\Phi(x))$, then

$$Z(\beta) = \sum_x \exp\left(\beta\Phi(x)\right), \tag{7.16}$$

and

$$\begin{aligned} \frac{dZ(\beta)}{d\beta} &= \sum_x \Phi(x)\exp\left(\beta\Phi(x)\right) \\ &= Z(\beta)\mathrm{E}_{x \mid \beta}\Phi(x). \end{aligned}$$

Solving this differential equation one obtains

$$\log\left(Z(\beta')/Z(\beta)\right) = \int_\beta^{\beta'} \mathrm{E}_{x \mid \widetilde{\beta}}\Phi(x)d\widetilde{\beta}, \tag{7.17}$$

where the integral can be approximated by estimating the expectation for a range of β-values and using a smoothing spline to construct a function that can be integrated numerically. The method seems to be stable and reasonably efficient, but the computational burden increases sharply with the number of unknown parameters. In our experiments we have for simplicity set appropriate values of the hyperparameters by trial and error.

7.4 Block Sampling in Restoration of Ultrasound Images

The examples in this section concerns a slightly harder problem than the one in Section 7.2, namely estimation of a true radio frequency image x based on observed ultrasound images, and using the Bayesian model described in Section 7.3. This has previously been done in Husby et al. (2001), but the goal now is to investigate if block sampling will lead to improved mixing and faster convergence than the single site random walk sampler used in Husby et al. (2001). One problem is that the full conditional for the log-variance field ν (Eq. (7.6)) is not Gaussian, hence we need to use a Gaussian approximation of the density as a proposal in a Metropolis-Hastings sampler. This approximation most be reasonably good in order to get decent acceptance rates.

We first use a synthetic dataset to get quantitative measures on the performance of the method, and to find appropriate ranges for the unknown hyperparameters β and δ. Secondly, we try to use the method for restoration of a real ultrasound image,

7.4.1 Simulated image

The dataset was constructed to have a structure similar to the real images considered in Section 8.4.2, and is based on the 64×64 binary two region image in Figure 7.4 (a). The area of the foreground region is 625 pixels. We then generated two log-variance fields using a Gaussian Markov random field approximation to a Gaussian field having an exponential correlation function with range 36. Whether this choice of correlation structure corresponds to the one found in real ultrasound images is if course hard to say, but it corresponds well with the assumption of smoothly varying scatter properties. The Gaussian fields were combined to yield the log-variance field in Figure 7.4 (b), with the background field having mean zero, and the foreground field having mean value equal to 1/2. A radio frequency field x was generated using Eq. (7.5) (Figure 7.4 (c)), and finally a simulated ultrasound image was generated using Eq. (7.3), with the point spread function having wavelength 1, and standard deviations 3/2 in both the radial and lateral directions. The image is shown in Figure 7.4 (d) after taking the absolute value and doing a log compression to enhance contrasts.

To find plausible values of the hyper parameters β and δ, the model in Eq. (7.6) has been used for estimating the true radio frequency and log-variance images for different values of the pair (β, δ). Figure 7.5 show posterior mean estimates of the log-variance field for the different parameter values, and we can clearly see that the parameters affect the properties of the model, as higher values of β and δ yield smoother images. However, the corresponding radio frequency images were visually very similar, and this indicates that variations of β and δ within the given range does not have any significant impact on the quality of the image restoration. To get a quantitative comparison between the models, we have used the double integral distance (Section 6.4) with $c = 10$ and $p = 2$ to measure the distance between the different image estimates and the true images. The results are shown in Table 7.1, indicating the the highest value of β gives the best estimate for ν. For x the differences are small, but the same conclusion seems to hold. Different values of c led to the same ranking between the images.

Moving on to the main experiment, we tried restoration using both the dual model Eq. (7.11) and Algorithm 2, and the original model Eq. (7.6) and Algorithm 1. We used $\beta = 10$ and $\delta = 1/2$ in both cases, and ran the block sampler for 1000 iterations and the single site sampler for 5000 iterations, since the block sampler used approximately 5 times more CPU time to do each update. To get reasonable acceptance rates for the block sampler, we updated the image in smaller blocks containing $50 - 100$ pixels. To avoid unwanted edge effects we used different, and over-lapping blocks for each iteration. Figure 7.6 (b) and (c) show the posterior mean estimates obtained from the output of the block sampler and single site sampler, respectively. Figure 7.7 shows the double integral distance between the samples and the truth for the block sampler (bottom curve), and for the first 1000 iterations of the single site sampler (top

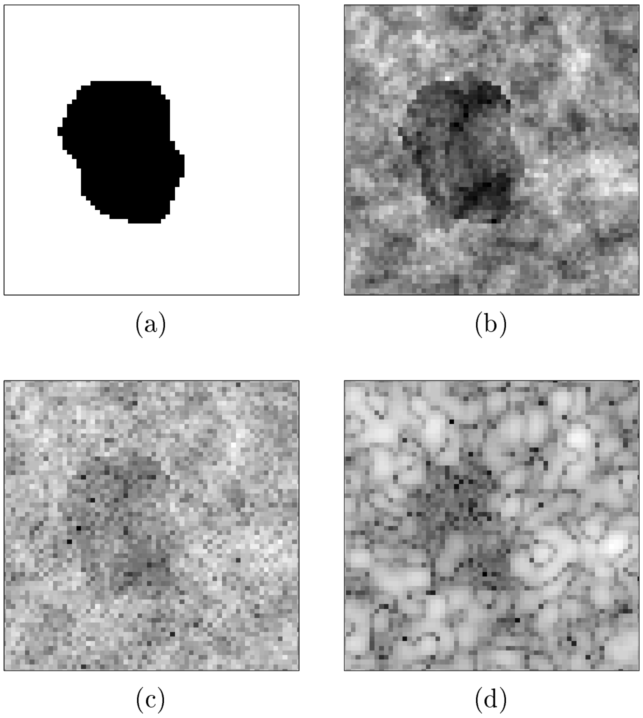


Figure 7.4: Simulated ultrasound image. (a) The binary image describing the anatomy. (b) The log-variance field representing the variations in acoustic properties. (c) The simulated true radio frequency image. (d) The corresponding observed image.

Table 7.1: The double integral distance between the true images and posterior mean estimates in Figure 7.5.

(β, δ)	ν	x
$(10, 1/2)$	5.53	459.0
$(10, 1)$	6.79	459.7
$(10, 4)$	9.33	463.3
$(5, 1/2)$	6.74	461.5
$(5, 1)$	8.13	462.7
$(5, 4)$	10.34	464.2
$(1, 1/2)$	12.36	462.2
$(1, 1)$	12.38	463.9
$(1, 4)$	12.68	464.7

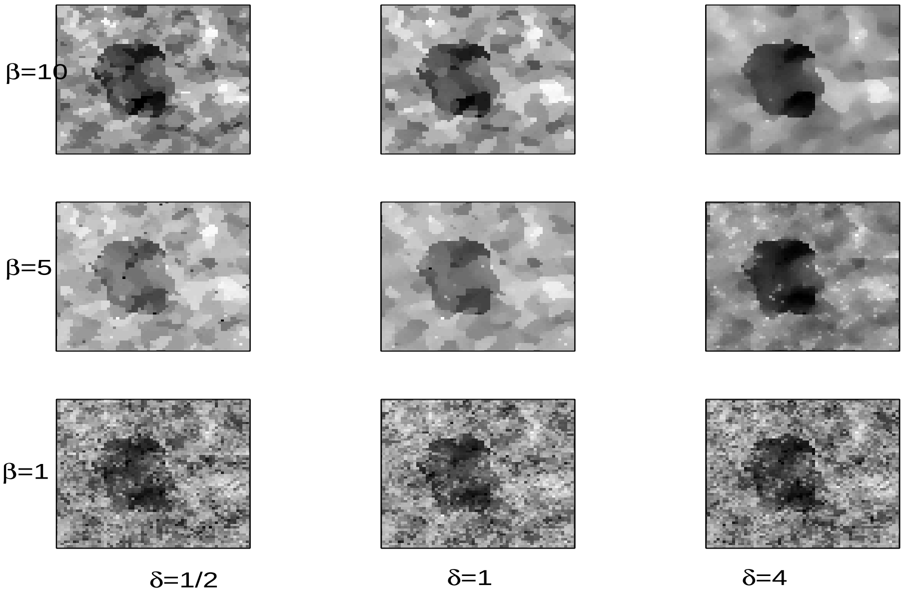


Figure 7.5: Posterior mean estimates of the log-variance field using different parameters β and δ. Row-wise from the top: $\beta = 10$, $\beta = 5$, and $\beta = 1$. Column-wise from the left: $\delta = 1/2$, $\delta = 1$, and $\delta = 4$.

curve).

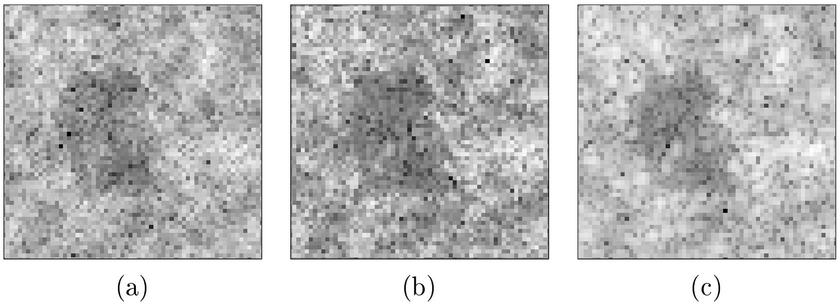


Figure 7.6: Restoration of simulated ultrasound image. (a) The true image. (b) Posterior mean estimate obtained using the dual model and block sampling. (c) Posterior mean estimate obtained using single site sampling.

We again see that the block sampler reaches a stable state very quickly, while the single site sampler seems to be moving very slowly in the same direction. Moreover, the posterior mean estimate from the block sampler (Figure 7.6 (b)) seems to be visually better than the estimate from the single site sampler (Figure 7.6 (c)).

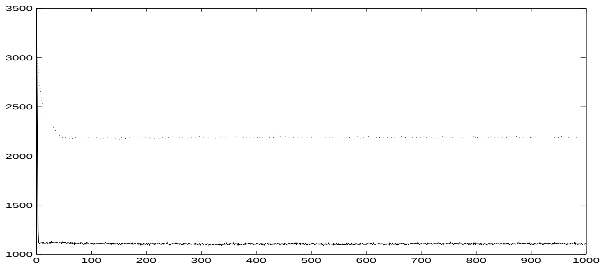


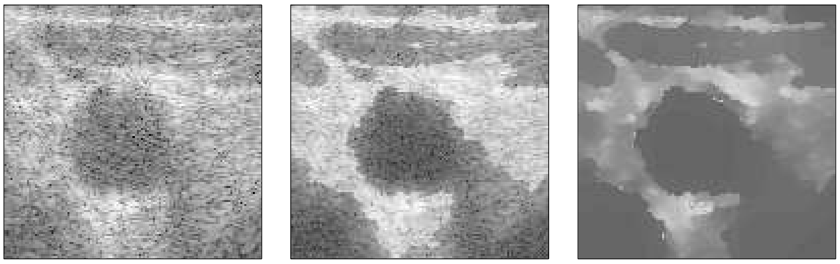
Figure 7.7: Restoration of simulated ultrasound image. A distance measure between the true image and the samples of the single site sampler (top curve) and the block sampler (bottom curve).

7.4.2 Real ultrasound image

In this last example we consider restoration of a radio frequency ultrasound image of a cross section through the carotid artery, shown plotted in polar coordinates in Figure 7.8 (a). We use the model in Eq. (7.7), where we for simplicity have set $\lambda = 1/2$ and kept β and δ fixed. We assume that the true image is degraded by the blur function in Eq. (7.4), and we have estimated the parameters as follows. The wavelength ω is found by inspecting the data, as the intensity peaks are regularly separated in most of the image. The parameters σ_{ra} and σ_{la} can be estimated in the frequency domain as in Hokland and Kelly (1996). The effect of mis-specifying the parameters of the point spread function is discussed in Langø et al. (2001).

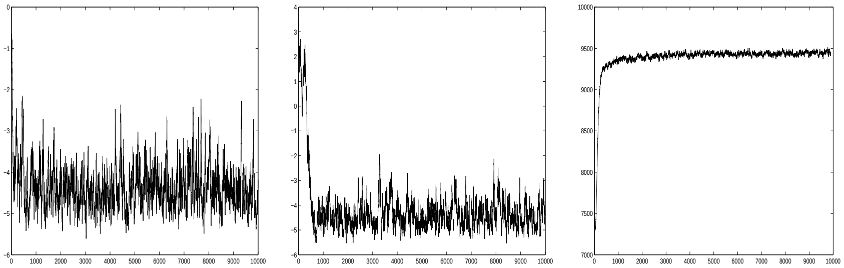
We have again tried both block and single site sampling, but we have so far had limited success with the former approach. The acceptance rate (7.15) tends to get too low, and we suspect the reason might be the large local variations in the radio frequency field x. Because of these variations, the posterior density (7.12) might be too skewed to be well approximated by the Gaussian density (7.13). Thus one needs methods for constructing better approximations, and recently there has been some promising work in that direction (Rue, Steinsland and Erland, 2001).

The following results are thus obtained using the single site sampler, that, although slow, seems to yield reasonable results. Posterior mean estimates of the radio frequency and log-variance images are shown in Figure 7.8 (a) and (b), respectively. The results are for $\beta = 5$ and $\delta = 2$, but other values in the range $1 \leq \beta \leq 10, 1/2 \leq \delta \leq 5$ gave similar results. Convergence diagnostics are shown in Figure 7.9, indicating good mixing. Obviously there is no way of comparing the restoration with the unknown true image, but the restoration seems to be less noisy and blurred, and with the important anatomical features still intact. However, for restorations to be more reliable, the hyperparameters β, δ, and λ should be integrated out in a fully Bayesian approach.



(a) (b) (c)

Figure 7.8: Image restoration using the edge-preserving model in Section 7.3 with parameters $\beta = 5$, $\delta = 2$, and $\lambda = 1/2$. (a) Log-compressed radio frequency image of a cross section through the carotid artery in polar coordinates. (b) A posterior mean estimate of the true radio frequency image. (c) The corresponding posterior mean estimate of the underlying log-variance field.



(a) (b) (c)

Figure 7.9: Convergence diagnostics for the experiment in Figure 7.8. Panels (a) and (b) show trace plots of the log-variance at different positions, while panel (c) shows a trace plot of the functional $f(\nu) = \beta \sum_m \omega_m \sum_i \phi(D_i^{(m)}\nu)$.

Finally, we note on a limitation of our approach. It is well known that point spread functions in ultrasound do not have exactly Gaussian shapes, and that they vary both laterally and radially in images (Ødegård, 1995). This is particularly a problem when the imaged tissue lies deep into the body. So even if the Gaussian approximation works reasonably well in our examples, the error introduced places limitations on the accuracy of the estimates, and on the quality of the restorations. Some attempts on estimation of 2D ultrasound pulse functions exist (Jensen and Leeman, 1994; Taxt and Strand, 2001), but this is an area that need further study. Incidentally, we tried to let some of the pulse function parameters vary spatially, but the results were not promising. There is always a danger of trying to model too many factors; this might lead to identifiability problems, and to severe sampling problems.

CHAPTER 8

A Model for Contour Detection in Ultrasound Images

8.1 INTRODUCTION

In this second example we consider contour detection in ultrasound images. We will restrict ourselves to images with one object, and consider estimation of the outline and cross-sectional area of the carotid artery. Images of the carotid artery are used in detection of atherosclerosis using the fact that diseased arteries are less likely to dilate in response to infusion of acetylcholine. Estimating the cross-sectional area of the artery before and after infusion is difficult because changes can be masked by blur, edge reflections, and image artifacts (speckle) introduced in the imaging process. An automated procedure should be able to assess these problems properly, as well as quantifying the uncertainty of the given answer, for instance by means of an interval estimate. Thus there is a real need for Bayesian methods with emphasis on realistic modelling of both object and data. Our approach will be fully Bayesian, since fixed parameter values might affect the interval estimates in unknown ways. Extension of our model to allow for an unknown number of objects is straightforward in principle, although it severely affects the computational complexity (Rue and Hurn, 1999).

The reasons for using an explicit shape model are twofold. Of course, when we are interested in the shapes of the imaged objects *per se*, it makes sense to model them explicitly, instead of eg. defining them through morphological operations on the texture Markov random fields. Unlike implicit models, for instance the one used in the previous section, an explicit shape model allows us to model apriori information on the structure of the imaged tissue. For instance, the number of objects, and their type, shape, and texture can be given informative prior distributions with appropriate variability.

In addition, shape models can be useful in restoration of ultrasound images, just because of the ease with which we can model a priori knowledge about the imaged tissue. The template or templates will define a partitioning of the imaged region, hopefully corresponding to the different organs and tissue types. Thus we have an explicit division of the imaged region into different scattering regimes, avoiding the computational difficulties of the implicit edge models in Chapter 4. Furthermore, the specular reflection along tissue interfaces (see eg. Figure 1.1) can be dealt with explicitly, and this will prove to be a great advantage. Of course, the use of explicit shape models introduces strong prior constraints, and is thus dependent on reasonable prior information.

There are two important aspects of this chapter. First, it shows the use of Bayesian methods for a rather complicated, real application, and demonstrates the importance and feasibility of using realistic models for both objects and data. A particularly important feature of our model is that the deformable template can be used for modelling the strong reflections occurring at tissue interfaces normal to the ultrasound beam. Secondly, the experiments we have performed indicate that it is important to take great care in designing the Markov chain sampler for complicated, hierarchical models. We have previously shown that block sampling in many situations performs better than single site sampling, but for our model it turns out that block sampling performs best when we update all or nearly all variables jointly. In fact, algorithms updating the variables sequentially sometimes seem to be mixing very slowly, and might thus give misleading results.

The chapter is organised as follows. We start by defining the deformable template model in Section 8.2.1, also discussing different parameterisations. The ultrasound likelihood model is redefined in Section 8.2.2, where we also briefly describe the standard likelihood model often used in imaging. This model will be used as a benchmark model. Sampling of the model is discussed in Section 8.3, while the largest part of the chapter, Section 8.4, shows experiments using both simulated and real ultrasound images. Here we discuss the effect of modelling edge reflections, and compare different sampling schemes.

8.2 Bayesian Model Formulation

8.2.1 Prior distribution for the contour

Our aim is to estimate the boundary of a blood vessel or an other elliptically shaped object such as an aneurism (see Figure 1.1). Objects with different shapes can be treated similarly, although in practice there is a limit to how complex the shapes can be.

We use the deformable template model defined in Section 5.3.2; we let the template e^0 be the m-sided polygon with vertices $(\cos 2\pi k/m, \sin 2\pi k/m)$, $k = 0, \dots, m-1$, and let t be the vector of transformations, $t^T = (t_0^T, t_1^T) = (t_{0,0}, \dots, t_{0,m-1}, t_{1,0}, \dots, t_{1,m-1})$. We assume that t_0 and t_1 are a priori independent, and for simplicity we let both be Gaussian with mean zero and a common tridiagonal precision matrix Q. Denote by $Q_t = I_2 \otimes Q$ the precision matrix for t. To keep the behaviour of the model approximately independent of the the number of edges, Q is parameterised as in Eq. (5.11). The unknown parameters κ and η are given hyperpriors $\Gamma(a_\kappa, b_\kappa)$ and $\Gamma(a_\eta, b_\eta)$, respectively. The closure constraint Eq. (5.2) can be written as $At = 0$, where

$$A = \begin{pmatrix} e_x^0 & e_y^0 \\ e_y^0 & -e_x^0 \end{pmatrix}, \tag{8.1}$$

and we write $\pi(t \mid At = 0)$ for the corresponding conditional distribution, which is also Gaussian with zero mean and covariance matrix

$$\widetilde{\Sigma}_t = Q_t^{-1} \left(I - A^T \left(A Q_t^{-1} A^T \right)^{-1} A Q_t^{-1} \right). \tag{8.2}$$

As noted in Section 3.3, in some images there are large variations in signal magnitude along the outline, due to the way the ultrasound beam is reflected. To avoid bias in estimates of the cross sectional area, it is important that the model is robust with respect to such image features. One simple solution would be to penalise interior angles that differ too much from $(m-2)\pi/m$, the interior angle of a circular polygon. This is easy to implement, but is somewhat ad-hoc, and works best when the contour is known to be close to a circle. Another approach is to modify or ignore the likelihood on the parts where data is spurious. Rue and Husby (1998) uses a binary Markov field ξ, called a destructive deformation field, living on the edges of the template, and having value 1 if data is missing on the edge, and 0 otherwise. The field is given a simple prior

$$\pi(\xi) \propto \exp\left(\beta_0 \sum_{k=0}^{m-1} \mathbb{I}_{[\xi_k = 1]} + \beta_1 \sum_{k=0}^{m-1} \mathbb{I}_{[\xi_k = \xi_{k+1}]} \right),$$

with subscript addition taken modulo m. On the parts of the contour where the destructive deformation field has value unity, the contour shape is essentially determined by the prior. The method proved to be well suited for modelling partly destroyed edges.

In this report we will try to model *how* specular reflection occur. As mentioned in Section 3.3, the ultrasound wave is partially reflected at interfaces between tissue types with different acoustic impedances, and this specular reflection may at least partly explain the weak echo at some parts of the artery wall. We model the total radio frequency signal as

$$x_s + \rho r_s, \quad s \in S,$$

where x_s is the diffuse component, r_s is the specular component, and the constant ρ is the relative magnitude of the specular and diffuse components. ρ can be estimated from the generalised spectrum of the data (Varghese, Donohue and Chatterjee, 1995), but this is computer intensive and we find it easier to treat it as an unknown parameter. where ρ is a scaling parameter.The specular component is non-zero at tissue interfaces only, where it follows the law

$$r_s = \frac{\frac{Z_2}{\cos\theta_2} - \frac{Z_1}{\cos\theta_1}}{\frac{Z_2}{\cos\theta_2} + \frac{Z_1}{\cos\theta_1}} \cos^\gamma \theta_1, \tag{8.3}$$

where we set the reflection exponent to $\gamma = 2$ (Hokland and Kelly, 1996), and use the acoustic impedances $Z = 162.0\ \mathrm{kg/(s \cdot cm^2)}$ for muscle and $Z = 166.6\ \mathrm{kg/(s \cdot cm^2)}$ for blood (Christensen, 1988).

Finally we note that it is not a contradiction to use piecewise linear curves to model a general continuous curve. The polygons are just a means of representing the (intractable) class of closed curves, and we can safely use polygons as approximations of the "true" curves as long as integrals calculated over the approximate model are close to integrals obtained from the true model. This point is made in Andreev and Arjas (1996) and Arjas (1996) who consider histogram approximations of probability densities. In fact, Andreev and Arjas (1996) show that if $\pi(\cdot)$ is a (prior) distribution on the space of density functions, then there exists a sequence $\{\pi_n(\cdot)\}_n$ of distributions on the space of piecewise constant densities, converging weakly to π, such that the corresponding posterior distributions $\pi_n(\cdot\,|\,y)$ converge weakly to the posterior $\pi(\cdot\,|\,y)$ corresponding to π.

8.2.2 Prior distribution for the log-variance field

Recall from Section 3.2 that under the diffuse scattering model, the radio frequency image consists just of zero mean Gaussian noise with a variance parameter determined by the acoustical properties of the imaged tissue. Thus we define a log-variance field ν that is piecewise smooth with homogeneous patches corresponding to the different tissue regions. In Chapter 7 we used an edge preserving smoothing prior for modelling ν, but in the present situation we use an *explicit* model for the tissue interface, hence there is no need for an implicit edge model such as the Geman & Yang model Eq. (4.10). The deformed template partitions the imaged region into two parts corresponding to the in- and outside of the artery wall, and it is natural to assume that this partitioning also corresponds well to the underlying subdivision of the log-variance field ν. Thus we can define two such fields, ν_0 and ν_1 corresponding to the outside and inside, respectively.

In the blood vessel images, the interior of vessel wall contains blood, which is certainly homogeneous and well represented by the diffuse scattering model in Section 3.2. For simplicity we assume the muscle tissue surrounding the vessel to be homogeneous as well, but with different scattering characteristics. Because of the homogeneity we believe the log-variance fields to be quite smooth, and we model them as Gaussian fields with exponential correlation functions, ie. we let $\nu_l \sim \mathrm{N}(\mu_l, \Sigma_l)$, where

$$\Sigma_{rs} = \sigma_l^2 \exp\left(-3\|r-s\|/\varrho_l\right), \quad l = 0, 1, \tag{8.4}$$

where ϱ_l is the range. Although Gaussian fields are natural candidates for smooth log-variance fields, they are not computationally convenient due to full matrices giving complexity of order $\mathcal{O}(n^3)$ for vital operations. We make use of the results in Rue and Tjelmeland (2001) to fit a GMRF with a covariance function very close to the exponential. The GMRF has only a 5×5 neighbourhood around each site, thus the computational complexity for the fitted GMRF is only $\mathcal{O}(n^2)$(Rue, 2001), giving a speed-up of $\mathcal{O}(n)$. The prior density for ν_l is

$$\pi(\nu_l \mid \tau_l, \mu_l) \propto \exp\left(-\frac{\tau_l}{2}\left(\nu_l - \mu_l 1_n\right)^T Q_l \left(\nu_l - \mu_l 1_n\right)\right), \quad l = 0, 1, \tag{8.5}$$

where $1_n = (1, \dots, 1)^T$, and μ_l and τ_l are unknown mean and precision parameters. The parameters μ_0, μ_1 are given uniform distributions on appropriate ranges $[L_j, U_j]$, $l = 0, 1$, while the precision parameters τ_0, τ_1 are given vague $\Gamma(c_l, d_l)$ hyperpriors. For simplicity we will in the following assume $Q_0 = Q_1 = Q_\nu$.

The surfaces ν_0 and ν_1 are only observed within their respective subsets of the image domain. Let $\mathcal{T}_t \subset S$ be the vertices enclosed by the template deformed by the transformation vector t, then the diffuse scattering signal x has conditional density

$$\begin{aligned}\pi(x \mid \nu_0, \nu_1, t) \propto \exp\Bigg(&-\frac{1}{2}\sum_{s \in \mathcal{T}_t \cap S}\left(x_s^2 \exp(-2\nu_{1,s}) + 2\nu_{1,s}\right)\\ &-\frac{1}{2}\sum_{s \in \mathcal{T}_t^{\complement} \cap S}\left(x_s^2 \exp(-2\nu_{0,s}) + 2\nu_{0,s}\right)\Bigg).\end{aligned} \tag{8.6}$$

This is similar to the somewhat simpler model used in Qian et al. (1996). For simplicity we will sometimes use the notation $\nu \equiv (\nu_0, \nu_1)$.

8.2.3 Observation model

As is in Chapter 7, we assume the data y_s at site s to be the superposition of signals from a small neighbourhood around s, and hence we model y as a convolution of $x + \rho r$ with the imaging system point spread function h, with additive independent zero mean Gaussian noise having precision λ:

$$\pi(y \mid x, r, \lambda) \propto \lambda^{n/2} \exp\left(-\frac{\lambda}{2}\sum_{s \in S}\left(y_s - \sum_k h_k(x_{s+k} + \rho r_{s+k})\right)^2\right), \tag{8.7}$$

We assign vague hyperpriors $\Gamma(c_\lambda, d_\lambda)$ and $\Gamma(c_\rho, d_\rho)$ to λ and ρ, respectively.

8.2.4 Benchmark likelihood

For reference we present the likelihood model usually used together with shape models in ultrasonic imaging (Glasbey, 1998; Lefebvre, Berger and Laugier, 1998; Mikic, Krucinski and Thomas, 1998; Hansen et al., 2000). This model does not take into account the physics of the imaging process, instead using the intensity value of the pixels directly. The pixel values are assumed to be independent and Gaussian distributed with mean and precision being location dependent, ie.

$$y_s \mid \tilde{\mu}_0, \tilde{\mu}_1, \tilde{\tau}_0, \tilde{\tau}_1 \sim \begin{cases} \mathrm{N}\left(\tilde{\mu}_0, \tilde{\tau}_0^{-1}\right), & s \in \mathcal{T}_t^{\complement} \cap S, \\ \mathrm{N}\left(\tilde{\mu}_1, \tilde{\tau}_1^{-1}\right), & s \in \mathcal{T}_t \cap S, \end{cases} \tag{8.8}$$

where the parameters $\tilde{\mu} = (\tilde{\mu}_0, \tilde{\mu}_1)$ and $\tilde{\tau} = (\tilde{\tau}_0, \tilde{\tau}_1)$ are given hyperpriors as above. For later reference we note that the likelihood density can be written as

$$\pi(y \mid t, \tilde{\mu}, \tilde{\tau}) \propto \tilde{\tau}_0^{n_0/2} \tilde{\tau}_1^{n_1/2} \exp\left(-\frac{1}{2} \sum_{l=0}^{1} \tilde{\tau}_l \left(S_l + n_l \left(\overline{y}_l - \tilde{\mu}_l\right)^2\right)\right), \tag{8.9}$$

where

$$n_1 = |\mathcal{T}_t \cap S|, \quad \overline{y}_1 = \frac{1}{n_1} \sum_{s \in \mathcal{T}_t \cap S} y_s^2, \quad S_1 = \sum_{s \in \mathcal{T}_t \cap S} \left(y_s - \overline{y}_1\right)^2,$$

and n_0, $\overline{y}_0$, and S_0 are defined similarly.

8.3 Sampling of the Posterior Distribution

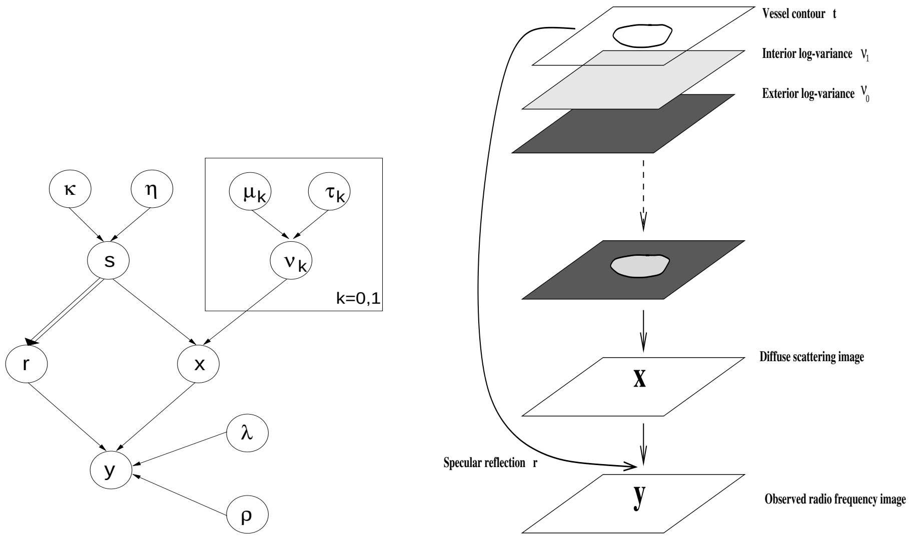
8.3.1 Introduction

A graphical representation of the model is shown in Figure 8.1. The full posterior distribution for all unknown parameters conditioned on the observed data has density

$$
\begin{aligned}
&\pi(x, \nu_0, \nu_1, t, \kappa, \eta, \mu_0, \mu_1, \tau_0, \tau_1, \lambda, \rho \mid y) \\
&\qquad \propto \exp\left(-\frac{\lambda}{2}\sum_{s\in S}\left(y_s - \sum_k h_k(x_{s+k} + \rho r_{s+k})\right)^2\right) \\
&\qquad \times \exp\left(-\sum_{s\in \mathcal{T}_t \cap S}\left(\frac{1}{2}x_s^2 \exp(-2\nu_{1,s}) + \nu_{1,s}\right)\right) \\
&\qquad \times \exp\left(-\sum_{s\in \mathcal{T}_t^{\complement} \cap S}\left(\frac{1}{2}x_s^2 \exp(-2\nu_{0,s}) + \nu_{0,s}\right)\right) \\
&\qquad \times \exp\left(-\sum_{l=0}^{1}\frac{1}{2}\tau_l \left(\nu_l - \mu_l 1_n\right)^T Q_l \left(\nu_l - \mu_l 1_n\right)\right) \\
&\qquad \times \left|\widetilde{\Sigma_t}^{-}\right|^{m/2} \exp\left(-\frac{1}{2}t^T \widetilde{\Sigma_t}^{-} t\right) \mathbb{I}_{[t \text{ is simple}]} \\
&\qquad \times \lambda^{n/2+c_\lambda-1}\rho^{c_\rho-1}\exp(-d_\lambda\lambda - d_\rho\rho) \\
&\qquad \times \prod_{l=0}^{1}\tau_l^{n/2+c_l-1}\exp(-d_l\tau_l)\mathbb{I}_{[L_l<\mu_l<U_l]} \\
&\qquad \times \kappa^{a_\kappa-1}\exp(-b_\kappa\kappa) \times \eta^{a_\eta-1}\exp(-b_\eta\eta).
\end{aligned}
\tag{8.10}
$$

This distribution is clearly analytically intractable, and so we have to resort to Markov chain Monte Carlo (McMC) methods to be able to do inference. Below we will briefly review the commonly used Metropolis-Hastings (MH) and Langevin-Hastings (LH) algorithms; for details and references we refer to Robert and Casella (1999).

The most common method is to update one parameter at a time in either a systematic or random scan, but it is well known that such algorithms have poor convergence and mixing properties compared to algorithms that update parameters in blocks (Carter and Kohn, 1994; Liu, Wong and Kong, 1994). In fact, Knorr-Held and Rue (2002) argue that updating all or nearly all parameters in one block may be necessary to get sufficient mixing and avoid estimation bias. However, such full block algorithms may be difficult to construct, and may be heavily computer intensive. Thus it is important to be careful when designing the sampling algorithms, while at the same time keeping and eye on the efficiency of the proposed algorithm. We will in the following discuss different algorithms



(a) (b)

Figure 8.1: Graphical representations of the contour detection model. The double arrow in (a) denotes a deterministic relationship.

for sampling from the posterior distribution Eq. (8.10), and try to compare their relative merits with respect to mixing and efficiency.

Metropolis-Hastings algorithm The Metropolis-Hastings algorithm proceeds as follows for a posterior distribution $\pi(x \mid y)$. Given the current state x we generate a new state x' from a proposal distribution $q(x' \mid x)$. The new state is accepted with probability

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x' \mid y)q(x \mid x')}{\pi(x \mid y)q(x' \mid x)}\right\},$$

otherwise the old state is retained.

We will in particular consider two special cases. If $q(x' \mid x)$ is a distribution with mean x we call the algorithm Random walk Metropolis-Hastings. On the other hand, if the proposal distribution q is independent of the current state x, the algorithm is called the Independence Metropolis-Hastings algorithm.

Langevin-Hastings algorithm An algorithm similar in spirit is obtained by using gradient information in the proposal distribution. If x is the current state, a new proposal is generated from a Gaussian distribution with mean

$x + \frac{h}{2}\nabla \ln \pi(x \mid y)$ and variance h. The acceptance probability becomes

$$\alpha(x, x') = \min \left\{ 1, \frac{\pi(x' \mid y) \exp\left(-\frac{1}{2h}\|x - x' - \frac{h}{2}\nabla \ln \pi(x' \mid y)\|^2\right)}{\pi(x \mid y) \exp\left(-\frac{1}{2h}\|x' - x - \frac{h}{2}\nabla \ln \pi(x \mid y)\|^2\right)} \right\}.$$

It is shown in Roberts and Rosenthal (1998) that the number of iterations required to obtain convergence is $\mathcal{O}(n^{-1/3})$, compared with $\mathcal{O}(n^{-1})$ for the random walk Metropolis-Hastings algorithm, so the Langevin-Hastings algorithm may be preferable, for high-dimensional problems. Note however that the gradient of the log posterior density should be readily available, a condition that is not always fulfilled.

8.3.2 Updating the edge transformation vector t

The most common way of sampling a deformable template model is to use the random walk Metropolis-Hastings algorithm. A change in shape is proposed by moving one or more vertices, and one then calculates the corresponding transformation vector, and accepts or rejects the move using the Metropolis-Hastings ratio. Other types of moves might be translation, rotation, or scaling of the whole template, or reordering of the vertices. See eg. Rue and Hurn (1999) for details on the different move types.

However, other algorithms are possible, for instance, Grenander and Miller (1994) use a discrete approximation of the Langevin diffusion to sample the template parameters. The advantage is that the use of data information might help to generate more efficient moves. On the other hand, the incremental construction of the algorithm makes it difficult to construct the large jumps that sometimes can be very efficient for escaping local minima.

For the reminder of this section we will discuss sampling of t using the Langevin-Hastings algorithm.

Updating the template using Langevin-Hastings The discrete Langevin equation for the posterior distribution $\pi(t \mid At = 0, x, \nu, \kappa, \eta)$ is given by

$$t' = t + \frac{h}{2}\nabla \ln \pi\left(t \mid At = 0, x, \nu, \kappa, \eta\right) + \sqrt{h}\,\vartheta_{2m},$$

where h is a scaling parameter, and ϑ_{2m} consists of $2m$ independent standard normal variables. This equation needs to be defined on the null-space of A, and we have therefore found it easier to reparameterize the equation as in Grenander and Miller (1994). Moreover, Møller, Syversveen and Waagepetersen (1998) argue in a similar problem that the reparameterisation given below often leads to better mixing, possibly because we remove the correlations between the components in the proposal.

The reparametrisation is done as follows: For $k = 0, \ldots, m-1$, define

$$\begin{aligned} c_k &= (\cos 2\pi k/m, \ldots, \cos 2\pi k(m-1)/m) \\ s_k &= (\sin 2\pi k/m, \ldots, \sin 2\pi k(m-1)/m), \end{aligned}$$

and let $\gamma_0 = (1 \quad 0)^T$ and $\gamma_1 = (0 \quad 1)^T$. Let F be a matrix having columns

$$f_k = \gamma_0 \otimes c_k + \gamma_1 \otimes s_k \qquad f_{k+m} = -\gamma_0 \otimes s_k + \gamma_1 \otimes c_k, \quad k = 0, \ldots, m-1. \tag{8.11}$$

Then the matrix F diagonalises the precision matrix Q_t, ie. $F^T Q_t F = \Lambda$, where Λ is a diagonal matrix containing the eigenvalues $\{\lambda_k\}$ of Q_t:

$$\lambda_k = \lambda_{k+m} = \frac{\kappa}{m} + 6\eta m^3 - 8\eta m^3 \cos\frac{2\pi k}{m} + 2\eta m^3 \cos\frac{4\pi k}{m}, \quad k = 0, \ldots, m-1. \tag{8.12}$$

To enforce the closure constraint $At = 0$, note that the row space of A (Eq. (8.1)) is spanned by the vectors $\gamma_0 \otimes c_1 + \gamma_1 \otimes s_1$ and $\gamma_0 \otimes s_1 - \gamma_1 \otimes c_1$, which are exactly the columns f_1 and f_{2m-1} of F. Thus closure is enforced by setting $\lambda_1 = \lambda_{2m-1} = 0$, and we can draw a sample t from the conditional distribution $\pi(t \mid At = 0)$ by setting $t = FDz$, where z consist of $2m-2$ independent standard normal variables, and

$$D = \begin{pmatrix} 1/\sqrt{\lambda_0} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 1/\sqrt{\lambda_2} & \ldots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 1/\sqrt{\lambda_{2m-2}} \\ 0 & 0 & \ldots & 0 \end{pmatrix}$$

Thus posterior simulations from the conditional distribution $\pi(t \mid At = 0, x, \nu, \kappa, \eta)$ can be obtained by transforming samples from the conditional distribution $\pi(z \mid x, \nu, \kappa, \eta)$.

To simplify the discussion, write $T(\kappa, \eta) = FD$, and assume for the moment that κ and η are fixed. The Langevin-Hastings algorithm for z is then as follows. Set

$$z' = z - \frac{h}{2} z + \frac{h}{2} T(\kappa, \eta)^T \nabla_t \ln \pi\,(x \mid t, \nu) + \sqrt{h}\vartheta_{2n-2} \tag{8.13}$$

$$t' = T(\kappa, \eta) z', \tag{8.14}$$

and accept the proposed state z' with probability

$$\alpha(z, z') = \min\left\{1, \frac{\pi(z')q(z \mid z')\pi(x \mid t', \nu)}{\pi(z)q(z' \mid z)\pi(x \mid t, \nu)}\right\}, \tag{8.15}$$

where $q(z' \mid z)$ is the density defined by Eq. (8.13). This will generate an aperiodic and irreducible Markov chain with the posterior as its stationary distribution.

Note that the algorithm above keeps the position c of the vertex v_0 fixed. To ensure proper mixing we let $(d_0, d_1) : c \mapsto c + (d_1\ d_0)^T$ define a translation of c. The parameters $(d_0, d_1) = d$ are updated using the Langevin equation $d' = d + (h/2)\nabla_d \ln \pi(x|t, \nu) + \sqrt{h}\vartheta_2$. For notational convenience we suppress the use of d in the following discussion.

Concerning the gradient of the likelihood, Grenander and Miller (1994) states a result in the case of a continuous curve and continuously observed data. The following result is similar, but with the number of edges fixed. As a matter of notation, we parameterise each edge by τ so that $e_j(\tau) = v_j + \tau(v_{j+1} - v_j)$, $\tau \in [0, 1]$.

Proposition 8

For the model defined in Eq. (8.6), let $H_l(x_i) = (1/2)\, x_i^2 \exp(-2\nu_{l,i}) + \nu_{l,i}$, $l = 0, 1$. Then for $k = 0, \ldots, m-1$ and $l = 0, 1$,

$$\frac{\partial}{\partial t_{l,k}} \ln \pi\,(x \mid t, \nu) = \sum_{j=k+1}^{m-1} A_j(t_{l,k}) \int_0^1 \left(H_1\left(e_j(\tau)\right) - H_0\left(e_j(\tau)\right)\right) d\tau \tag{8.16}$$

with the Jacobians

$$A_j(t_{0,k}) = \begin{vmatrix} e^0_{k,x} & e_{j,x} \\ e^0_{k,y} & e_{j,y} \end{vmatrix} \qquad A_j(t_{1,k}) = \begin{vmatrix} -e^0_{k,y} & e_{j,x} \\ e^0_{k,x} & e_{j,y} \end{vmatrix}.$$

Furthermore, for $l = 0, 1$,

$$\frac{\partial}{\partial d_l} \ln \pi\,(x \mid t, \nu) = \sum_{j=0}^{m-1} J_j(d_l) \int_0^1 \left(H_1\left(e_j(\tau)\right) - H_0\left(e_j(\tau)\right)\right) d\tau$$

where $J_j(d_0) = e_{j,y}$ and $J_j(d_1) = -e_{j,x}$.

Proof. We sketch a proof. As before, let e^0 be the template, and e be the template deformed by t, having vertices v and edges e. Fixing the first vertex v_0, the others are given by

$$v_k = v_0 + \sum_{j=0}^{k-1} e_j = v_0 + \sum_{j=0}^{k-1} \begin{pmatrix} 1 + t_{0,j} & -t_{1,j} \\ t_{1,j} & 1 + t_{0,j} \end{pmatrix} \begin{pmatrix} e^0_{j,x} \\ e^0_{j,y} \end{pmatrix}. \tag{8.17}$$

For notational convenience, denote the log-posterior $\ln \pi(x \mid t, \nu)$ by $f(t)$. For a small perturbation t_ϵ of t, the difference $f(t_\epsilon) - f(t)$ must be computed over the m parallelograms with edges e_k and $\nabla_t v_k$. Using Eq. (8.17), the gradients are given by

$$\frac{\partial}{\partial t_{0,j}} v_k = \begin{pmatrix} e^0_{j,x} \\ e^0_{j,y} \end{pmatrix} \mathbb{I}_{[k>j]} \qquad \frac{\partial}{\partial t_{1,j}} v_k = \begin{pmatrix} -e^0_{j,y} \\ e^0_{j,x} \end{pmatrix} \mathbb{I}_{[k>j]}$$

$$\frac{\partial}{\partial d_0} v_k = \begin{pmatrix} d_0 \\ 0 \end{pmatrix} \qquad \frac{\partial}{\partial d_1} v_k = \begin{pmatrix} 0 \\ d_1 \end{pmatrix},$$

and the expression for the Jacobians follows. $\square$

As mentioned above, in Section 8.3, it is often desirable to update a variable jointly with its hyperparameters. In the case of the edge transformation vector t, this can be done as follows: We propose new values for κ and η by scaling the current values by numbers drawn from the distribution $U[1/f, f]$, for some $f > 1$. We then set

$$z' = z - \frac{h}{2} z + \frac{h}{2} T(\kappa, \eta)^T \nabla_t \ln \pi\left(x \mid t, \nu\right) + \sqrt{h}\vartheta_{2n-2} \tag{8.18}$$

$$t' = T(\kappa', \eta') z', \tag{8.19}$$

and accept the proposed state (z', κ', η') with probability

$$\alpha(z', \kappa', \eta' \mid z, \kappa, \eta) = \min\left\{1, \frac{\pi(z')\, q(z \mid z', \kappa', \eta') \pi\left(x \mid t', \nu\right)}{\pi(z)\, q(z' \mid z, \kappa, \eta)\, \pi\left(x \mid t, \nu\right)} \frac{\pi(\kappa')\, \pi(\eta')\, \kappa\eta}{\pi(\kappa)\, \pi(\eta)\, \kappa'\eta'}\right\}, \tag{8.20}$$

where $q(z' \mid z, \kappa, \eta)$ is the density defined by Eq. (8.18). Note that the normalising constants still cancel even if the transformation matrix T changes.

8.3.3 Updating the diffuse scattering field x

The full conditional distribution for the diffuse scattering parameter x is given by

$$\pi(x \mid y, \nu, t, \lambda, \rho) \propto \exp\left(-\frac{1}{2} x^T \left(\lambda H^T H + V\right) x + \lambda x^T \left(Hy - \rho H^T H r\right)\right), \tag{8.21}$$

where V is a diagonal matrix with elements

$$V_{ss} = \exp\left(-2\nu_{0,s}\right) \mathbb{I}_{[s \in \mathcal{T}_t^{\complement} \cap S]} + \exp\left(-2\nu_{1,s}\right) \mathbb{I}_{[s \in \mathcal{T}_t \cap S]}.$$

Thus x given all other parameters is a GMRF which can be Gibbs sampled using the methods in Rue (2001).

The hyperparameters λ and ρ can be drawn from their full conditionals

$$\lambda \mid \ldots \sim \Gamma\left(c_\lambda + \frac{n}{2},\, d_\lambda + \frac{1}{2}\|y - H(x+\rho r)\|^2\right)$$
$$\rho \mid \ldots \sim \Gamma\left(c_\rho,\, d_\rho + (y - Hx)^T\, Hr\right)$$

Alternatively, all three parameters can be updated jointly. We then propose new values for the hyperparameters by scaling the current values by random numbers uniformly distributed on the interval $[1/f, f]$ for some $f > 1$. A proposal for x is then draw from the distribution Eq. (8.21) conditional on the new values λ' and ρ'. The new set (x', λ', ρ') of parameters is accepted with probability

$$\begin{aligned} \alpha(x', \lambda', \rho' \mid x, \lambda, \rho) &= \min\left\{1, \frac{\pi(x', \lambda', \rho' \mid \ldots)\pi(x \mid \lambda, \rho, \ldots)\, q(\lambda \mid \lambda')q(\rho \mid \rho')}{\pi(x, \lambda, \rho \mid \ldots)\pi(x' \mid \lambda', \rho', \ldots)\, q(\lambda' \mid \lambda)q(\rho' \mid \rho)}\right\} \\ &= \min\left\{1, \frac{\pi(\lambda')\pi(\rho')}{\pi(\lambda)\pi(\rho)}\frac{\lambda\rho}{\lambda'\rho'}\right\}, \end{aligned} \tag{8.22}$$

otherwise the old state (x, λ, ρ) is retained.

8.3.4 *Updating the log-variance field ν*

The conditional distribution for the log-variance field ν_l is given by

$$\pi\left(\nu_l \mid x, t, \tau_l, \mu_l\right) \propto \exp\left(-\frac{1}{2}\tau_l\left(\nu_l - \mu_l 1_n\right)^T Q_\nu \left(\nu_l - \mu_l 1_n\right) - \sum_{s \in \mathcal{S}_l} f\left(\nu_{l,s}, x_s\right)\right), \tag{8.23}$$

where

$$\mathcal{S}_0 = \mathcal{T}^{\complement} \cap S, \quad \mathcal{S}_1 = \mathcal{T} \cap S, \quad \text{and} \quad f(\nu, x) = \frac{1}{2}x^2 e^{-2\nu} + \nu.$$

It is not possible to sample directly from this distribution, so we use either a Random walk Metropolis-Hastings algorithm with a Gaussian proposal distribution with covariance matrix $\sigma_\nu^2 I_n$, or a Langevin-Hastings algorithm.

A third option, using a Gaussian approximation of the density Eq. (8.23) as a proposal distribution in a Metropolis-Hastings algorithm did not work very well. This idea has been used in Rue (2001) and Knorr-Held and Rue (2002), but we suspect that the distribution in Eq. (8.23) is to skewed in some directions to be well enough approximated by a Gaussian distribution. This might be due to the large variations in the magnitude of the x_i's, see the discussion in Section 7.4.2.

The hyperparameters τ_l are either drawn from their full conditionals

$$\tau_l \mid \ldots \sim \Gamma\left(a_l + \frac{n}{2},\, b_l + \frac{1}{2}\left(\nu_l - \mu_l 1_n\right)^T Q_\nu \left(\nu_l - \mu_l 1_n\right)\right),$$

or updated jointly with μ_l and ν_l, in which the new state $(\nu_l', \tau_l', \mu_l')$ is accepted with probability

$$\alpha(\nu_l', \tau_l', \mu_l' \mid \nu_l, \tau_l, \mu_l) = \min\left\{1, \frac{\pi(\nu_l' \mid x, t, \tau_l', \mu_l')\pi(\tau_l')\tau_l}{\pi(\nu_l \mid x, t, \tau_l, \mu_l)\pi(\tau_l)\tau_l'}\right\}. \tag{8.24}$$

8.3.5 Sampling schemes

We have implemented three schemes for sampling from the posterior distribution (8.10), with the intention of investigating whether the choice of sampling scheme can affect the robustness and performance of our proposed method. In Scheme 1 each of the parameters t, x, ν_0 and ν_1 are updated jointly with their hyperparameters as described above in Section 8.3.2, Section 8.3.3, and Section 8.3.4, respectively. The parameters were updated with only one hyperparameter at a time; this enables us to tune the scaling f to obtain acceptance rates around 25%.

Scheme 2 differs from Scheme 1 in that we update the template t and the diffuse scattering field x in one block jointly with one of the hyperparameters κ, η, λ, and ρ in turn. The parameters ν_0 and ν_1 are updated as in Scheme 1. Finally, in Scheme 3 *all* parameters t, x, ν_0 and ν_1 are updated jointly in a single block, again jointly with one of the hyperparameters in turn. The purpose of the last two schemes is to reduce the effect of the possibly strong correlations between the parameters. Grouping the parameters also simplifies the computation of acceptance rates, hence for an image having 64×128 pixels an 550 MHz PC used approximately 0.18 seconds per iteration for Scheme 1, and 0.16 seconds per iteration for Scheme 2 and Scheme 3.

8.4 Experiments

In this section we evaluate the performance and robustness of the model from Section 8.3 by applying it to both simulated and real data sets. Of course, for the simulated image the diffuse scattering model holds by construction, but the experiments are useful for testing different sampling strategies, and to get a feel for their efficiency and robustness. The real data sets are ultrasound images of cross sections through the carotid artery.

8.4.1 Contour detection in simulated ultrasound images

In the first section we present some preliminary experiments performed on the simulated ultrasound image in Figure 7.4. The experiments include comparison between the random walk and Langevin-Hastings algorithms for sampling the deformation parameter t and the log-variance field ν.

Sampling of the deformable template Our first experiment considers sampling of the edge transformation vector t. The deformable template model and the posterior distribution of the edge transformation vector are described in Section 8.2.1, and the different sampling strategies are presented in Section 8.3. We want to compare the random walk and Langevin-Hastings algorithms for sampling of the template t, and have kept all other variables fixed to exclude other sources of variation.

The table below shows results from 20000 iterations of the random walk (RW) and Langevin-Hastings (LH) algorithms, with hyperparameters updated jointly with the transformation vector. We have used a burn in of 3000 iterations, after which every 10th sample is stored.

Table 8.1: Summary results based on 20000 iterations from the posterior distribution for the edge transformation vector t. The estimated autocorrelations $\mathrm{A}(k)$ are based on every 10th sample of the Markov chain, and the empirical variance EV is based on the initial positive sequence estimator (Geyer, 1996). α denotes the acceptance rate.

		Mean	Sdv	A(1)	A(5)	A(10)	EV	α
η	RW	$.67 \cdot 10^{-4}$	$.48 \cdot 10^{-4}$	.914	.683	.542	10^{-6}	.310
η	LH	$1.47 \cdot 10^{-4}$	$.66 \cdot 10^{-4}$	.994	.971	.942	10^{-6}	.131
κ	RW	89.4	6.78	.444	.078	.074	55.5	.175
κ	LH	100.2	6.85	.997	.988	.974	11785	.134
area	RW	628	4.11	.943	.775	.625	616.8	.243
area	LH	622	5.41	.978	.938	.895	1956.0	.133

The results indicate that the random walk algorithm is superior to the Langevin-Hastings algorithm for this problem. We believe that the reason might be found in the implementation of the Langevin algorithm. One needs to evaluate the likelihood gradient (8.16), and for the gradient to be continuous we need a differentiable interpolation of the data over the image domain. We have chosen to use the data without interpolation, and this might lead to discontinuities in the gradient causing the sampler to get stuck in local minima. However, for this simple object model the random walk algorithm seems to be very efficient, and much simpler to implement. Also note that the Langevin algorithm is most useful for high-dimensional parameter vectors, see Roberts and Rosenthal (1998).

Hyperpriors for κ and η We have tried different hyperpriors for κ and η to investigate to what extent hyperpriors affect the result. The experiments were run on the data set in Figure 7.4(d), and the results are based on 15000 iterations after a burn in of 5000 iterations.

We recall that η and κ are given Gamma hyperpriors with parameters ϵ_η, δ_η and ϵ_κ, δ_κ, respectively. An interpretation of the parameters is given in Section 5.3.3, on which we base the choice of hyperparameters. The different choices for ϵ_κ and δ_κ are given in Table 8.2, along with summary statistics from the different Markov chains. The results indicate that our model is stable over the chosen range of hyperparameters; the behaviour of the template does not change even if the posterior distribution for κ changes significantly (see Figure 8.2 for the marginal posteriori distributions for κ). The choice of hyperprior does however seem to have an effect on mixing, as the autocorrelation plots in Figure 8.4 indicate. It is likely that the least informative prior allows the sampler to move more freely, hence we will use the hyperprior $\Gamma(3, 1/30)$ in the subsequent experiments.

Table 8.2: Parameter estimates using different hyper-priors for κ and having $\epsilon_\eta = 1/2$ and $\delta_\eta = 1/2$ fixed. The results are based on 15000 iterations of the random walk sampler after a burn in 0f 5000 iterations.

$(\epsilon_\kappa, \delta_\kappa)$	mean	sd	mean area	CI area	α
(90, 1)	91.8	9.05	631	(623, 643)	.274
(30, 1/3)	91.4	14.8	626	(617, 635)	.301
(9, 1/9)	77.8	21.6	629	(620, 639)	.215
(3, 1/30)	100.9	30.6	627	(617, 636)	.291

Table 8.3: Parameter estimates using different hyper-priors for η and having $\epsilon_\kappa = 3$ and $\delta_\kappa = 1/30$ fixed. The results are based on 15000 iterations of the random walk sampler.

$(\epsilon_\eta, \delta_\eta)$	mean	sd	mean area	CI area	α
(1/10, 1/10)	$17.7 \cdot 10^{-5}$	$7.0 \cdot 10^{-5}$	629	(618, 640)	.285
(1/2, 1/2)	$4.12 \cdot 10^{-5}$	$2.3 \cdot 10^{-5}$	627	(620, 635)	.247

For η we tried two different sets of hyperparameters, shown in Table 8.3. Again, the results indicate that while the choice of hyperparameters affects the marginal posterior for η, inference about the template itself is not affected. Choosing the $\Gamma(1/2, 1/2)$-prior seems to lead to a better-mixing chain, see the autocorrelation plots in Figure 8.4 (e) and (f), and is thus the one we will use throughout.

Finally we note that there is a limit to how un-informative the hyperpriors for κ and η might be. Especially for real data carrying sparse information about the shape of the imaged object, the template needs to be constrained to avoid

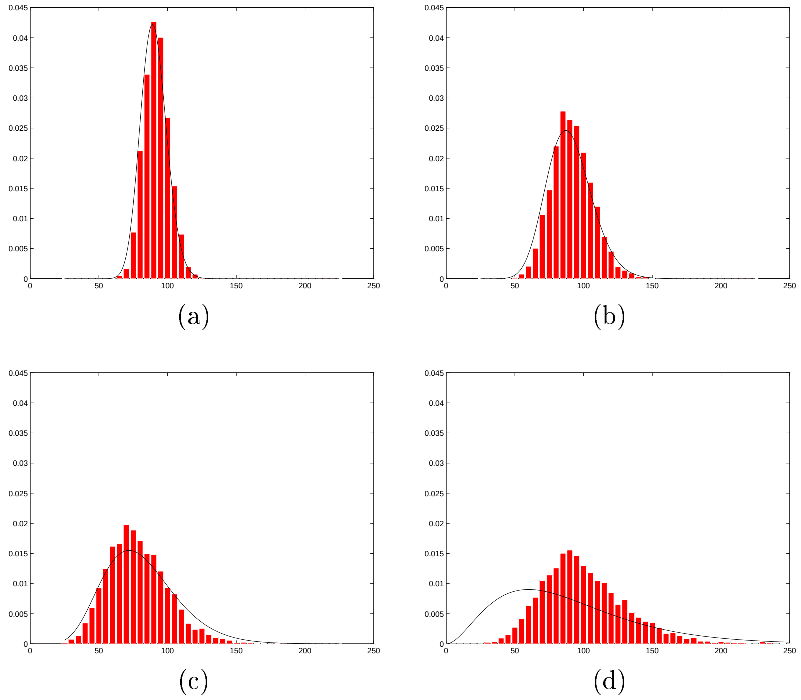


Figure 8.2: Estimated marginal posterior densities for κ. The curves shown are the prior distributions (a) $\Gamma(90, 1)$, (b) $\Gamma(30, 1/3)$, (c) $\Gamma(9, 1/9)$, and (d) $\Gamma(3, 1/30)$.

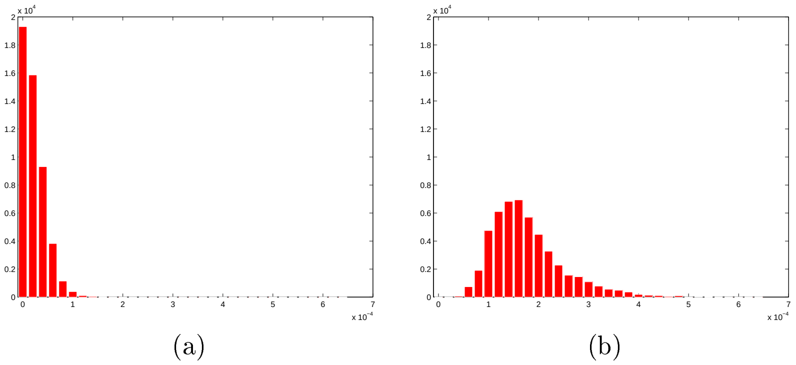


Figure 8.3: Estimated marginal posterior densities for η using prior distributions (a) $\Gamma(1/2, 1/2)$, and (b) $\Gamma(1/10, 1/10)$.

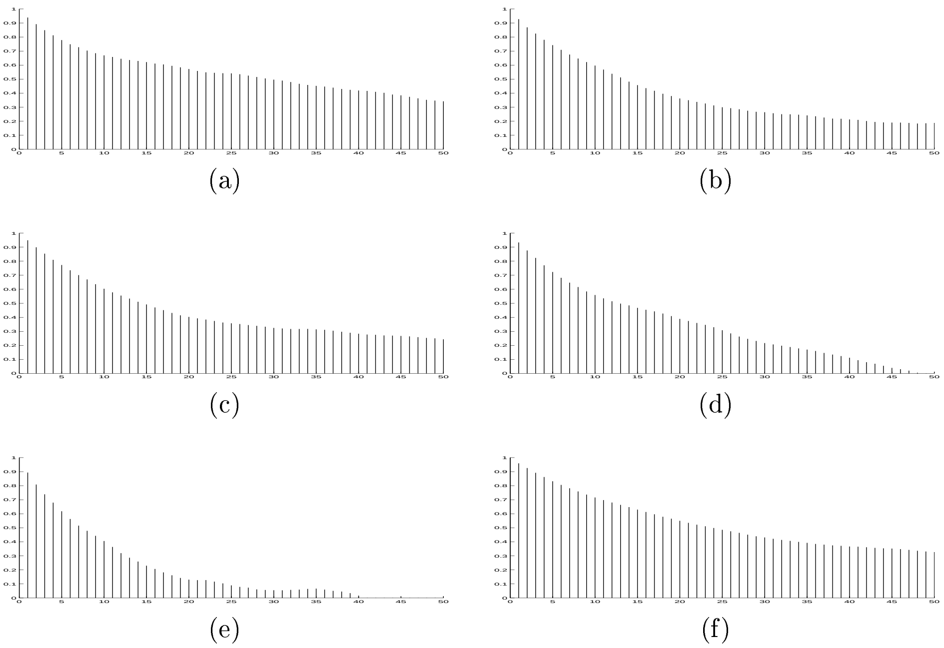


Figure 8.4: Estimated autocorrelations for the area of the object in Figure 7.4, based on every 10th scan of the random walk algorithm, and with different hyper-priors for κ and η. Two upper rows: $\eta \sim \Gamma(1/2, 1/2)$, and κ having priors (a) $\Gamma(90, 1)$, (b) $\Gamma(30, 1/3)$, (c) $\Gamma(9, 1/9)$, and (d) $\Gamma(3, 1/30)$. Bottom row: $\kappa \sim \Gamma(90, 1)$, and (e) $\eta \sim \Gamma(1/2, 1/2)$, (f) $\eta \sim \Gamma(1/10, 1/10)$.

problems. The constraint encode in the hyperpriors of course reflect prior belief that the object is close to having a circular or elliptic shape.

Sampling of the log-variance field This experiment compares random walk (RW) and Langevin-Hastings (LH) sampling for the log-variance fields ν_0 and ν_1. The results are based on 5000 iterations with a burn in of 500 iterations, starting with flat images with all zeros. The hyperparameters where updated jointly with the fields, as described in Section 8.3.4. The summary results in Table 8.4 show that there is not much difference, although the Langevin algorithm seems to give slightly better mixing. It should however be noted that it uses approximately 50% more CPU time, and is thus less efficient. We will still use the Langevin algorithm in the following examples, since mixing is especially important in real applications.

Table 8.4: Summary results based on 5000 iterations from the posterior distribution for the log-variance fields ν_0 and ν_1. The true values are given below in Table 8.5.

		Mean	Sdv	A(1)	A(5)	A(10)	EV	α
μ_0	LH	.435	.0686	.302	.00789	.0299	.000003	.543
μ_0	RW	.437	.0714	.297	.0276	.0045	.000380	.490
μ_1	LH	-.000427	.0552	.269	-.0223	-.0312	.000099	.496
μ_1	RW	.00647	.0581	.400	.0159	.0126	.000190	.471
τ_0	LH	.748	.0163	.413	-.00299	.0344	.00002	.183
τ_0	RW	.734	.0160	.310	.0233	.0121	.00168	.161
τ_1	LH	1.075	.0238	.344	.0423	-.0390	.000024	.172
τ_1	RW	1.070	.0240	.347	.0663	.0283	.000202	.169

Performance of sampling schemes We tried sampling schemes 1 and 3 (Section 8.3.5), and in addition a fourth scheme where all parameters and hyperparameters where updated by themselves in a sequential fashion. The results did not differ much for this data set, but the full block sampler (Scheme 3) seemed to mix faster; see the autocorrelation plots in Figure 8.5. We state the results for this sampler only. Table 8.5 shows estimates for all parameters of the model; they seem to be in good agreement with the true values. The object outline was estimated consistently well for all sampling schemes, and Figure 8.6 shows the distance transform average of the samples collected after a burn-in period.

8.4.2 Contour detection in ultrasound images of the carotid artery

To test the feasibility of the model in real applications, we consider recorded radio frequency ultrasound images of the carotid artery. The images are shown in Figure 1.1(d) and (e), and the primary goal is to obtain interval estimates

Table 8.5: Simulated ultrasound image. Estimates for the area of the object and for the model parameters.

Par.	Mean	Credibility interval	Truth	Initialisation
Area	628	$(598, 658)$	625	470
μ_0	.447	$(.310, .581)$	0.5	0.0
μ_1	$-6.270 \cdot 10^{-4}$	$(-.114, .112)$	0.0	0.0
τ_0	.746	$(.716, .777)$	0.75	1.0
τ_1	1.073	$(1.037, 1.129)$	1.0	1.0
λ	.741	$(.051, 2.724)$	1.0	1.0
κ	182.00	$(52.00, 391.29)$	-	90.0
η	$1.681 \cdot 10^{-3}$	$(.513, 3.531) \cdot 10^{-3}$	-	0.007

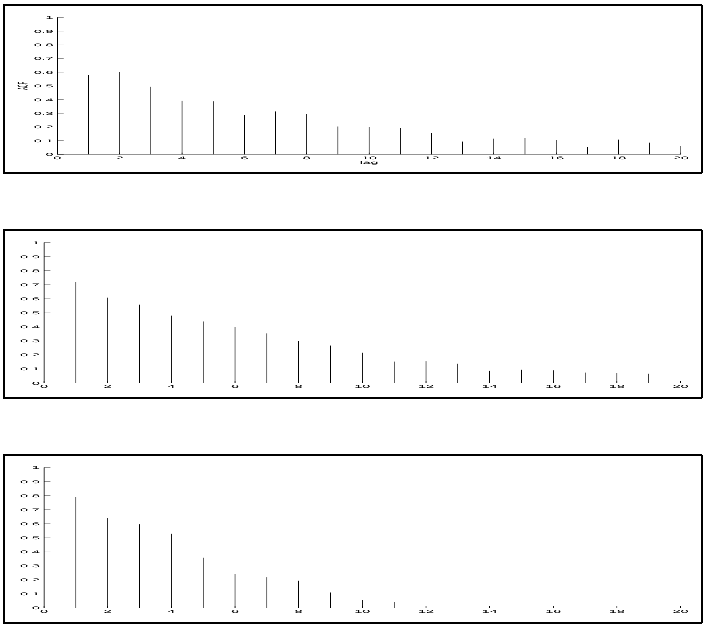
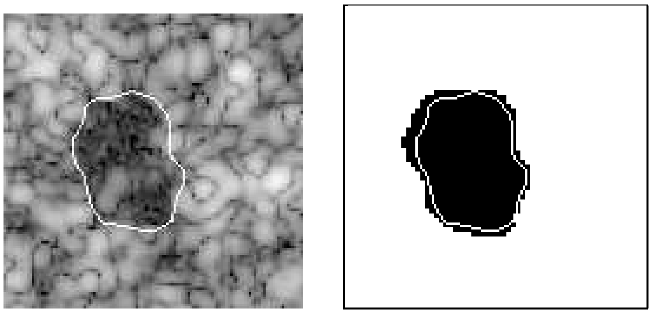


Figure 8.5: Estimated autocorrelations for the cross-sectional vessel area using every 100th sample. From top to bottom: Scheme 4, Scheme 1, and Scheme 3.



(a) (b)

Figure 8.6: Contour estimate. The distance transform estimate based on samples from the posterior distribution shown together with (a) the data, and (b) the underlying truth.

of the cross sectional vessel areas. Compared to the simulated image, the effect of blurring and noise seems to be much more pronounced. The strong reflections at tissue boundaries approximately perpendicular to the incoming pulse are particularly striking, as are the correspondingly weak signal at boundaries parallel to the beam. These effects pose an additional challenge compared to the "ideal" image in Figure 7.4.

We chose to use essentially the same hyper-priors as above, except for κ, which we gave a slightly more informative $\Gamma(9, 1/9)$ prior, reflecting the prior knowledge that the blood vessels have a fairly regular shape, and partly accounting for missing data in the images. The hyperparameter η was given a $\Gamma(1/2, 1/2)$ prior as above, and the precision parameters τ_0 and τ_1 of the log variance fields were given $\Gamma(1, 1)$-priors, while the means were taken to be uniformly distributed on the interval $[-5, 5]$. Finally, we chose a $\Gamma(1, 1)$ prior for the noise precision λ, and a $\Gamma(1, 1/2)$ prior for the specular reflection parameter ρ. Our results are of course conditional on the chosen hyperparameters, but we found the interval estimate for the vessel area to be insensitive to different but sensible choices of the hyperparameters. In all experiments the template was initialised as a circle located in the centre of the image. Better initialisations could of course be found using some pre-processing technique, but in most cases convergence was fast enough for the chosen initialisation to be sufficient.

Performance of the sampling schemes Performance of the different sampling schemes was monitored by inspecting trace plots and autocorrelations for the parameters; especially the cross-sectional area, which is the quantity we are most interested in. The estimated autocorrelation functions for the hyper-parameters decay rapidly, while the autocorrelation function for the cross-sectional area reveal slow convergence for this particular quantity. This is not surprising given the complexity and high dimension of the model. More interestingly, we found the interval estimate to depend on the choice of sampling scheme. Scheme 1, in particular, performed poorly, exhibiting slow convergence and poor mixing. Repeated runs revealed that the Markov chain is prone to get stuck in local minima, or even drifting off. This behaviour is not surprising, as there is considerable interdependency between the diffuse scattering field x and the template t. Thus updating one parameter conditional on the other may cause the chain to move very slowly.

The other two schemes perform better in that their behaviour is consistent over repeated runs, and that the estimated contour seems reasonable. However, inspection of autocorrelation plots reveal that Scheme 2 mixes better than Scheme 3, where all parameters are updated jointly, see Figure 8.7. In the last scheme the number of parameters to update is more than doubled and the dependency between the ν-fields and the other fields does not seem to be strong enough to counterbalance the increase in complexity. In conclusion, blocking x and t together seems to give a huge improvement in sampling robustness, while blocking *all* variables does not seem to give an additional gain. The estimated autocorrelation function for the Scheme 2 is shown in Figure 8.7, indicating reasonable good mixing of the chain.

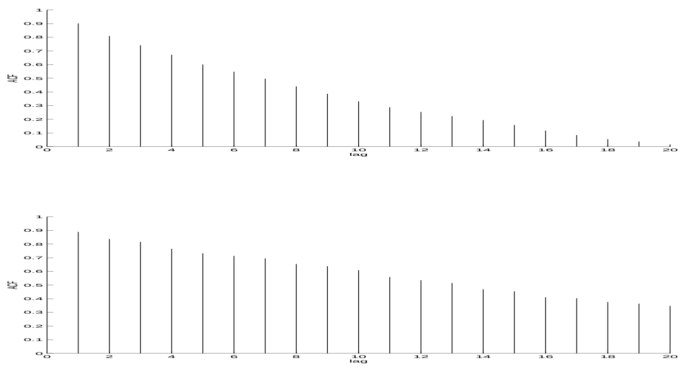
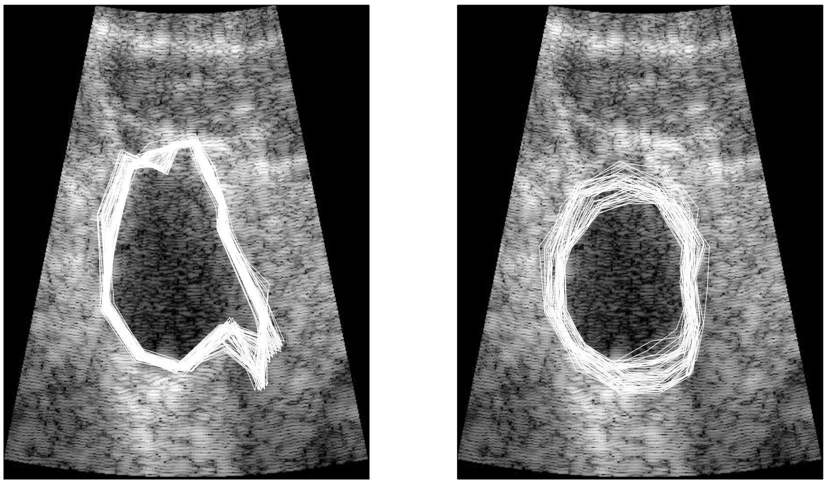


Figure 8.7: Estimated autocorrelation function for the cross-sectional area using Scheme 2 (top) and Scheme 3 (bottom). The functions are calculated using every 100th sample after a burn-in of 25,000 iterations.

Effect of the specular reflection The specular reflection component (8.3) is important for obtaining robust contour estimates. Comparing runs without the specular reflection component ($\rho \equiv 0$) with those with specular reflection, the former seems to lead to wider interval estimates for the cross-sectional vessel area. This is due to greater uncertainty in the sections of the contour where the data is spurious. Figure 8.8 shows why: We have plotted approximately 50 samples taken with a separation of 100 iterations collected late in runs without specular reflection (left image) and with specular reflection (right image). The images clearly demonstrate the gain in modelling the reflection effect occurring at tissue interfaces.

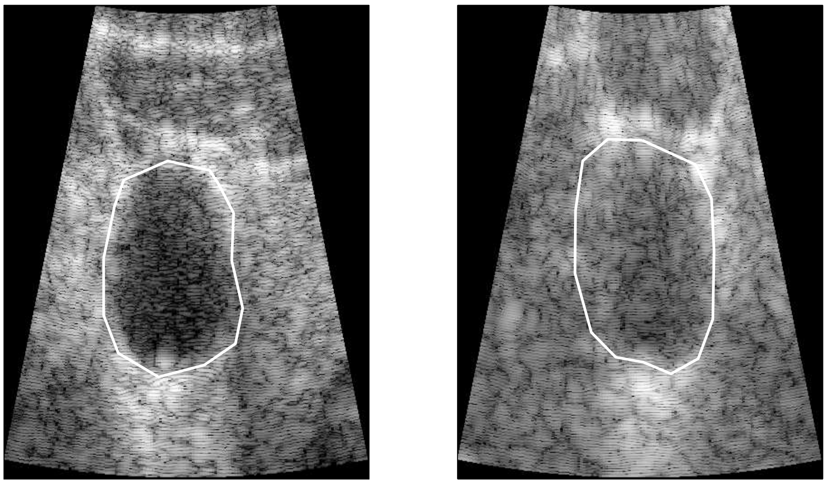


(a) (b)

Figure 8.8: Samples from the posterior distribution for the vessel wall. The samples are collected with a separation of 100 iterations late in the MCMC run.

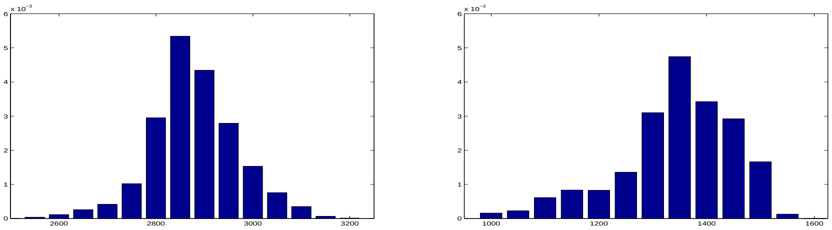
Results Figure 8.10 show histograms of the vessel areas for the images in Figure 1.1, using the model for specular reflection and the prior settings described above. The corresponding 95% credibility intervals are $[2690, 3086]$ for Figure 1.1 (d) and $[1062, 1530]$ for Figure 1.1 (e). We have also computed mean estimates of the vessel walls by using the distance average for random closed sets Section 6.3.2. To get a feel for the uncertainty we have also plotted approximately 50 samples taken with a separation of 100 iterations late in the MCMC run, see Figure 8.8(b) and Figure 8.11(a). The results have been evaluated by cardiologists, who have found them to be in good agreement with their knowledge of vessel shapes.

Benchmark likelihood For comparison we tried contour detection using the standard likelihood model described in Section 8.2.4. In stead of the radio frequency image we used the log compressed intensity image as the data. The model did not prove to be robust for this kind of data. Figure 8.11 shows a sample from the posterior distribution taken after $200,000$ updates. For the



(a) (b)

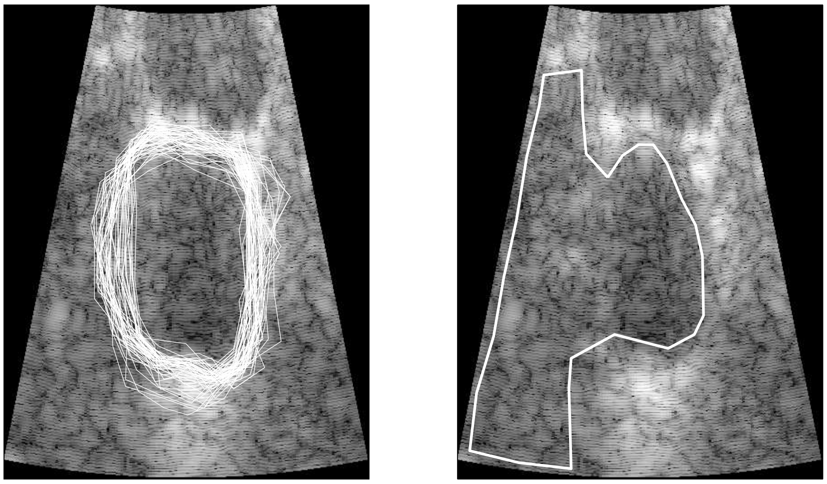
Figure 8.9: Mean estimates of the vessel wall.



(a) (b)

Figure 8.10: (a) and (c) show histograms of the cross-sectional vessel areas for the images in Figure 8.9 (a) and Figure 8.9 (b), respectively.

hyper-priors we used the same parameters as above in Section 8.4.2, and even fixing the parameters κ and η at values giving strong prior information did not affect the behaviour of the sampler. Using a destructive deformation field as described in Section 8.2.1 and Rue and Husby (1998) would possibly amend the problem, but that is more of sampling "trick" than a physically justifiable model feature. Also, with no prior information on the intensities of the back- and foreground regions, we would expect the mode of the distribution to be represented by a curve lying strictly at the interior of the bright regions in the image, thus underestimating the true area.



(a) (b)

Figure 8.11: Sample from the posterior distribution using the simple intensity likelihood in Section 8.2.4.

Chapter 9

Discussion

This report describes Bayesian inference in medical image models, but many of the conclusions have implications for other spatial models. We have discussed non-linear, or edge-preserving, models, and shown how to block sample such models when the observation model is Gaussian or close to Gaussian. We have done experiments indicating that block sampling gives consistently better estimates than single site sampling, and should therefore be the preferred way of sampling for such models.

The second theme describes a Bayesian hierarchical model for contour detection in ultrasound images. We take a fully Bayesian approach, and present interval estimates for the cross-sectional area of an imaged blood vessel. An essential feature of the approach is a physically justifiable Markov random field model for the radio frequency ultrasound data. Previous work and the experiments in this report indicate that the model gives a reasonably good description of the data, and we firmly believe that having a good likelihood is the key for obtaining reliable and reasonably accurate estimates of the quantities of interest. Experiments using a simpler likelihood model seems to support this claim. As object model we use a deformable template, which is well known from other applications. Note that deformable templates are a natural choice when we want to assess the uncertainty of our estimates. Standard methods such as segmentation and deterministic edge detecting algorithms do not provide a natural framework obtaining such uncertainty estimates. Moreover, the combination of the likelihood and object model seems to be particularly fruitful in ultrasound applications, as the the template helps us to directly model the strong reflections occurring at tissue interfaces. This mode of reflection can be very dominant in some images, but is hard to model without having an explicit edge model. Besides being used for object recognition, template models may therefore be helpful in speckle reduction algorithms.

To our knowledge this is the first attempt of using complex hierarchical likelihood models together with deformable template priors, and we have shown that the combination is indeed feasible, although the sampling tends to get involved and time consuming. We have furthermore demonstrated that block sampling is possible for the model, and our results indicate that updating all or nearly all variables in a single block leads to better mixing and hence to more reliable parameter estimates. This is an important point, which is consistent with the results in Knorr-Held and Rue (2002).

Acknowledgement

The author thank T. Langø at Sintef Unimed for providing the ultrasound data, and A. Baddeley, M. A. Hurn and H. Rue for helpful comments and suggestions.

Bibliography

Amit, Y. (1994). A nonlinear variational problem for image matching, *SIAM Journal on Scientific Computing* **15**(1): 207–224.

Amit, Y., Grenander, U. and Piccioni, M. (1991). Structural image restoration through deformable templates, *Journal of the American Statistical Association* **86**: 376–387.

Amit, Y. and Piccioni, M. (1991). A non-homogeneous Markov process for the estimation of Gaussian random fields with non-linear observations, *Annals of Probability* **19**: 1664–1678.

Andersen, K. E., Brooks, S. P. and Hansen, M. B. (2000). A Bayesian approach to crack detection in electrically conducting media, *Technical Report R-00-2019*, Department of Mathematical Sciences, Aalborg University, Denmark.

Andreev, A. and Arjas, E. (1996). A note on histogram approximation in Bayesian density estimation, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford University Press, Oxford, UK, pp. 487–490.

Angelsen, B. A. J. (2000). *Ultrasound Imaging – Waves, Signals, and Signal Processing*, Emantec, Trondheim.

Arjas, E. (1996). Discussion of the paper by J. A. Hartigan, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford University Press, Oxford, UK, pp. 221–222.

Baddeley, A. J. (1992). Errors in binary images and a L^p version of the Hausdorff metric, *Nieuw Archief voor Wiskunde* **10**: 157–183.

Baddeley, A. J. and Molchanov, I. (1998). Averaging of random sets based on their distance functions, *Journal of Mathematical Imaging and Vision* **8**(1): 79–92.

Baddeley, A. J. and Van Lieshout, M. N. M. (1993). Stochastic geometry models in high–level vision, *in* K. V. Mardia and G. K. Kanji (eds), *Statistics and Images*, Vol. 20, Abingdon: Carfax Publishing, chapter 11, pp. 235–256.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching, *Computer Vision, Graphics and Image Processing* **46**: 1–21.

Bakircioglu, M., Grenander, U., Khaneja, N. and Miller, M. I. (1998). Curve matching on brain surfaces using Frenet distance metrics, *Human Brain Mapping* **6**(5): 329–332.

Barone, P. (1999). Fast deconvolution by a two-step method, *SIAM Journal of Scientific Computing* **21**(3): 883–899.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B* **36**: 192–236.

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B* **48**: 259–302.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**(1): 3–66.

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments (with discussion), *Journal of the Royal Statistical Society, Series B* **61**(4): 691–746.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions, *Biometrika* **82**(4): 733–746.

Blake, A., Bascle, B., Isard, M. and MacCormick, J. (1998). Statistical models of visual shape and motion, *Philosophical Transactions of the Royal Society of London Series A* **356**(1740): 1283–1301.

Blake, A. and Isard, M. (1998). *Active Contours*, Berlin: Springer-Verlag.

Blake, A. and Zisserman, A. (1987). *Visual Reconstruction*, MIT Press, Cambridge, MA.

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions, *Statistical Science* **1**: 181–242.

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**: 567–585.

Bookstein, F. L. and Green, W. D. K. (1992). Edge information at landmarks in medical images, *in* R. A. Robb (ed.), *Visualization in Biomedical Computing*, Vol. 1808, SPIE, pp. 242–258.

Boothby, W. (1986). *An Introduction to Differentiable Manifolds and Riemannian Geometry*, Academic Press.

Borgefors, G. (1986). Distance transformations in digital images, *Graphical Models and Image Processing* **34**: 344–371.

Bouman, C. and Sauer, K. (1993). A generalized Gaussian image model for edge-preserving MAP-estimation, *IEEE Transaction on Image Processing* (2): 296–310.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika* **81**(3): 541–553.

Chambolle, A. (1995). Image segmentation by variational methods – Mumford-Shah functional and the discrete approximation, *SIAM Journal of Applied Mathematics* **55**(3): 827–863.

Charbonnier, P. (1994). *Reconstruction d'image: Régularization avec prise en compte des discontinuités*, PhD thesis, Univ. Nice, Sophia Antipolis, France.

Charbonnier, P., Blanc-Feraud, L., Aubert, G. and Barlaud, M. (1997). Deterministic edge-preserving regularization in computed imaging, *IEEE Transaction on Image Processing* **6**(2): 298–311.

Chen, M., Shao, Q. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer Verlag, New York.

Christensen, D. A. (1988). *Ultrasonic Bioinstrumentation*, Wiley.

Christensen, G. E. (1994). *Deformable shape models for anatomy*, PhD thesis, Department of Electrical Engineering, Washington University, St. Louis, MO.

Christensen, G. E., Rabbitt, R. D. and Miller, M. I. (1996). Deformable templates using large deformation kinematics, *IEEE Transaction on Image Processing* **5**(10): 1435–1447.

Cohen, L. (1991). Note on active contour models and balloons, *Computer Vision, Graphics and Image Processing* **53**(2): 211–218.

Cootes, T. F., Hill, A., Taylor, C. J. and Haslam, J. (1994). Use of active shape models for locating structures in medical images, *Image Vision Computing* **12**(6): 355–366.

Cootes, T. F. and Taylor, C. J. (1999). A mixture model for representing shape variation, *Image Vision Computing* **17**: 567–573.

Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J. (1995). Active shape models - their training and application, *Computer Vision Image Understanding* **61**(1): 38–59.

Davatzikos, C. (1996). Spatial normalization of 3-d brain images using deformable models, *Journal of Computer Assisted Tomography* **20**(4): 88–97.

de Figueiredo, M. T. and Leitao, J. M. N. (1992). Bayesian estimation of ventricular contours in angiographic images, *IEEE Transactions on Medical Imaging* **11**(3): 416–429.

de Souza, K. M. A., Kent, J. T. and Mardia, K. V. (1999). Stochastic templates for aquaculture images, *Journal of the Royal Statistical Society, Series C* **48**(2): 211–227.

Delfour, M. C. and Zolesio, J. (1994). Shape analysis via oriented distance functions, *Journal of Functional Anlysis* **123**: 129–201.

Dryden, I. L. and Mardia, K. V. (1991). General shape distributions in a plane, *Advances in Applied Probability* **23**: 259–276.

Dryden, I. and Mardia, K. (1998). *Statistical Shape Analysis*, John Wiley and Sons, Chichester.

Duncan, J. S. and Ayache, N. (2000). Medical image analysis: Progress over two decades and the challenges ahead, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(1): 85–106.

Dupuis, P., Grenander, U. and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching, *Quarterly of Applied Mathematics* **LVI**(3): 587–600.

Friel, N. and Molchanov, I. (1998). Distances between grey-scale images, *Mathematical morphology and its applications to image and signal processing*, Vol. 12 of *Comput. Imaging Vision*, Amsterdam, The Netherlands, pp. 283–290.

Frigessi, A. and Rue, H. (1997). Bayesian image classification with Baddeley's delta loss, *Journal of Computer Graphics and Statistics* **6**(1): 55–73.

Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis* **15**: 183–202.

Gee, J. C. (1999). On matching brain volumes, *Pattern Recognition* **32**(1): 99–111.

Geman, D. and Reynolds, G. (1992). Constrained restoration and the recovery of discontinuities, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(3): 367–383.

Geman, D. and Yang, C. (1993). Nonlinear image recovery with half-quadratic regularization and FFTs, *Technical report*, Dept. of Math. and Stat., Univ of Massachusetts, Amherst.

Geman, D. and Yang, C. (1995). Nonlinear image recovery with half-quadratic regularization, *IEEE Transaction on Image Processing* **4**(7): 932–946.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geman, S. and McClure, D. (1987). Statistical methods for tomographic image reconstruction, *Proc. 46th Sess. Int. Stat. Inst. Bulletin ISI*, Vol. 52.

Georgiou, G. and Cohen, F. (1998). Statistical characterization of diffuse scattering in ultrasound images, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.* **45**(1): 57–64.

Geyer, C. (1996). Markov chain Monte Carlo lecture notes, *Technical report*, School of Statistics, University of Minnesota.

Geyer, C. J. (1991). Estimating normalizing constants and reweighting mixtures in Markov chain Monte Carlo, *Technical Report 568*, School of Statistics, University of Minnesota.

Glasbey, C. A. (1998). Ultrasound image segmentation using stochastic templates, *Journal of Computing and Information Technology* **6**: 107–116.

Glasbey, C. A. and Mardia, K. V. (1998). A review of image-warping methods, *Journal of Applied Statistics* **25**(2): 155–171.

Glasbey, C. A. and Mardia, K. V. (2001). A penalised likelihood approach to image warping (with discussion), *Journal of the Royal Statistical Society, Series B* **63**(3): 465–492. To appear.

Goodman, J. (1975). Statistical properties of laser speckle patterns, *in* J. Dainty (ed.), *Laser Speckle and Related Phenomena*, Springer Verlag, Berlin.

Green, P. J. (1990). Bayesian reconstruction from emission tomography data using a modified EM algorithm, *IEEE Transactions on Medical Imaging* **9**(1): 84–93.

Green, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* **82**(4): 711–732.

Grenander, U. (1967). Toward a theory of patterns, *Symposium on Probability Methods in Analysis*, Loutraki, Greece;, Berlin: Springer-Verlag.

Grenander, U. (1969). Foundations of pattern analysis, *Quarterly of Applied Mathematics* **27**: 1–55.

Grenander, U. (1981). *Regular Structures: Lectures in Pattern Theory*, Vol. III, Springer, New York.

Grenander, U. (1993). *General Pattern Theory*, Oxford University Press.

Grenander, U., Chow, Y. and Keenan, D. M. (1991). *Hands: a Pattern Theoretic Study of Biological Shapes*, Research Notes on Neural Computing, Springer Verlag, Berlin.

Grenander, U. and Manbeck, K. M. (1993). A stochastic model for defect detection in potatoes, *Journal of Computer Graphics and Statistics* **2**: 131–151.

Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems (with discussion), *Journal of the Royal Statistical Society, Series B* **56**(4): 549–603.

Grenander, U. and Miller, M. I. (1998). Computational anatomy: An emerging discipline, *Quarterly of Applied Mathematics* **LVI**(4): 617–694.

Guyon, X. (1995). *Random Fields on a Network – Modelling, statistics, and applications*, Springer Verlag, New York.

Hansen, M. B., Møller, J. and Tøgersen, F. A. (2000). Bayesian contour detection in a time series of ultrasound images through dynamic deformable template models, *Preprint*, Aalborg University, Denmark.

Heap, T. and Hogg, D. (1996). Automated pivot location for the Cartesian-polar hybrid point distribution model, *in* R. Fisher and E. Trucco (eds), *7th British Machine Vision Conference*, BMVA Press, Edinburgh, UK, pp. 276–285.

Hebert, T. and Leahy, R. (1989). A generalized EM algorithm for 3D Bayesian reconstruction from Poisson data using Gibbs priors, *IEEE Transactions on Medical Imaging* **8**(2): 194–202.

Higdon, D. M., Bowsher, J. E., Johnson, V. E., Turkington, T. G., Gilland, D. R. and Jaszczak, R. J. (1997). Fully Bayesian estimation of Gibbs hyerparameters for emission computed tomography data, *IEEE Transactions on Medical Imaging* **16**(5): 516–526.

Higdon, D. and Yamamoto, S. (2000). Bayesian image analysis in scanning magnetoresistance microscopy, *Technical report*, Duke University.

Hobolth, A. and Jensen, E. B. V. (2000). Modelling stochastic changes in curve shape, with application to cancer diagnostics, *Advances in Applied Probability (SGSA)* **32**(2): 344–362.

Hobolth, A., Kent, J. T. and Dryden, I. L. (1999). On the relationship between edge and vertex modelling, *Technical report*, University of Aarhus, Denmark.

Hobolth, A., Pedersen, J. and Jensen, E. B. V. (2000). A continuous parametric shape-model, *Technical Report 13*, Department of Mathematical Sciences, University of Aarhus, Denmark.

Hobolth, A., Pedersen, J. and Jensen, E. B. V. (2001). A deformable template model with special reference to elliptical templates, *Technical Report 14*, Department of Mathematical Sciences, University of Aarhus, Denmark.

Hokland, J. (1995). *Speckle reduction, restoration, and volume visualization in medical ultrasonics*, PhD thesis, Department of Physiology, University of Bergen, Bergen, Norway.

Hokland, J. and Kelly, P. (1996). Markov models of specular and diffuse scattering in restoration of medical ultrasound images, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **43**(4): 660–669.

Hurn, M. A. (1998). Confocal fluorescence microscopy of leaf cells: An application of Bayesian image analysis, *Journal of the Royal Statistical Society, Series C* **47**(3): 361–377.

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* **18**(6): 657–662.

Hurn, M., Steinsland, I. and Rue, H. (2001). Parameter estimation for a deformable template model, *Statistics and Computing* **11**: 325–334.

Husby, O., Lie, T., Langø, T., Hokland, J. and Rue, H. (2001). Bayesian 2d deconvolution: A model for diffuse ultrasound sc attering, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.* **48**(1): 121–130.

Insana, M. F., Wagner, R. F., Garra, B. S., Brown, D. G. and Shawker, T. H. (1986). Analysis of ultrasound image texture via generalized Rician statistics, *Optical Engineering* **25**: 743–748.

Iwai, T. and Asakura, T. (1996). Speckle reduction in coherent information processing, *Proc. IEEE*, Vol. 0084, pp. 7765–791.

Jain, A. K., Zhong, Y. and Dubuisson-Jolly, M. (1998). Deformable template models: A review, *Signal Processing* **71**: 109–129.

Jalobeanu, A., Blanc-Feraud, L. and Zerubia, J. (1998). Estimation d'hyperparamètres pour la restauration d'images satellitaires par une méthode MCMCML, *Technical Report No 3469*, INRIA.

Jensen, J. and Leeman, S. (1994). Nonparametric estimation of ultrasound pulses, *IEEE Transactions on Biomedical Engineering* **41**(10): 929–936.

Joshi, S. (1997). *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain Submanifolds*, PhD thesis, Department of Electrical Engineering, Washington University, St. Louis, MO.

Joshi, S. C., Miller, M. I., Christensen, G. E., Banerjee, A., Coogan, T. A. and Grenander, U. (1995). Hierarchical brain mapping via a generalized Dirichlet solution for mapping brain manifolds., *in* R. A. Melter, A. Y. Wu, F. L. Bookstein and W. D. K. Green (eds), *Vision Geometry IV*, Vol. 2573, SPIE, pp. 278–289.

Kao, C.-M., Pan, X., Hiller, E. and Chen, C.-T. (1998). A Bayesian approach for edge detection in medical ultrasound images, *IEEE Transaction on Nuclear Science* **45**(6): 3089–3096.

Kass, M., Witkin, A. and Terzeopoulos, D. (1988). Snakes: Active contour models, *International Journal of Computer Vision* **1**(4): 321–331.

Kendall, D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces, *Bull. London Math. Soc.* **16**: 81–121.

Kent, J. T. (1994). The complex Bingham distribution and shape analysis, *Journal of the Royal Statistical Society, Series B* **56**(2): 285–299.

Kent, J. T., Dryden, I. L. and Anderson, C. R. (2000). Using circulant symmetry to model featureless objects, *Biometrika* **87**(3): 527–544.

Kent, J. T. and Mardia, K. V. (1994). The link between kriging and thin-plate splines, *in* F. P. Kelly (ed.), *Probability, Statistics and Optimisation*, John Wiley & Sons.

Kent, J. T., Mardia, K. V. and Walder, A. N. (1996). Conditional cyclic Markov random fields, *Advances in Applied Probability (SGSA)* **28**: 1–12.

Kervrann, C. and Heitz, F. (1999). Statistical deformable model-based segmentation of image motion, *IEEE Transaction on Image Processing* **8**(4): 583–588.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping, *Scandinavian Journal of Statistics* . (to appear).

Langø, T. (2000). *Ultrasound guided surgery: Image processing and navigation*, PhD thesis, Norwegian University of Science and Technology, Trondheim, Norway.

Langø, T., Lie, T., Husby, O. and Hokland, J. (2001). Bayesian 2d deconvolution: Effect of using spatially invaria nt point spread functions, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.* **48**(1): 131–141.

Lauritzen, S. (1996). *Graphical Models*, Oxford University Press, Oxford, UK.

Lefebvre, F., Berger, G. and Laugier, P. (1998). Automatic detection of the boundary of the calcaneus from ultrasound parametric images using an active contour model; clinical assessment, *IEEE Transactions on Medical Imaging* **17**(1): 45–52.

Lewis, T., Owens, R. and Baddeley, A. J. (1999). Averaging feature maps, *Pattern Recognition* **32**(9): 1615–1630.

Liu, J. S., Wong, W. H. and Kong, A. (1994). Covariance of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes, *Biometrika* **81**(1): 27–40.

Lizzi, F. L., Greenbaum, M., Felappa, E. J. and Elbaum, M. (1983). Theoretical framework for spectrum analysis in ultrasonic tissue characterization, *J. Acoust. Soc. Am.* **73**: 1366–1373.

Mardia, K. V. and Dryden, I. L. (1999). The complex Watson distribution and shape analysis, *Journal of the Royal Statistical Society, Series B* **61**(4): 913–926.

Mardia, K. V., Kent, J. T., Goodall, C. R. and Little, J. A. (1996). Kriging and splines with derivative information, *Biometrika* **83**: 207–221.

Mardia, K. V., Qian, W., Shah, D. and de Souza, K. M. A. (1997). Deformable template recognition of multiple occluded objects, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(9): 1035–1042.

McInerey, T. and Terzopoulos, D. (1996). Deforamble models in medical image analysis: A survey, *Medical Image Analysis* **1**(2): 91–108.

Mikic, I., Krucinski, S. and Thomas, J. D. (1998). Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates, *IEEE Transactions on Medical Imaging* **17**(2): 274–284.

Miller, M. I., Christensen, G. E., Amit, Y. and Grenander, U. (1993). Mathematical textbook of deformable neuroanatomies, *Proceedings of the National Academy of Science* **90**(24): 11944–11948.

Miller, M. I., Srivastava, A. and Grenander, U. (1995). Conditional-mean estimation via jump-diffusion processes in multiple target tracking/recognition, *IEEE Transactions on Signal Processing* **43**(11): 2678–2690.

Møller, J., Syversveen, A. R. and Waagepetersen, R. (1998). Log Gaussian Cox processes, *Scandinavian Journal of Statistics* **25**(3): 451–482.

Mulet-Parada, M. and Noble, J. A. (2000). 2D+T acoustic boundary detection in echocardiography, *Medical Image Analysis* **4**(1): 21–30.

Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems, *Communications on Pure and Applied Mathematics* **42**(5): 577–685.

Nicholls, G. and Fox, C. (1998). Prior modelling and posterior sampling in impedance imaging, *in* A. Mohammad-Djafari (ed.), *Bayesian Inference for Inverse Problems*, Vol. 3459 of *Proceedings of SPIE*, SPIE, pp. 116–237.

Ødegård, L. (1995). *Phase aberration correction in medical ultrasound imaging*, PhD thesis, The Norwegian Institute of Technology, Trondheim, Norway.

Perrin, O. and Senoussi, R. (1999). Reducing non-stationary stochastic processes to stationarity by a time deformation, *Statistics and Probability Letters* **43**(4): 393–307.

Pham, D. L., Cu, C. Y. and Prince, J. L. (2000). Current methods in medical image segmentation, *Annual Review of Biomedical Engineering* **2**: 315–.

Phillips, D. B. and Smith, A. F. M. (1994). Bayesian faces via hierarchical template modeling, *Journal of the American Statistical Association* **89**(428): 1151–1163.

Piccioni, M., Scarlatti, S. and Trouvé, A. (1998). A variational problem arising from speech recognition, *SIAM Journal of Applied Mathematics* **58**(3): 753–771.

Qian, W., Titterington, D. M. and Chapman, J. N. (1996). An image analysis problem in electron microscopy, *Journal of the American Statistical Association* **91**(435): 944–952.

Ripley, B. D. and Sutherland, A. I. (1990). Finding spiral structures in images of galaxies, *Philosophical Transactions of the Royal Society of London A* **332**: 477–485.

Robert, C. P. and Casella, G. (1999). *Monte Carlo statistical methods*, Springer-Verlag New York.

Roberts, G. O. and Rosenthal, J. S. (1998). Optimalscaling of discrete approximations to Langevin diffusions, *Journal of the Royal Statistical Society, Series B* **60**(2): 255–268.

Rosati, M. (2000). Asymptotic behaviour of a Geman and McClure discrete model, *Applied Mathematics and Optimization* **41**(1): 51–85.

Rue, H. (1995). New loss functions in Bayesian imaging, *Journal of the American Statistical Association* **90**(431): 900–908.

Rue, H. (1997). A loss function model for the restoration of grey level images, *Scandinavian Journal of Statistics* **24**(1): 103–114.

Rue, H. (1999). Baddeley's delta metric, *in* S. Kotz, C. B. Read and D. L. Banks (eds), *Encyclopedia of Statistical Sciences*, Vol. Update Volume 3, Wiley, pp. 158–162.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields with applications, *Journal of the Royal Statistical Society, Series B* **63**(2): 325–338.

Rue, H. and Hurn, M. A. (1999). Bayesian object identification, *Biometrika* **86**(3): 649–660.

Rue, H. and Husby, O. K. (1998). Identification of partly destroyed objects using deformable templates, *Statistics and Computing* **8**(3): 221–228.

Rue, H., Steinsland, I. and Erland, S. (2001). Approximating hidden Gaussian Markov random fields. In preparation.

Rue, H. and Syversveen, A. R. (1998). Bayesian object recognition with Baddeley's delta loss, *Advances in Applied Probability (SGSA)* **30**(1): 64–84.

Rue, H. and Tjelmeland, H. (2001). Fitting Gaussian Markov random fields to Gaussian fields, *Scandinavian Journal of Statistics* . (to appear).

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association* **87**(417): 108–119.

Schultz, R. R. and Stevenson, R. L. (1995). Stochastic modelling and estimation of multispectral image data, *IEEE Transaction on Image Processing* **4**(8): 1109–1119.

Sozou, P., Cootes, T. F., Taylor, C. J. and Mauro, E. D. (1995). A non-linear generalisation of point distribution models using polynomial regression, *Image Vision Computing* **13**(5): 451–457.

Stoica, R., Descombes, X. and Zerubia, J. (2000). A Markov point process for road extraction in remote sensed images, *Technical Report 3923*, INRIA.

Taxt, T. (1995). Restoration of medical ultrasound images using 2-dimensional homomorphic deconvolution, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.* **42**(4): 543–554.

Taxt, T. and Strand, J. (2001). Two-dimensional noise-robust blind deconvolution of ultrasound images, *IEEE Trans. on Ultrason. Ferroelec. Frec. Contr.* **48**(4): 861–866.

Thompson, M. P. and Toga, W. (1996). Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations, *Medical Image Analysis* **7**: 271–294.

Trouvé, A. (1998). Diffeormorphism groups and pattern matching in image analysis, *International Journal of Computer Vision* **28**: 213–221.

Varghese, T., Donohue, K. D. and Chatterjee, J. P. (1995). Specular echo imaging with spectral correlation, *Proc. IEEE Ultrason. Symp.*, pp. 1315–1318.

Wagner, R., Insana, M. and Brown, D. (1987). Statistical properties of radio-frequency and envelope-detected signals with application to medical ultrasound, *J. Opt. Soc. Am, A* **4**(5): 910–922.

Weir, I. S. (1997). Fully Bayesian reconstructions from single-photon emission computed tomography, *Journal of the American Statistical Association* **92**(437): 49–60.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin.

Wood, A. T. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$, *Journal of Computational and Graphical Statistics* **3**(4): 409–432.

Younes, L. (1998). Computable elastic distances between shapes, *SIAM Journal on Applied Mathematics* **58**(2): 565–586.

Younes, L. (1999). Optimal matching between shapes via elastic deformations, *Image Vision Computing* **17**(5-6): 381–389.

Yuille, A. L., Hallinan, P. W. and Cohen, D. S. (1992). Feature extraction from faces using deformable templates, *International Journal of Computer Vision* **8**(2): 133–144.

Zhong, Y. and Jain, A. K. (1997). Object localization using color, texture and shape, *Proc. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, Springer, Berlin, Venice, pp. 279–294.

Zong, X., Laine, A. and Geiser, E. (1998). Speckle reduction and contrast enhencement of echocardiograms via multiscale nonlinear processing, *IEEE Transactions on Medical Imaging* **17**(4): 532–540.