Statistical Estimation of the Parameters of a PDE

Colin Fox and Geoff Nicholls

December 2001

Abstract

We consider the class of image recovery problems where the desired image appears as the spatially-varying coefficients of a partial differential equation (PDE), the data consists of measured values on the boundary, and the forward map can only be adequately simulated by solving the PDE subject to boundary conditions. This is the natural mathematical setting for non-invasive imaging using strongly scattered waves. We give a gentle introduction to the statistical (Bayesian) approach to solving this class of inverse problems, and present some computational examples using sampling algorithms. The statistical approach quantifies the inherent uncertainty in images recovered from incomplete noisey data using knowledge of the forward map the measurement process and noise statistics. The single most likely image, usually found by applying the inverse of the forward map to the measured data, does not give a good reconstruction in this class of problems because that image is unrepresentative of the bulk of feasible images. Instead it is necessary to summarize the feasible images by calculating expectations over the posterior distribution. One advantage of this route to solving inverse problems is the ability to quantify accuracy within the recovered image, while a major disadvantage is computational expense. We use the illustrative example of imaging electrical conductivity and give examples of reconstruction of an unknown conductivity from simple synthetic data. In principle, high-level models may be incorporated relatively easily in the sampling algorithms, and we give a glimpse of some image models that are currently used.

1 Introduction

Non-invasive imaging used for medical diagnostics and geophysics often uses scattering of waves to probe the object under investigation. These techniques are non-invasive because the wave scattering is measured remotely from the object, and the object is irradiated with waves from a source remote to the object.

A common example is X-ray tomography where a person is irradiated from one side using a point X-ray source and the partially-transmitted waves are measured using a photographic plate on the opposite side to the source, as depicted

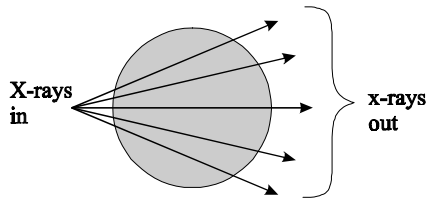

Figure 1: A schematic of the measurement process in X-ray tomography.

in figure 1. Because X- rays propagate through tissue in essentially straight-line paths, largely independent of the particular tissue being imaged, the measurement process may be treated as linear. That is, to a good approximation, the mapping from unknown spatially-varying X-ray absorptivity to the measured values is a linear relation.

However, many potential imaging modalities such as optical diffusion 'tomography' involve strong scattering of the waves. In these cases the paths of propagation depend implicitly on the object being imaged. Then the wave field that propagates and scatters must be modelled by requiring that it satisfies a partial differential equation (PDE) in which the unknown object properties appear as spatially-varying coefficients. Three examples of governing PDEs, corresponding to three different types of energy being propagated, are given in the following table.

object property	governing PDE	PDE classification
electrical conductivity	$\nabla \cdot (\sigma \nabla \phi) = s$	elliptic
acoustic impedance	$\nabla \cdot (\sigma \nabla p) = \frac{\sigma}{c^2} \ddot{p}$	hyperbolic
thermal conductivity	$\nabla \cdot (\sigma \nabla u) = \dot{u}$	parabolic

In each case the object property being sought is denoted by σ and appears as the spatially-varying coefficient in the PDE governing the propagation of energy, or 'waves'. The measurements made are of the boundary values of electrical potential ϕ, sound pressure p, and temperature u, respectively in the three cases. If the scattered field can be well approximated as a small change about a known field then the (linear) Born approximation may be used to simulate the measurement process. Otherwise, the measurement process must be simulated by solving the PDE subject to boundary conditions that correspond to the wave irradiation. In that case the mapping from object property, σ, to measurements is nonlinear. This latter property presents a basic difficulty in these inverse problems and, as we will see later, is one reason why straightforward application of algorithms that have been successful for linear inverse problems, such as X-ray tomography, have been unsuccessful in solving this class of inverse problems.

A second source of difficulty can be understood by including in the analysis the statistical detail of errors inherent in the measurement process. Inclusion of the error process is critical for practical solution inverse problems such as these,

where the forward map has a large range of sensitivities, though it is often only treated in an *ad hoc* way in many mathematical analyses. A probabilistic model for the error process results in a probabilistic model for the measurement process and inversion is then naturally a problem of statistical inference for parameter estimation from data. A robust solution can often be achieved by also introducing prior information in the form of a probability density over allowable reconstructions. These three steps; accurately modelling the measurement process, including the statistics of the measurement error and using a prior density, characterize the 'Bayesian' approach to inverse problems.

Strictly, the Bayesian analysis ends with the formation of the 'posterior' probability density function over parameter space. However, useful image recovery requires presenting just one result (or at most a few) which is best done by summarizing the posterior. How best to summarize the posterior remains a topic of research and emerging ideas. A few decades ago available computing resources limited the methods available to optimization algorithms designed to find the mode of the posterior distribution, giving image recovery methods that are equivalent to regularization. We give examples of regularized inversion in section 3.1.1. The widespread use of the mode has lead to unquestioning acceptance that the mode is the best way to summarize the posterior. However that is not the case for the class of problems considered here, and we give reasons for that and specific counter examples in section 3.1.4. Instead, summarizing the posterior requires considering the bulk of states in the support of the posterior, which typically entails calculating expectations over the posterior distribution. These high-dimensional integrals are computationally expensive, and much of the authors' research in the past few years has been to find efficient algorithms for problems where the forward map requires solution of a PDE. We give examples of such calculations and thereby demonstrate that the required calculations are tractable.

The simplest case of the title problem is imaging electrical conductivity in 2 dimensions, though it shares with the others the primary computational difficulty that the space part of the PDE is the elliptic, Laplacian-like, operator. In this paper we use imaging of electrical conductivity as a running example, and show how that inverse problem can be solved within an inferential computation using Markov chain Monte Carlo. In section 2 we define an idealized measurement set for conductivity imaging, in section 4 we present a statistical formulation of the inverse problem, and in section 4.3 we present reconstructed images, and other statistics. It is of formal interest to note that the other imaging problems, such as imaging from ultrasound backscatter, can be shown to be equivalent to a sequence of problems of this type [11].

While the expense of computing expectations is a major drawback of the Bayesian approach to inverse problems, there are several advantages of this method. One advantage is that other summary statistics may be computed at the same time as the estimate of the image, such as the variance of the recovered image values which gives a measure of ambiguity in the image. Examples of the image pixel variance are presented in section 4.3. A second advantage is that present, Markov chain, sampling algorithms allow the use of more sophisticated

image models than do gradient-based optimization algorithms. In particular the sampling algorithms can work with image space models that may be discrete, or discontinuous, or even variable dimension. We conclude by giving a few examples of mid- and high-level image-space models in section 5.

1.1 Properties of Inverse Problems for Image Recovery

In our work the spatially-varying parameters of the PDE form an image and we often refer to the inverse problem as one of *image recovery*. The *data* is the set of measurements from which the image is recovered, which will be either far-field measurements of scattered waves or of appropriate boundary data. The *forward problem* (or map) is the mapping from an actual image to the corresponding data (strictly the noise-free data) while the inverse problem is to recover the true image from measured data.

The following table sets out some contrasting properties between the forward and inverse problems.

Forward Problem	**Inverse Problem**
image $\longrightarrow$ data	data $\longrightarrow$ image
physical model (PDE)	implicit
direct computation	indirect
well posed	ill posed
unique	never unique

The forward problem is found by modelling the measurement process using a physical theory for wave scattering, and (as we saw in section 1) takes the form of a PDE parameterized by the image. Simulating the forward map is a well-determined direct computation (even though the PDE solution may be computed via an implicit numerical algorithm). In contrast, the inverse problem is normally only implicitly defined by the requirement that the recovered image be consistent with the data and so image recovery is an indirect calculation. Because the forward problem corresponds to a repeatable cause-effect sequence, the forward map is unique and well posed in a natural metric. That causal relationship also necessarily involves a loss of information (or increase in entropy) [6] which makes the forward and inverse maps fundamentally different. The inverse problem then requires recovering lost information which shows up by the inverse problem being ill-posed, in our case discontinuous even in the case of complete noise-free data. In the practical case where we try to recover a continuously varying function on the basis of a finite data set, the inverse problem cannot have a unique solution.

Later we will distinguish between the idealized data resulting from the forward map, and the actual measured data which involves measurement imprecision.

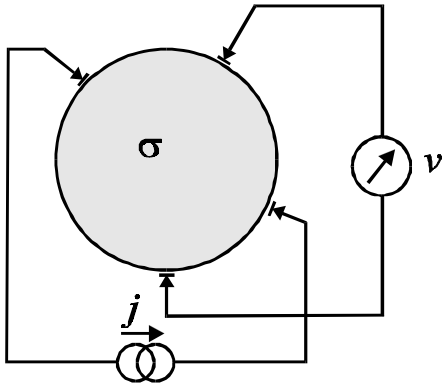

Figure 2: A schematic of a typical measurement for imaging of electrical conductivity. Current j is asserted on a pair of electrodes and the resulting voltage v is measured on a second electrode pair.

2 Conductivity Imaging Measurement Set

In this section we define a simplified model for electrical conductivity imaging (sometimes called electrical impedance tomography or EIT) to use as an illustrative inverse problem. In section 4 we formulate this problem in terms of statistical inference and give numerical examples of mean images using synthetic data.

Figure Fig 2 shows a simplified schematic of a typical measurement process for imaging the conductivity within an object occupying the circular region Ω. Electrodes are placed around the boundary of the object, $\partial\Omega$, at positions $x_1, x_2, \cdots, x_E$. The schematic shows a current j being asserted on a pair of electrodes with the resulting voltage v measured on a second electrode pair. More generally we may assert a vector of currents $j = \left(j\left(x_1\right), j\left(x_2\right), \cdots, j\left(x_E\right)\right)^T$ and measure the resulting vector of voltages $v = \left(\phi\left(x_1\right), \phi\left(x_2\right), \cdots, \phi\left(x_E\right)\right)^T$ with respect to some reference potential. If N different current vectors are asserted, and the resulting voltages measured, then the data consists of the set of current-voltage pairs

$$d = \left\{j^{(n)}, v^{(n)}\right\}_{n=1}^{N}.$$

The unknown internal conductivity $\sigma\left(x\right)$ is related to these measurements via Neumann boundary-value problem (BVP)

$$\begin{aligned} \nabla \cdot \sigma\left(x\right) \nabla \phi\left(x\right) &= 0 && x \in \Omega \\ \sigma\left(x\right) \frac{\partial \phi\left(x\right)}{\partial n\left(x\right)} &= j\left(x\right) && x \in \partial\Omega \end{aligned}$$

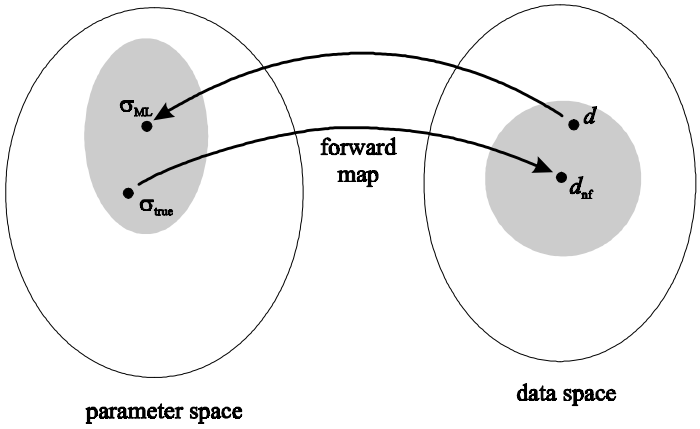

Figure 3: A schematic showing parameter space Σ_Ω and data space D. Shown are the true image $\sigma_{\rm true}$, noise-free data $d_{\rm nf}$, measured data d and the most-likely image $\sigma_{\rm ML}$. The grey region in data space shows the range of possible measured data, while the grey region in parameter space shows the set of possible images consistent with the data d.

along with a reference for ϕ such as $\int_{\partial\Omega}\phi dl = 0$. Thus the data consists of realizations of the inverse of the Dirichlet to Neumann map on the boundary[16], denoted Γ_σ. The inverse problem is to find $\sigma(x)$ from d. Note that simulation of the inverse of the Dirichlet-to-Neumann map requires solving a linear PDE, however the relationship between the unknown conductivity σ and the measurements is not linear and, hence, the inverse problem is non-linear.

3 Statistical Model of Inverse Problems

In this section we give two descriptions of the relationships between image and data in inverse problems in the presence of noise. First an intuitive description and then the formal Bayesian model found by quantifying the error in the measurement process and the resulting inherent uncertainty in the images recovered from measured data.

Figure Fig 3 gives a schematic of the distributions of possible data and recovered images when measurement error is considered. True image $\sigma_{\rm true}$ results in ideal, noise-free, data $d_{\rm nf}$ under the forward map. Because of measurement inaccuracy we do not actually measure data $d_{\rm nf}$, but some slightly noisy version. The range of possible measured data, after corruption by noise, is shown as a grey region around $d_{\rm nf}$. In a particular experiment we measure a single data set, say d. If an inverse to the forward map exists we may apply the inverse to the measured data d, or equivalently use a model-fitting procedure, to find image $\sigma_{\rm ML}$ which is typically the 'maximum likelihood' estimate of the image. Because

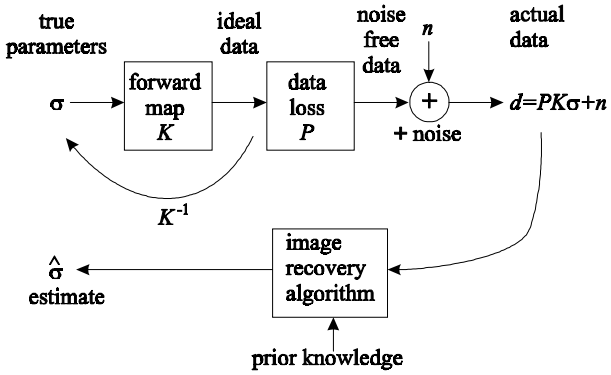

Figure 4: A block model of the measurement and imaging process.

of the measurement error on d, there is in fact a range of feasible images each being consistent with the measured data up to a feasible noise vector, and these are shown by the grey region around σ_{ML}. In practical image recovery problems we begin with the actual measurements d, can calculate σ_{ML} and the range of feasible images, but do not have access to σ_{true} or d_{nf}.

A point that we would like to emphasize in this paper, is that the image σ_{ML} calculated by straightforward inversion (or even regularized inversion) of the noisy data d, is generally a poor estimate of the true parameters for the class of title problems. We will examine this issue further in section 3.1.4.

The uncertainties in measured data and recovered image can be quantified by modelling the various stages in the measurement process as shown in the block diagram in figure 4. The true parameters, or image, is the starting point of the chain. We do not measure σ directly, but rather some function of σ where the function is called the forward map, denoted K. In this paper we take K to be the map from σ to ideal data – that is data consisting of all possible measurable properties of σ without any measurement error. The focus of many theoretical studies in inverse problems is investigation of the existence and uniqueness (and perhaps continuity) of the inverse map K^{-1}. In many cases in this class of inverse problems, K^{-1} is unbounded and so its existence is not necessarily relevant to the performance of image recovery from actual data. Further, with actual measurement equipment we measure a finite data set even though there may be infinitely many possible measurables. The reduction from all measurables to the set of quantities actually measured is often called "data loss" though in mathematical terms it is a projection operator, P. It is usual to refer to the true values of the measured quantities as the noise-free data. It is unavoidable that this data is corrupted by measurement error with the difference between the measured data and the noise free data being some noise vector n. The probability density function (pdf) of the noise, $f_N(n)$ say, is a property of the measurement equipment and it is this pdf that makes the

measurement process probabilistic. The measured data, d, is a known function of two unknowns, signal and noise. In order to fix ideas, suppose the noise is additive, $d = PK\sigma + n$. Since the noise n is a random variable, and d depends on n, then also the data d is a random variable with pdf given by $f_{D|\Sigma}(d|\sigma) = f_N(d - PK\sigma)$ (the Jacobian for the change of variables from n to d is unity here). Thus the measurement process corresponds to drawing samples from a known probability distribution that is parameterized by σ (and perhaps other parameters) and the job of recovering σ corresponds to classical statistical inference for estimating the parameter σ. The pdf $f_{D|\Sigma}$ is called the likelihood function for the conductivity parameter.

A common case of noise pdf is when the measurement equipment has a linear response and measurement error comes from many sources. Then the noise is often normally distributed with zero mean, i.e. $n \sim N(0, s^2)$ with standard deviation s. As a result the data is also normally distributed with $d \sim N(PK\sigma, s^2)$, i.e. the unknown image is a parameter setting the mean of the distribution. For the case of zero mean normally distributed noise, the likelihood is explicitly

$$f_{D|\Sigma}(d|\sigma) \propto \exp(\|d - PK\sigma\|^2 / 2s^2).$$

The measured data is finite dimensional. If the image is represented as a function of real variables, it follows that there are infinitely many images that are consistent with the data and the image recovery problem does not have a unique solution. This property also holds for discrete versions of the inverse problem. For example, in conductivity imaging of circular objects, linearized about the constant image, there a few more than 100 singular values relatively larger than the (good) measurement error of 1 part in 1000 [8]. Hence, only about 100 components of σ are effectively measured, even though many more pixels are required to represent σ if the measurement process is to be accurately modeled. So even in the discrete case, the inverse problem is inherently non unique.

The extra information, not available from the measurement process, that needs to be included in the image recovery algorithm if a single value of σ is to be recovered is often called prior knowledge. The nature of this information can take various forms, with a corresponding range of mathematical defensibility. The most defendable type of extra information are physical constraints on the image such as the positivity of brightness, or electrical conductivity. More generally we may know that the physical system is invariant under some group of transformations [14], and hence so must be the representation of the image. Regularization is often used in non-statistical approaches to inverse problems and can be couched as prior knowledge, though it usually is determined by a mathematically expedient requirement that the recovered image be smooth, or small. The least defendable, though perhaps the most useful and the one that would be useful to encode mathematically, are preferences of an expert who would discard images containing obvious artifacts simply because they are highly unlikely, though strictly possible.

To simplify notation we will suppose that the space Σ_Ω of conductivity images is finite. The probability distribution of feasible images, consistent with the actual data, may be quantified using Bayes' rule for conditional probabilities. The 'posterior' distribution for σ conditional on measuring d is

$$\Pr(\sigma|d) = \frac{f_{D|\Sigma}(d|\sigma)\Pr(\sigma)}{\sum_{\sigma\in\Sigma_\Omega} f_{D|\Sigma}(d|\sigma)} \qquad \text{(Bayes' rule)}$$

where $\Pr(\sigma)$ is the prior probability distribution for images. In the 'subjectivist' formulation of probability the prior and posterior distributions for σ, $\Pr(\sigma)$and $\Pr(\sigma|v)$, are quantified representations of our state of knowledge about the true value of σ [15].

The explicit use of a prior pdf is often a conceptual sticking point for some people, though its inclusion is unavoidable. Since $\Pr(\sigma)$ is a distribution, its functional form depends on the coordinate system (or more formally the measure) chosen. For any sufficiently smooth prior, there is a corresponding coordinate system in which $\Pr(\sigma)$ is constant. So Laplace's "principle of insufficient reason", which states that the prior should be set to a constant if no good reason to the contrary exists, is equivalent to a quite strong statement that in the coordinate system chosen the prior takes a constant value. Under a different measure, the prior would not be constant. So choice of a prior distribution has the same status as choosing the coordinate system for representing σ, which is unavoidable. Our preference is to choose a representation for σ that is convenient for computation and for expressing any knowledge about the image, and then explicitly write the prior distribution.

3.1 Solutions are Summary Statistics

Once a particular data set has been measured, all information about the desired parameters σ is contained in posterior distribution $\Pr(\sigma|d)$. In image recovery, σ typically has many hundreds or thousands of parameters (e.g. pixels) and the range of the posterior distribution is that high-dimensional space. Hence it is not convenient, nor useful, to convey the whole posterior distribution to an end user. Instead, summary statistics that characterize the posterior distribution should be found, perhaps consisting of a single image or a few images that indicate the bulk of feasible images in the support of the posterior.

3.1.1 Traditional Solutions - modes

For many decades, image recovery has been performed by finding, and reporting, the mode of the posterior distribution. This approach has the intuitive appeal that the mode is the single most-likely image – however, as we will see later, that does not mean that the mode is representative of the bulk of feasible images. Specifically the modes are the maximum Likelihood estimate (ML) and the maximum a posteriori estimate (MAP), given by

$$\hat{\sigma}_{\mathrm{ML}} = \arg\max_{\sigma\in\Sigma_\Omega} f_{D|\Sigma}(d|\sigma),$$
$$\hat{\sigma}_{\mathrm{MAP}} = \arg\max_{\sigma\in\Sigma_\Omega} f_{D|\Sigma}(d|\sigma)\Pr(\sigma).$$

For example, if the noise is additive Gaussian with zero mean and variance s^2, and we use a simple Gaussian prior of the form $\Pr(d\sigma) \propto \exp\left(-\|\sigma\|^2/2\lambda^2\right) d\sigma$, then

$$\hat{\sigma}_{\mathrm{MAP}} = \arg\min \|d - PK\sigma\|^2 + \theta\|\sigma\|^2 \tag{1}$$

where $\theta = s^2/\lambda^2$ is called the regularizing parameter. This estimate has been discovered many times, in a number of settings and is variously known as Tikhonov regularization, Prussian-Hat cleaning (in radio-astronomy), Weiner filtering (when implemented as a filter), and is closely related to Backus-Gilbert inversion. In the limit $\theta \to 0$ the map from d to $\hat{\sigma}_{\mathrm{MAP}}$ is the Moore-Penrose inverse, while the case $\theta = 0$ gives $\hat{\sigma}_{\mathrm{ML}}$ which is the least-squares solution first suggested by Gauss in 1880.

Even though equation 1 can be derived within the statistical framework, we make a distinction between this regularized solution, that seeks to recover just a single image independent of whether or not that image is representative of the bulk of feasible images, and statistical methods that base the recovered image on the full set of feasible images.

3.1.2 Examples of ML and MAP Solutions

Figure 5 shows two examples of image recovery calculated using equation 1, taken from a text used at Auckland for a graduate course on inverse problems [13]. The two top sub-images show images that have been blurred by convolution with a known blurring function, and then corrupted by the addition of Gaussian noise. The right-top image has also had 50% of the degraded pixels set to black. In both cases the un-blurred image is estimated using equation 1.

Because the blurring function has no spectral zeros, the forward map for the left-hand example is invertible, and the inverse may be calculated using Fourier deconvolution. The application of the exact inverse, giving the maximum Likelihood estimate, is shown in the left-mid sub-image. It is clear that the exact inverse does not give a useful reconstruction. The reason for the poor performance of the exact inverse is simply because many of the singular values of the forward map are very small, leading to large amounts of noise in the reconstruction. The MAP estimate, with θ chosen using the L-curve criterion [7], is shown in the lift-bottom sub-image and provides a readable reconstruction.

The forward map leading to the right-top sub-image is linear though not invertible since there are half the number of data pixels as in the original image pixels. However, the unblurred image may still be estimated using Equation. 1 and the resulting MAP estimate is shown in the right-bottom with θ chosen using the L-curve criterion.

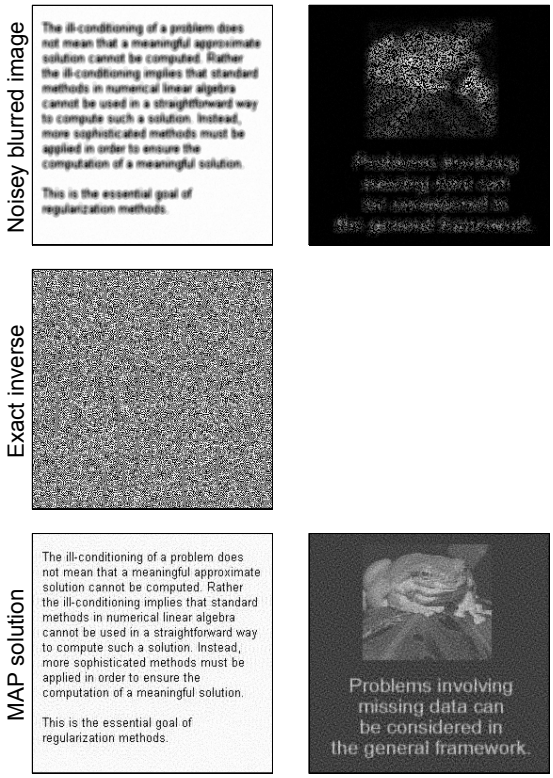

Figure 5: A tableau of images giving two examples of image recovery as the modes of the likelihood and posterior. The top sub images show noisey images. The bottom two subimages show de-blurred images calculated via regularized inversion for a suitable choice of regularizing parameter. The left mid image shows the effect of applying the inverse of the blurring function to the top-left image.

As can be seen, good results can be achieved using regularized inversion, or equivalently the MAP estimate, with suitable choice of regularizing parameter θ. In examples like the ones here, where the variables are continuous, the forward map is linear, and the noise is Gaussian, using a Gaussian prior leads to a symmetric posterior where the mode is also the mean and it is hard to see how the MAP estimate could be improved upon. However, in general these conclusions do not hold for the title problem, as we will see in section 3.1.4.

3.1.3 Inferential Solutions

Since the posterior distribution gives the probability density over all images consistent with the data and prior knowledge, any question about the true image is best answered as an expectations over the posterior distribution. So if $g : \Sigma_\Omega \to \mathbb{R}$ is some function of the image, the expected value of g is

$$\mathrm{E}\left[g\left(\sigma\right)\right] = \sum_{\sigma \in \Sigma_\Omega} \Pr\left(\sigma|d\right) g\left(\sigma\right). \tag{2}$$

For example, if g is an indicator function for an image showing that the patient has cancer, then Equation. 2 gives the probability of cancer based on the measurements and the prior information.

We give a sampling algorithm for computing the expectation in 4.2 and numerical examples in section 4.3

3.1.4 Comparison of mode and mean

Our main point in this section is to show that in high-dimensional inverse problems (like image recovery), the mode of the distribution does not necessarily represent the bulk of feasible solutions and hence does not necessarily give a good approximation to expectations, such as the mean. We next give an argument why the mode need not be a good approximation to the mean, and then an example where the difference is marked.

Consider the schematic of a posterior distribution with two modes shown in figure 6. For the sake of the argument, assume that the maximum value of $\Pr\left(\sigma|d\right)$, which defines the mode, is twice the value of the lower local maximum. Also assume the width, say L, of the global peak is half the width of the lower local peak *in each of the coordinate directions* $\sigma_1, \sigma_2, \cdots$. Then the probability mass (or number of feasible solutions) associated with the peak around the mode is about $P_{\max} L^n$ compared to $(P_{\max}/2)(2L)^n$ for the lower peak. So there is 2^{n-1} times *more* probability mass centered around the lower, broader, peak. For high-dimensional problems where n is large, this metric factor can be huge. Even for a coarse 100×100 pixel image, the lower peak would contain $2^{10000-1} \approx 10^{3010}$ times more feasible reconstructions than the peak around the mode. The function value is virtually irrelevant in comparison to the width of a peak when calculating high dimensional integrals. Note that there did not need to be two peaks for this to occur. If a distribution with a single mode becomes broader as the function value decreases in some direction, then the bulk of the

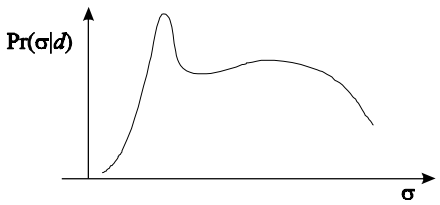

Figure 6: A schematic of a posterior distributiuon with two modes.

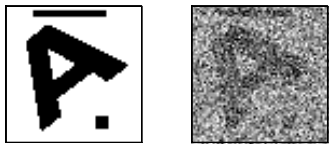

Figure 7: True binary image (left) and gray-scale image showing the image after pixel-wise addition of Gaussian noise (right).

probability mass will lie away from the mode. So we see that there is no reason that the mode of a high-dimensional skewed distribution should have anything to do with the mean of that distribution.

The following is an image-recovery problem in which the mode of the posterior is very different to the mean. Consider the problem of recovering a binary (black and white) image after pixel-wise addition of zero-mean Gaussian noise with known variance. This toy example is chosen because the mode and mean can be exactly calculated [9]. Interested readers should view that paper for further details.

Figure 7 shows the true image (left) and the pixel-wise degraded version (right). We specify a prior distribution by modelling the image on the pixel lattice as an Ising Markov random field, with distribution

$$\Pr(x) \propto \exp\left(\theta \sum_{i=1}^{N^2} \sum_{j \sim i} \delta_{x_i, x_j}\right) \tag{3}$$

where the sum over $j \sim i$ is a sum over all neighbours on the pixel lattice and $\delta_{a,b}$ is the indicator function for the event $a = b$. In Fig. 8 we show the MAP state along with the mean, a sample from the posterior and the marginal posterior mode (MPM), for the smoothing parameters $\theta = 0.125$, 0.25, 0.375, 0.5, 0.625. The MPM shows each pixel as the mode of the marginal distribution of that pixel, and hence takes the value that the pixel most frequently took in the samples drawn from the posterior. For the case of binary images, the MPM is just the thresholded mean, $x_{\mathrm{MPM}} = [x_{\mathrm{mean}}]$.

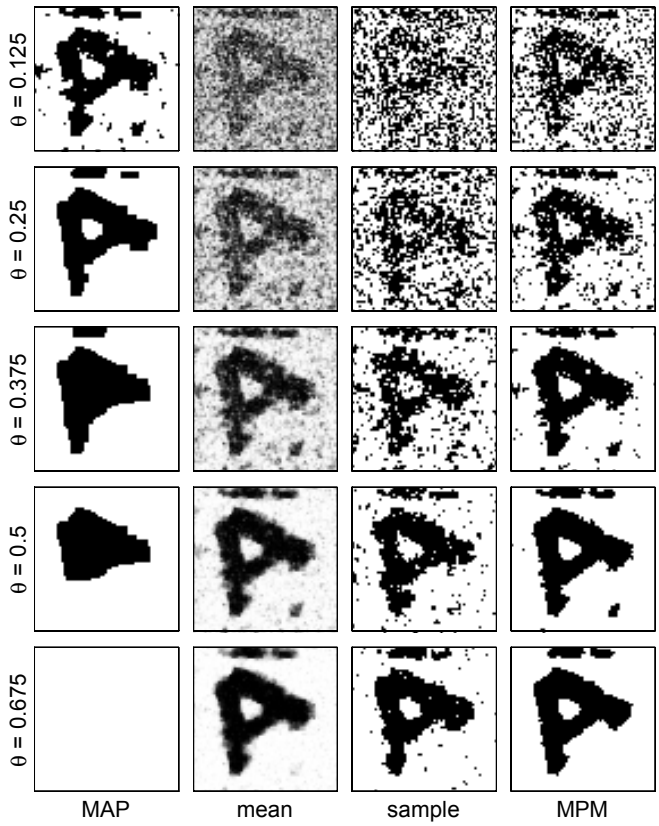

Figure 8: Tableau of maximum *a posteriori* (MAP) state, mean, a single sample from the posterior, and the marginal posterior mode (MPM) for a range of smoothing parameters θ.

Note that as the smoothing parameter θ increases, x_{MAP} becomes smoother and goes from being a reasonable recovered image at $\theta \approx 0.25$, first loosing the centre of the A, then the "legs" and finally, for all θ greater than some critical value in the range $(0.5, 0.675)$, $x_{\mathrm{MAP}} = \mathbf{1}$, i.e. the all-white state. The mean image indicates that the position of the bulk of posterior probability mass, clusters increasingly around better reconstructions as θ increases, with good reconstructions for $\theta \gtrsim 0.5$. While the mean is not a feasible reconstruction for binary images, for large $\theta \gtrsim 0.5$ it is "close" to the MPM which does provide a good recovered image in that range. The sample images show what a typical state from the posterior looks like.

The MAP state never looks like a typical state across the range of θ, and for larger θ is something of an outlier. For θ in the range $(0.125, 0.25)$ x_{MAP} gives a recovered image with the square feature beneath the A being present and with the image not too spotty. But then one needs to choose θ carefully, reflecting the standard practice of having to be particular about the smoothing parameter in regularized inversion. At larger smoothing parameters, $\theta \gtrsim 0.5$ for this example, when the prior is doing an excellent job of shaping the posterior so that the bulk of posterior probability mass contains smooth images that fit the data well and themselves make good reconstructions, the MAP state is a hopeless reconstruction precisely because it is entirely unrepresentative of typical samples. For $\theta \gtrsim 0.675$ this situation is extreme: Then the MAP state is an extreme outlier and is completely useless, while the posterior is dominated by states from which a good recovered image could be formed.

For the case of recovering a pixel-wise degraded binary image, we find that regularized inversion does not give good reconstructions and shows sensitivity to θ precisely because the single most-likely state is unrepresentative of the bulk of feasible images. On the other hand the bulk of posterior probability is clustered around good estimates of the true image and statistics that summarize the posterior can achieve good image recovery that is insensitive to choices in the prior such as our smoothing parameter θ. It is clear that such statistics exist for this example, and we expect also for the majority of image recovery problems. We conjecture that those properties of the MAP estimate occur in other image recovery problems where the mode is not necessarily the mean. In other than very special cases of these high-dimensional non-linear inverse problems, the single most likely image (σ_{ML} or σ_{MAP}) does not give good image estimates. Since the forward map in our case is nonlinear, the posterior distribution will in general be skewed, and so there is no reason that the mode should be sought as an estimate of the coefficients of a PDE. Therefore, in non-invasive imaging from strongly scattered waves it is vital to explicitly quantify the noise process and the resulting set of feasible images, and to produce images based on the set of all feasible images.

4 Conductivity imaging via Statistical Inference

In this section we give a comprehensive example of solving the conductivity imaging inverse problem using statistical inference.

4.1 Bayesian Formulation for Conductivity Imaging

We wish to recover a conductivity image given the current and electrical potential measurements described in section 2. The unknown true currents and potentials are not of direct interest, however the likelihood of any particular conductivity image depends on them, and so we are obliged to make them additional objects of inference. Since the true boundary current and boundary potential are related by the inverse of the Dirichlet-to-Neumann map, Γ_σ, for a given conductivity, σ, we need only parameterize a state by conductivity and one of current or potential. We will use current. Let $\rho^{(n)}$ denote a generic current distribution associated with the n'th measurement set, R the space of generic currents, V and J the space of measured boundary voltages and currents, respectively, so the data space is $D = J, V$.

Let $P(\sigma, d\rho|d) = \Pr(\sigma, d\rho|d)$ denote the posterior distribution for the unknown conductivity and the unknown current. In terms of the likelihood and prior,

$$P(\sigma, d\rho|d) \propto f_{D|\Sigma,R}(d|\sigma,\rho)\Pr(\sigma, d\rho).$$

To compute the likelihood of some conductivity image σ, and unknown true current $\rho^{(n)}$ associated with the n'th measurement set, we must compute the true boundary potential $\phi^{(n)} = \Gamma_\sigma\left(\rho^{(n)}\right)$ and compare $\phi^{(n)}$ with $v^{(n)}$ and $\rho^{(n)}$ with $j^{(n)}$. If noise on current and potential measurements are independent, then their joint likelihood is

$$f^{(n)}_{J,V|\Sigma,R}(j^{(n)}, v^{(n)}|\sigma,\rho^{(n)}) = f^{(n)}_{V|\Sigma,R}(v^{(n)}|\sigma,\rho^{(n)}) f^{(n)}_{J|R}(j^{(n)}|\rho^{(n)})$$

For example, if all noise is i.i.d. Gaussian, with zero mean, and variances s^2_ρ and s^2_ϕ are known,

$$j^{(n)}|\rho^{(n)} \sim \mathrm{N}(\rho^{(n)}, s^2_\rho),$$

$$v^{(n)}|\sigma,\rho^{(n)} \sim \mathrm{N}(\Gamma_\sigma(\rho^n), s^2_\phi),$$

and the likelihood is

$$f_{J,V|\Sigma,R}(j, v|\sigma,\rho) = \prod_{n=1}^{N} f^{(n)}_{J,V|\Sigma,R}(j^{(n)}, v^{(n)}|\sigma,\rho^{(n)}). \tag{4}$$

Some prior distribution must be asserted for σ and ρ. For example, if the conductivity is known to take one of two values throughout a pixel image of

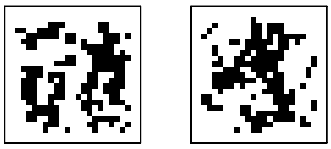

Figure 9: Two samples drawn from the prior distribution for conductivity.

conductivity values, and the current is asserted at E point electrodes, we might represent our prior preference for σ and ρ by the Ising binary Markov random field for σ (c.f. equation 3) and a uniform density on $\mathbb{R}^E$ for the current. Figure 9 show two samples drawn from the prior distribution for conductivity. It can be seen that this prior prefers connected regions of a single value, with the undesirable preference for vertical and horizontal edges. We use this prior as an illustration, only, of what can be achieved using a simple model, and present some more sophisticated models in section 5.

The joint posterior distribution, $P(\sigma, \rho|j, v) \propto f_{J,V|\Sigma,R}(j, v|\sigma, \rho) \Pr(\sigma)$, then gives the basis of all inference about σ (or ρ). To simplify our following presentation of algorithms, we suppose uncertainty in current measurements may be ignored, and the conductivity image is the sole object of inference. The posterior distribution becomes simply

$$P(\sigma|v) \propto f_{V|\Sigma}(v|\sigma) \Pr(\sigma).$$

It is straightforward to remove this assumption within the framework we lay out below.

4.2 Markov chain Monte Carlo

In the following discussion it is assumed that the conductivity at each pixel takes one of a finite set of values. It is straightforward to generalize this treatment to the case of continuously-valued pixel conductivities.

The expectation in equation 2 may be calculated using Monte Carlo integration as follows. If $\{X_t, t = 1, 2, \ldots, n\}$ are distributed according to the posterior distribution, $P(\cdot|v)$, then

$$\mathrm{E}\left[g(\sigma)\right] \approx \frac{1}{n} \sum_{t=1}^{n} g(X_t). \tag{5}$$

The task then is to draw samples from the given posterior $P(\cdot|v)$. The Markov chain Monte Carlo (MCMC) algorithm achieves sampling by generating $\{X_t\}_{t=0}^{\infty}$ as a Markov chain of random variables $X_t \in \Sigma_\Omega$, with a t-step distribution $\Pr(X_t = \sigma|X_0 = \sigma^{(0)})$ that tends to $P(\sigma|v)$, as $t \to \infty$. Thus the algorithm produces a random walk through the space of feasible images with the long-term

probability that the walk will visit a particular image tending to the desired posterior distribution.

The Metropolis-Hastings (MH) algorithm for generating such a Markov chain of states is:

1. given state $\sigma_t$at time t generate candidate state σ' from a proposal distribution $q\left(\sigma'|\sigma_t\right)$
2. Accept candidate with probability

$$\alpha\left(\sigma'|\sigma_t\right) = \min\left(1, \frac{P(\sigma'|v)q\left(\sigma_t|\sigma'\right)}{P(\sigma_t|v)q\left(\sigma'|\sigma_t\right)}\right)$$

3. If accepted, $X_{t+1} = \sigma'$ otherwise $X_{t+1} = \sigma_t$
4. Repeat

The proposal distribution $q\left(.|\sigma_t\right)$ can be any distribution that ensures the Markov chain $X_t, t = 0, 1, 2\ldots$ is irreducible and aperiodic. Since, by construction, the chain is reversible with respect to $P(\sigma|v)$, it is ergodic with equilibrium $P(\sigma|v)$. In practice the efficiency of an MCMC algorithm depends on the proposal distribution. In the next section we specify the proposal distribution that we have used for conductivity imaging.

4.2.1 Three-Move Metropolis Hastings

At the proposal step of the MH algorithm, choose one of 3 "moves" with probability ζ_p, $p = 1, 2, 3$. Each move specifies a particular way of modifying the current state to produce a candidate state. Depending on which move is chosen, a proposal distribution $q_p(\sigma'|\sigma)$ is determined. The corresponding acceptance probability is given by setting $q = q_p$ in the above expression for $\alpha(\sigma'|\sigma_t)$. The algorithm thereby defines three transition probabilities $\Pr^{(p)}(X_{t+1} = \sigma_{t+1}|X_t = \sigma_t), p = 1, 2, 3$ corresponding to the three possible stochastic update types. These transition probabilities are reversible w.r.t. $P(\sigma|v)$. The overall transition probability is

$$\begin{aligned}&\Pr(X_{t+1} = \sigma_{t+1}|X_t = \sigma_t)\\ &\qquad = \sum_{p=1}^{3} \zeta_p \Pr{}^{(p)}(X_{t+1} = \sigma_{t+1}|X_t = \sigma_t).\end{aligned}$$

If at least one of the moves is irreducible on Σ_Ω, then the equilibrium distribution is $P(\sigma|v)$.

We have found the following set of moves effective for MCMC in an EIT setting. A pixel n is a *near-neighbour* of pixel m if their lattice distance is less than or equal to $\sqrt{8}$. An *update-edge* of conductivity image σ is a pair of near-neighbouring pixels of unequal conductivity.

Move 1 *Flip a pixel.* Select a pixel m at random and assign σ_m a new conductivity σ'_m chosen uniformly at random from the set of allowed conductivity values.

Move 2 *Flip a pixel near a conductivity boundary.* Pick an update-edge at random from the set of current update edges. Pick one of the two pixels in that edge at random, pixel m say. Proceed as in Move 1.

Move 3 *Swap conductivities at a pair of pixels.* Pick an update-edge $\langle m, n \rangle$ at random from the set of update edges of the current state. Set $\sigma'_m = \sigma_n$ and $\sigma'_n = \sigma_m$.

4.3 Numerical Experiments

In this section we provide some numerical examples of calculating expectations over the posterior distribution for conductivity imaging, using simple synthetic data sets. The three examples correspond to different models for the conductivity. The examples come from Nicholls and Fox (1998) and interested readers are directed to that paper for computational details. In each example we start with a known conductivity distribution defined on a 25×25 pixel lattice, simulate measurements at 16 electrode locations inside a strip of width 2 pixels neighbouring the boundary, and then use MCMC with the forward map simulated by solving the PDE on the pixel lattice to sample the posterior.

Figure 10 shows an experiment in which the true conductivity at each pixel takes one of three values, and the same model is used for the recovered image. Subfigure A shows the true conductivity. Subfigures B1 and B2 show samples taken from the posterior distribution, generated by the Markov chain after burn in. Because the states resemble the true image, it is reasonable to conclude that the chain has reached equilibrium. Each of these samples correctly recovers the main features of the true conductivity, in particular the background and the two inclusions have the correct conductivity though there is some ambiguity in the location of boundaries. The mean image, shown in subfigure C, is a good estimator for the true conductivity – though it is not actually in allowable image space. Subfigure D shows the pixel variance, with darker shades indicating higher variance. This image shows that the background conductivity is essentially determined while the diagonal boundary of the lower inclusion is not well positioned.

Figure 11 shows a more realistic conductivity model that has two levels. Each pixel is taken as coming from one of three types of material with each material having a range of allowable conductivities. The material type is therefore equivalent to the conductivity model used in the previous example, but the conductivity is not three valued. Subfigure A shows the true conductivity. Subfigures B and C show a sample taken from the posterior distribution, where B gives the material type and C gives the sampled conductivity based on B. The mean conductivity image is shown in subfigure C, while subfigure D shows the conductivity variance, with darker shades indicating higher variance. Again

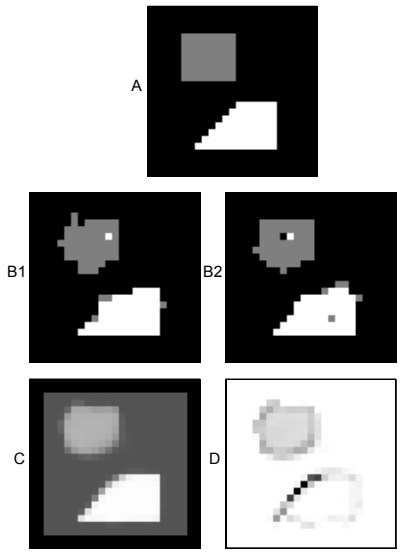

Figure 10: Recovering conductivity information for a synthetic data set with three conductivity levels, computed from A. The simulation uses 16 electrodes, with 4 evenly spaced on each side. Conductivities 1, 2 and 3 are black grey and white respectively. The states in B1 and B2 are samples from the posterior, the mean conductivity is shown in C, and pixel conductivity variance is shown in D, where darker shades indicate greater variance.

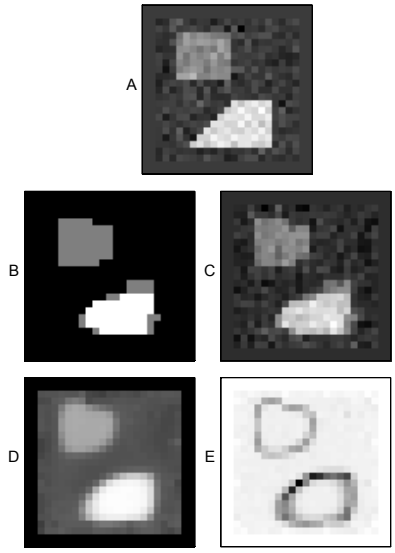

Figure 11: Recovering conductivity information for a synthetic data set, computed from A with continuous conductivity pixel variables. Lighter shades indicate higher conductivity. A two-level model is used consisting of a material type model and a distribution of conductivity for each material type. The state in B gives sampled reconstructed material type classifications while C gives sampled reconstructed conductivity values. Mean conductivity values are shown in D, and pixel conductivity variance is shown in E.

we see that the mean is a reasonable reconstruction with the ambiguity in the reconstruction quantified by the variance image.

Figure 12 shows that certain conductivity patterns can screen internal structure. The true conductivity patterns in subfigures A and B are similar except that in A the conductivities going inwards go medium-low-high (i.e. 2, 1, 3) while in B the sequence is low-medium-high (i.e. 1, 2, 3).The lack of contrast in the interior makes B harder to recover, which corresponds to the high variances displayed in subfigure H.

5 Some image models

The nature of the MCMC algorithm makes it feasible in principle to use image models, or parameterizations, that give image spaces that are discrete, or disconnected, or even variable dimension. This allows a substantially broader class of image models than can be handled by gradient-based optimizations traditionally used to implement regularized inversion. Since image representation and specifying prior information are two faces of the same process, a signifi-

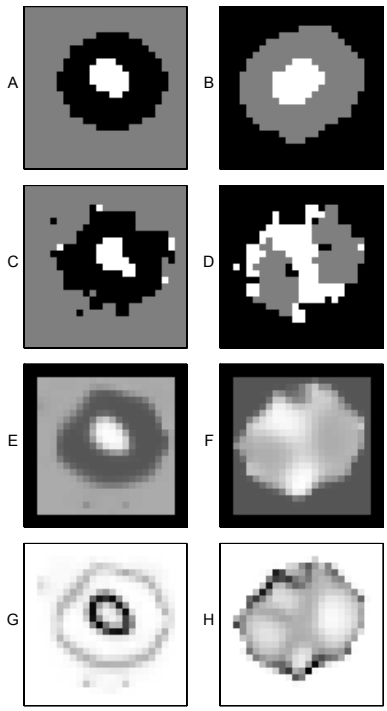

Figure 12: Indications that shielding leads to increased uncertainty in sampled reconstructed conductivity. Recovering conductivity information for a synthetic data set with three conductivity levels. Subplot A gives true conductivity for C,E, and G, while B gives true conductivity for D, F, and H. A, B: true conductivity. C, D: sample reconstructed conductivity from posterior. E, F: mean conductivity. G, H: pixel conductivity variance.

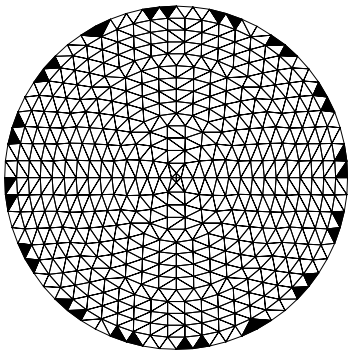

Figure 13: A finite-element grid for a circular region showing the elements used to model electrode placement.

cant advantage afforded by the richer choice of image models is that relatively natural prior image information may be specified. For example it is relatively straightforward to restrict allowable images to ones with a constant background containing blob-like regions, and even restrict the number of blobs – though there are difficulties in achieving a computationally efficient algorithm. The ability to specify this type of prior information gives a substantial improvement over the mathematically-expedient measures of image quality typically used in regularization algorithms.

5.1 Finite-size electrodes

Imaging electrical conductivity from actual measurements requires accurate modelling of the measurement process including details of electrode geometry and contact. In this first example we show that a detailed finite-element model may be used within an MCMC calculation. Figure 13 shows a finite-element discretization used for imaging within a circular region, with the elements used to model finite electrodes coloured black. In this example the circular region has radius 1, the synthetic data was generated from an image with constant background conductivity and a circular inclusion, having radius 0.1 centered at coordinates (0,0.6), with twice the background conductivity. The image was modelled as having one of the two known conductivity levels within each element, with an Ising-type prior, and the likelihood was calculated by solving the PDE on the same finite-element mesh. Figure 14 shows output traces of the sample prior and likelihood when the Markov chain was run from the initially constant, background, conductivity, while figure 15 shows a sample from the chain after 4650 steps, when the chain appears to be in equilibrium. It is interesting to note that the likelihood converges to its equilibrium distribution more slowly than does the prior, showing that there are many states consistent with the data that do not make good reconstructions, in the sense of the prior.

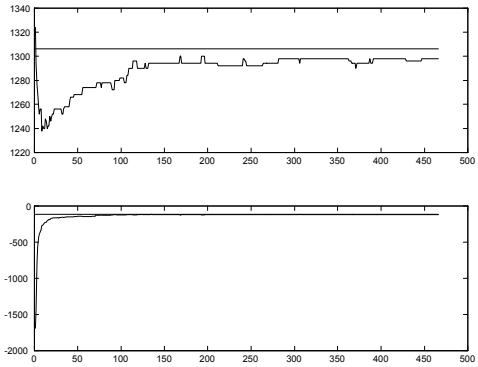

Figure 14: Sample prior (top graph) and Likelihood. The horizontal lines give the value of the prior and Likelihood of the true image.

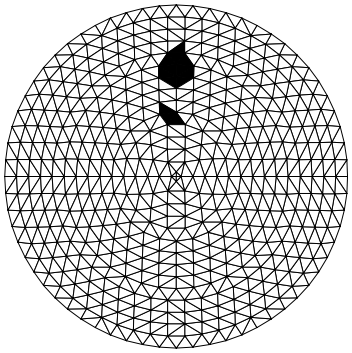

Figure 15: Sample number 46500

The sample in figure 15 indicates that the circular inclusion has been found though the spurious elements set to the higher conductivity shows that it is not unambiguously defined.

5.2 Higher-level models

In each application area, the conductivity in the imaged region will typically have some predictable form. For example, if we makes images of oranges we may expect to see pips. How might this prior knowledge be used to inform the reconstruction? In the Bayesian inferential framework the place for knowledge of this kind is the prior probability distribution over some appropriate space of conductivity functions. How should we define this space? The parameterization is most important, since the information we possess will be more easily represented in some parameterizations than others.

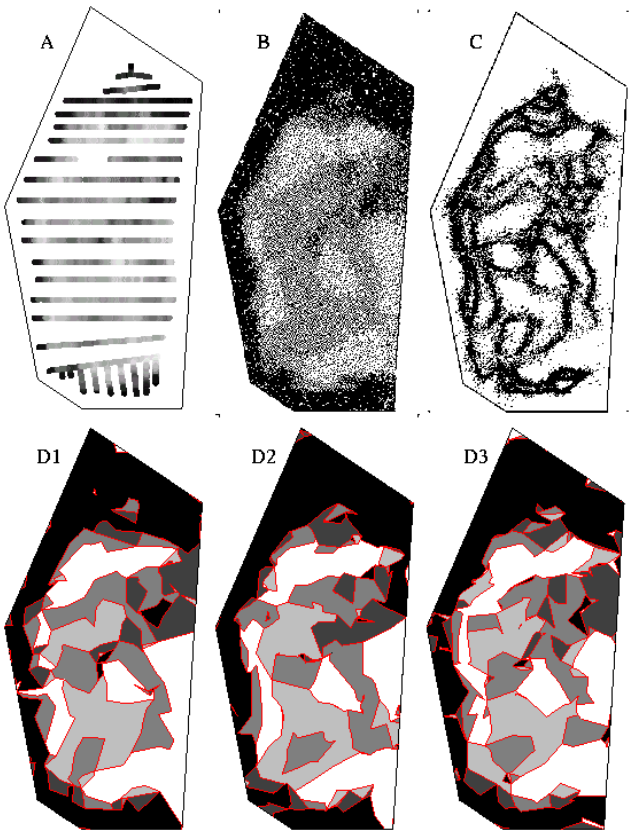

Figure 16: Intermediate level model: (A) data, 1746 resistivity readings over neolithic hill fort, (B) posterior mean resistivity, (C) posterior edge length density, (D1-3) samples from posterior.

In the language of the pattern recognition literature (Marr, 1982, Grenander 1993) a pixel based representation of the conductivity image is a *low level* parameterization. It is easy to represent local prior knowledge in this parameterization. For example we can write down a probability distribution favouring locally smooth pixel-based images such as the Ising Markov random field.

An image representation based on regions bounded by variable polygonal boundaries, which meet at vertices located in the plane continuum, is an example of an *intermediate level* representation. In this example we should decide on a set of basis functions to give the conductivity function over the polygonal faces. Nicholls, (1998) gives a probability distribution on such random face-coloured polygonal graphs, with piecewise constant face colouring, and uses it in a Bayesian analysis of measured resistivity data. Figure 16 is taken from that paper. It is easy in such an intermediate level parametrization to express con-

straints on the topology of the graph of boundaries, which would be inconvenient in a pixel-based representation.

Examples of *high level* parameterizations are given by Grenander and Miller (1994) and Baddeley and van Lieshout (1993) who consider "object" processes. The image model is a composition of unions of templates each representing entire objects and having an associated group of allowable transformations or deformations.

6 Summary

The Bayesian inferential framework provides a comprehensive analysis of inverse problems including consideration of the forward map, measurement process and measurement error. Image modelling and inclusion of prior knowledge are achieved through specifying a prior probability density over the image parameterization. The posterior density, being the product of likelihood and prior, contains all information about the unknown image and is the sole focus of inference.

Expectations over the posterior can be calculated using the MCMC algorithm, giving a solution to the inverse problem (up to algorithmic efficiency). Since the MCMC algorithm uses simulation of the forward map to evaluate the likelihood (and not inverses of the forward map), if you can simulate the forward map then you can sample and calculate expectations over the posterior, i.e., 'solve' the inverse problem.

In high-dimensional inverse problems that are discrete and/or non-linear, such as recovering the coefficients of a PDE from boundary measurements, there is no reason that the mode of the posterior – traditionally reported in regularized inversion – should represent the bulk of feasible images. We gave an example where this single most-likely image is not a good reconstruction and is an outlier in the set of feasible images. Hence in image recovery problems such as estimating the coefficients of a PDE, it is necessary to calculate images that summarize the bulk of feasible images. This is the reason that the extra machinery of Bayesian inference is required compared to, say, the existing optimization-based regularization methods. Which summary statistics give best reconstructions and the details of efficient algorithms remain open questions. However, it seems likely that Bayesian inference will ultimately give comprehensive solutions to the class of title problems.

Le Cam, the Berkeley satistician, purportedly dismissed inverse problems by observing that: the forward problem is probability, and the inverse problem is statistics. We concur with the analysis, but not the sentiment. There remains much work to be done to make Bayesian inference into a robust off-the-shelf technique for solving real imaging problems.

References

[1] U. Grenander. *General Pattern Theory.* Oxford University Press, 1993.

[2] D. Marr. *Vision.* W.H. Freeman and Co, San Franscico, 1982.

[3] G. K. Nicholls, "Bayesian image analysis with Markov chain Monte Carlo and coloured continuum triangulation mosaics," *Journal of the Royal Statistical Society B* **60**, pp. 643–659, 1998.

[4] U. Grenander and M. Miller. Representation of knowledge in complex systems. *J. R. Statist. Soc. B*, 56(3), 1993.

[5] A.J. Baddeley and M.N.M van Lieshout. Stochastic geometry models in high level vision. In KV Mardia and GK Kanji, editors, *Statistics and Images, Vol 1*, volume 20 of *J. Applied Statistics*, pages 231–256. 1993.

[6] U. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging.* Institute of Physics Puplishing, Bristol and Philadelphia, 1998.

[7] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems.* Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[8] C. Fox, Conductance Imaging, PhD thesis, Cambridge University, 1988.

[9] C. Fox and G. K. Nicholls, "Exact MAP States and Expectations from Perfect Sampling: Greig, Porteous and Seheult revisited," In A. Mohammad-Djafari editor, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIP conference proceedings volume 568, pp 252-263, 2001.

[10] G. K. Nicholls and C. Fox, "Prior Modelling and Posterior sampling in Impedance Imaging," In A. Mohammad-Djafari editor, *Bayesian Inference for Inverse Problems*, SPIE conference proceedings volume 3459, pp 116-127, 1998.

[11] C. Fox and G. K. Nicholls, "Physically-based likelihood for ultrasound imaging," In A. Mohammad-Djafari editor, *Bayesian Inference for Inverse Problems*, SPIE conference proceedings volume 3459, pp 116-127, 1998.

[12] K. E. Andersen, S. P. Brooks and M. B. Hansen, "Bayesian Inversion of Geoelectrical Resistivity Data", to appear, JRSS B.

[13] S. M. Tan, C. Fox and G. K. Nicholls, "Inverse Problems" Course Text in Electronic Form available at http://www.phy.auckland.ac.nz/Staff/smt/453707SC.html

[14] E. T. Jaynes, "Prior Probabilities" IEEE Transactions on Systems Science and Cybernetics, SSC-4, Sept. 1968, 227-241.

[15] J. M. Smith and A. F. M. Smith, "Bayesian Theory", John Wiley & Sons, Chichester, 1994.

[16] B. Pavlov, "S-matrix and Dirichlet-to-Neumann operators", In:Scattering: Scattering and Inverse Scattering in Pure and Applied Science, vol 2, Eds R. Pike and P. Sabatier, Academic Press, San Diego, 2002.