

A Gibbs sampler for conductivity imaging and other inverse problems

Colin Fox

Electronics Group, Physics Department, University of Otago, PO Box 56, Dunedin,
New Zealand

ABSTRACT

Gibbs samplers have many desirable theoretical properties, but also have the pesky requirement that conditional distributions be available. We show how conditional densities can be evaluated for the posterior distribution in conductivity imaging - virtually for free in coordinate directions and very cheaply in other 'special' directions. The analysis actually applies to a broad class of non-invasive imaging techniques that utilize strong scattering of energy, and leads to efficient iterative algorithms whether implementing inference or optimization. The resulting Gibbs sampler draws an independent conductivity image in only a little more compute time than required for optimization.

Keywords: Gibbs sampling, conductance imaging, EIT, inverse problems, Bayesian inference, uncertainty quantification

1. INTRODUCTION

The observation that Gibbs samplers are essentially identical to stationary iterative linear solvers (see Refs. 1 and 2 for details and further references) has given a new impetus to investigating Gibbs sampling as a route to performing computational inference in inverse problems. Stationary linear iterative samplers developed in the 1950's are now considered to be very slow, because of their geometric rate of convergence, and many acceleration techniques have been developed that are now commonplace. These include the polynomial acceleration techniques introduced in the 1960's, and the Krylov space methods that are now amongst the state-of-the-art for solving large-scale problems. Both of those acceleration techniques can be applied to Gibbs sampling,^{1,2} hence the renewed interest. Here we discuss the design of an efficient Gibbs sampler for use in imaging from strong wave scattering, that can then be accelerated using those techniques.

1.1 Imaging from Wave Scattering

Non-invasive imaging typically uses scattering of waves to probe the object under investigation. These techniques are non-invasive when the wave scattering is measured remotely from the object, and the object is irradiated by a source remote to the object.

A well-known example is x-ray imaging where an object is irradiated from one side using a point x-ray source and the partially-transmitted waves are measured on the opposite side to the source. This presents a relatively simple imaging problem (at least in the idealized case of a narrow-band source) since the path of propagation is unaffected by the unknown spatially-varying material properties being imaged. Then the forward map may be modelled as a *fixed* (linear) operator. This feature of x-ray imaging is related to the high energy, ionizing, nature of x-rays which as a side-effect causes damage to the material through which it propagates – or more correctly by which it is absorbed.

In this paper we are interested in cases where the scattering within the object is strong, i.e., where the path that the waves take is significantly affected by propagation through the object. This occurs for low-energy, non-ionizing, radiation which is a less-destructive source of waves for the purpose of imaging. However, the path that each ray takes is implicitly determined by the propagation medium leading to a non-linear imaging problem in contrast to imaging from weak scattering where the forward map is linear. However, when the material properties

Further author information: C.F.: E-mail: fox@physics.otago.ac.nz, Telephone: +64 3 479 7806

Image Reconstruction from Incomplete Data VII, edited by Philip J. Bones, Michael A. Fiddy,
Rick P. Millane, Proc. of SPIE Vol. 8500, 850006 · © 2012 SPIE
CCC code: 0277-786/12/\$18 · doi: 10.1117/12.931111

are not affected by the wave field, then the forward map is governed by an operator that, while not constant, is a *linear function* of the unknown material properties. We consider that case in detail, and show how it leads to significant efficiencies when imaging is performed iteratively.

Three examples of imaging from wave scattering, corresponding to three different types of energy being propagated, are given in the following table.

quantity being imaged	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 quantity being imaged is denoted by $\sigma(x)$ and appears as the spatially-varying coefficient in the PDE governing the propagation of energy, or ‘waves’. The measurements are made of the boundary values of electrical potential ϕ , ultra-sound pressure p , and temperature u , respectively in the three cases. In each case the measurement process is simulated by solving the PDE subject to boundary conditions that correspond to the wave irradiation.

The simplest case is imaging electrical conductivity, though it shares with the others the space part of the PDE which is an elliptic, Laplacian-like, operator which depends *linearly* on σ . In fact, the PDE is an example of a $A^T C A$ system³ where A is a ‘matrix’ that contains geometric information only, and C is a diagonal matrix of material properties. In this case, $A \equiv \nabla$ and $C \equiv \text{diag}(\sigma)$ is a ‘matrix’ with the conductivity values on the diagonal. As Strang³ explains, the equilibrium equations of many physical systems can be put in this form.

The finite-element method (FEM) discretization of the PDE inherits these properties. Then A and C are discrete analogs of the infinite-dimensional operators in the PDE, and the matrix $K = A^T C A$ is called the *stiffness matrix*. Through the remainder of this paper we consider a naturally discrete problem of this type. However, all the analysis performed also holds for the PDE case, though with greater technical requirements.

1.2 Bayesian Formulation and Gibbs Sampling

In a given non-invasive imaging setup, the externally applied wave field corresponds to asserting boundary conditions on the governing PDE, while measurements correspond to values of the field at fixed locations. We denote the resulting *forward map* from unknown σ to measured data d by $F : \sigma \mapsto d$. For the systems considered above, this data simulation requires solving a system of the type $K\phi = f$ for ϕ , where f is a forcing vector, with data d given by some components of ϕ . The stiffness matrix K depends linearly on σ via $K = A^T C(\sigma)A$, giving a non-linear inverse problem for σ given d .

Strictly, the forward map F defines the mapping to *noise-free* data. In practical circumstances the measured data is subject to measurement errors, and the mathematical model for the forward map is subject to modeling errors. A general stochastic model for the practical measurement process is

$$d = G(\sigma, v) \quad (1)$$

where σ is the unknown physical properties, and v is a random variable accounting for variability between ‘identical’ experiments. In the simplest formulation the stochastic part is attributed to measurement error that is additive and independent of σ so that

$$G(\sigma, v) = F(\sigma) + v$$

where $v \sim \pi_n(\cdot)$ comes from the *noise* distribution. Then, when the forward map is treated as certain, the distribution over data conditioned on σ is given by

$$\pi(d|\sigma) = \pi_n(d - F(\sigma)). \quad (2)$$

The likelihood function for given data d is the same function considered as a function of the unknown variables σ , i.e., $l(\sigma|d) = \pi(d|\sigma)$. Hence, formulating the likelihood function requires modeling the forward map as well as the distribution over measurement errors. Evaluation of the likelihood function requires simulation of the forward map, and hence is typically computationally expensive.

Given measurements d , the focus for inference, at least in parameter estimation, is the posterior distribution given by Bayes' theorem

$$\pi(\sigma|d) = \frac{l(\sigma|d)\pi(\sigma)}{\pi(d)} \propto l(\sigma|d)\pi(\sigma) \quad (3)$$

where $\pi(\sigma)$ is the prior distribution and $\pi(d)$ is often called the evidence. The most frequently used posterior density has the form

$$\pi(\sigma|d) \propto \exp \{-\chi(d - F(\sigma)) - \rho(\sigma)\} \quad (4)$$

where χ and ρ are relatively simple functions. For example $\chi(y) = y^T \Sigma^{-1} y / 2$ when the noise comes from a Gaussian process with known covariance matrix Σ , while ρ is an energy function when the prior model for σ is a Markov random field (MRF). Regularized inversion is the same as maximizing this functional (with fixed hyperparameters), leading to the common erroneous assertion that regularization is a Bayesian method.

The Markov chain Monte Carlo (MCMC) methods for exploring the posterior distribution use sampling algorithms to draw (random) samples $\sigma^{(k)}$ from the posterior distribution. Then, the mean of *any* function $g(\sigma)$ of σ can be estimated by the sample average

$$E(g|d) = \int g(\sigma) \pi(\sigma|d) d\sigma \approx \frac{1}{N} \sum_{k=1}^N g(\sigma^{(k)}) . \quad (5)$$

Furthermore, the law of large numbers guarantees that (under quite mild assumptions) the variance of the sample average behaves like $\propto N^{-1}$.

Perhaps the simplest sampling algorithm is the Gibbs sampler⁴⁻⁶ that repeatedly samples from (block) conditional distributions of π . We consider the simplest, and original, version of Gibbs sampling in which one iteration consists of conditional sampling along each coordinate direction in sequence. This algorithm is very simple:

ALGORITHM 1 (GIBBS SAMPLER). *At state $\sigma^{(n)} = \sigma$, simulate $\sigma^{(n+1)}$ as follows:*

1. *Select one component σ_i .*
2. *Update σ_i by sampling from the conditional distribution for σ_i , i.e. $\pi(\sigma_i|\sigma_{-i}, d)$.*

Here σ_{-i} denotes all components of σ other than the i^{th} . Performing the selection in step (1) in sequence gives the standard Gibbs sampler, whereas selecting i at random gives the *random scan* Gibbs sampler.

If instead of conditional sampling in step (2) we performed conditional optimization, Alg. 1 would give the Gauss-Seidel optimization algorithm. This connection is explored in Ref. 2 and underlies the application of acceleration techniques that were developed for computational optimization.

Efficient implementation of a Gibbs sampler relies on the conditional distributions being available in some convenient form that makes drawing samples simple. The ideal situation is where the conditional distributions have a simple analytic form that allows sampling via the inverse cumulative distribution function method, or by some other transformation method. We have not been able to implement these methods for EIT, however we have the next best thing which is to be able to evaluate conditional densities cheaply. Considering Eqn. 4, and assuming that the prior density is cheap to evaluate, the expense of evaluating conditional densities is set by the cost of evaluating the forward map F along coordinate directions. That is, when at state $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ we wish to evaluate

$$F(\sigma_1, \sigma_2, \dots, \sigma_{i-1}, \sigma_i + \Delta, \sigma_{i+1}, \dots, \sigma_n) \quad (6)$$

for any change Δ to the i^{th} component σ_i .

A number of efficient conditional sampling algorithms are available in the case that the conditional density may be evaluated cheaply. When the conditional distribution is known to be log-convex, one can use the adaptive direction sampler,⁷ or improved versions.⁸ Unfortunately the conditional distributions in EIT are not always log-convex so instead we employ multiple steps of a Metropolis-Hastings algorithm that has small computational cost compared to other parts of the algorithm.

2. IMAGING AN ELECTRICAL NETWORK

We consider recovering the positive resistance values in a network of resistors, from noisy measurements made at the boundary of the network. We choose this example because the resulting computational problem allows us to concisely demonstrate the development of a Gibbs sampler, while having the same structure as the FEM discretized equations for electrical impedance tomography (EIT). Indeed, any FEM system can be written as an equivalent network problem, in which a 'resistance' connects adjacent nodes in the FEM mesh with the value given by the elements neighbouring the edge.

2.1 Nodal Equations

In a network of resistors, it is usual to think of resistors occupying the edges of an undirected graph. Consider the case where there are n nodes, numbered $1, 2, \dots, n$. A resistor (if any) between nodes l and m is denoted $r_{(l,m)}$, and the set of such edges is denoted D . We will find it more useful to work in terms of conductances

$$\sigma_{(l,m)} = \begin{cases} 1/r_{(l,m)} & (l,m) \in D \\ 0 & \text{otherwise} \end{cases}$$

is the conductance between nodes indexed by l and m .

Figure 1 shows a resistor network, in this case a square lattice of resistors, when there are $N = 4$ resistors per side. Also shown is case when the lower right node is taken as the electrical reference node, and a current is being injected into one boundary node (and removed from the reference node).

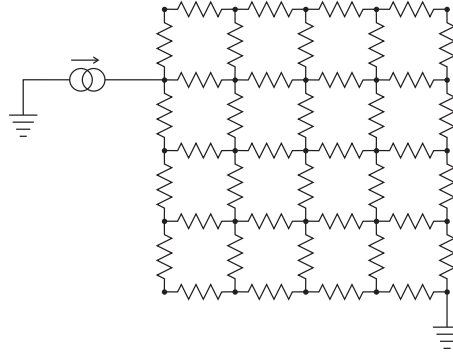


Figure 1. The resistor network for the case $N = 4$. Also shown are the reference node and current being injected at one node on the boundary.

The vector of voltages $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$ at nodes is related to the vector of currents injected into nodes $\mathbf{i} = (i_1, i_2, \dots, i_n)^T$ (and removed from the reference node) by a combination of Ohm's and Kirchoff's laws. To form the system (admittance) matrix Y we first arbitrarily assign a '+' and '-' end to each resistor, then form the $k \times n$ geometry matrix

$$A_{lm} = \begin{cases} +1 & \text{'+' end of resistor } l \text{ is connected to node } m \\ -1 & \text{'-' end of resistor } l \text{ is connected to node } m \\ 0 & \text{else} \end{cases} \quad (7)$$

then

$$Y = A^T C A \quad (8)$$

where C is the $k \times k$ matrix with conductances on the diagonal. The nodal equations are then

$$Y \mathbf{v} = \mathbf{i}. \quad (9)$$

subject to the reference voltage that we write as

$$\mathbf{v} \cdot \mathbf{c} = 0 \quad (10)$$

for some vector $\mathbf{c}$. For computational purposes it is noteworthy that Y is sparse, symmetric, and positive semi-definite when all $r_{(l,m)} \in (0, \infty)$.

2.2 Forward Map

Electrical measurements are made on the network by injecting currents into nodes and measuring the resulting nodal voltages. To maintain the parallel with EIT as a non-invasive imaging technique, we restrict the nodes used for current injection or voltage measurement to a subset of nodes on the boundary – though in reservoir modeling applications measurements are often made at a set of internal nodes. Let $E \subset \{1, 2, \dots, n\}$ be the set of boundary nodes used for injecting current or measuring voltages. We refer to these nodes as electrodes. At all other nodes $m \in \{1, 2, \dots, n\} \setminus E$ current is conserved, i.e. $i_m = 0$, and the voltage v_m is unknown.

Since nodal voltages $\mathbf{v}$ are a linear function of applied currents $\mathbf{i}$, all possible measurements are made by applying the (usual) basis of current vectors at electrodes $\mathbf{i}^k = e_k$ for $k \in E$, and measuring the resulting voltage at all electrodes, v_l for $l \in E$. Hence noise-free measurements consist of the block of the inverse of the admittance matrix $\{Y_{l,k}^{-1} : l, k \in E\}$, which we denote using the (Matlab-like) notation Y_{EE}^{-1} . The inverse problem may be stated as: given a block of the inverse of the reduced admittance matrix, is it possible to recover the full admittance matrix of the form in equation 8 and hence the resistor values?

Calculating the forward map $A : \mathbf{r} \mapsto Y_{EE}^{-1}$ requires solving the matrix equation in (9) for each current vector $\mathbf{i}^k$, a total of $|E|$ solutions, where $|E|$ denotes the number of electrodes. For typical networks containing thousands of resistors, and tens of electrodes, this step dominates computational cost of evaluating the posterior density.

3. EFFICIENT CONDITIONAL UPDATING

The system equations 9 and 10 present a difficulty for manipulation as the system matrix Y is singular. However, a modified form of these equations, that has the same solution, allows working with a modified system matrix that is non-singular. The non-singular form is then suitable for direct application of the Woodbury formula, that gives an explicit expression for the forward map when the state σ is updated. This expression simplifies for $A^T C A$ systems in a way that allows practical computation, particularly in coordinate direction.

3.1 Non-singular matrix form

When the resistor network is connected, and all resistors have positive values, it follows that the admittance matrix has just one zero eigenvalue corresponding to the constant vector. This null space is precisely resolved by asserting a reference voltage, as in Eqn. 10, to give a unique solution. It follows that the modified system equations

$$(Y + \mathbf{c}^T \mathbf{c}) \mathbf{v} = \mathbf{i} \quad (11)$$

give the same solution, without need for a constraint. (This is a kind of *penalty method* as used in BVPs, however it is not the usual type as no Lagrange multiplier is needed, and the solution is exact.) The modified system matrix

$$K = Y + \mathbf{c}^T \mathbf{c} \quad (12)$$

is non-singular, and retains the desirable properties of Y , such as being expressible in the form $A^T C A$ where now matrices A and C are augmented by the vector $\mathbf{c}$ and the diagonal value 1, respectively.

The system matrix K , unlike Y , is invertible. Any efficient numerical method, that exploits the fact that K is symmetric positive definite, may be used to solve the system in Eqn. 11.

3.2 Updating via the the Woodbury Formula

Consider now a change to the conductances, $\Delta\sigma$, occurring to the current set of conductances σ . The (modified) system matrix for σ is $K = A^T C A$, and the updated matrix for $\sigma + \Delta\sigma$ is

$$K + A^T C_\Delta A \quad (13)$$

where C_Δ is a diagonal matrix with $\Delta\sigma$ on the diagonal.

Writing

$$U = A^T \quad (14)$$

and

$$W = A^T C_\Delta A \quad (15)$$

allows the inverse of the updated matrix to be written using the Woodbury formula

$$(K + U W^T)^{-1} = K^{-1} - \left[K^{-1} U (I + W^T K^{-1} U)^{-1} W^T K^{-1} \right] \quad (16)$$

$$= K^{-1} - \left[K^{-1} A^T (I + C_\Delta A K^{-1} A^T)^{-1} C_\Delta A K^{-1} \right]. \quad (17)$$

As we have shown previously,⁹ the outer terms K^{-1} on the right-hand side do not mean that we need to calculate the inverse of K since all terms required for operating by K^{-1} are available once the solution at state σ has been calculated. However, the inner inverse

$$(I + C_\Delta A K^{-1} A^T)^{-1} \quad (18)$$

requires further consideration. For this we expand K^{-1} as

$$K^{-1} = J_1^{-1} C^{-1} J_2^{-T} \quad (19)$$

for some matrices J_1 and J_2 that are not uniquely determined. This transforms Eqn 18 to

$$(I + C_\Delta A J_1^{-1} C^{-1} J_2^{-T} A^T)^{-1}. \quad (20)$$

The expression $K K^{-1} = I$ gives the requirements

$$A^T C A J_1^{-1} C J_2^{-T} = I \quad \text{and} \quad J_1^{-1} C^{-1} J_2^{-T} A^T C A = I. \quad (21)$$

which suggests the simplification when

$$A J_1^{-1} = I \quad \text{and} \quad J_2^{-T} A^T = I$$

where notionally I is the appropriate identity in each expression, though needs to be interpreted with some care. The matrix J_2 may be formed uniquely as the Moore-Penrose generalized inverse of A , while J_1 is not actually needed as the term $A J_1^{-1}$ may simply be canceled. This the required inverse

$$(I + C_\Delta C^{-1} J_2^{-T} A^T)^{-1}, \quad (22)$$

where J_2 is known, is of size $r \times r$ when r resistor values are changed. This expression offers remarkable performance when r is small, with the case of coordinate directions ($r = 1$) allowing evaluation of updated solutions for virtually no cost.

In terms of the Gibbs sampler, this cheap evaluation allows a Metropolis within Gibbs algorithm to be implemented efficiently. However, we find that simply using the Woodbury formula to ‘dead reckon’ the solution for many steps leads to numerical instability. Therefore we do periodically recalculate the solution using a standard matrix solution of Eqn 11.

4. DISCUSSION

We have laid out a formalism for the efficient conditional evaluation of the posterior distribution of the conductance imaging problem. In particular we have outlined a solution algorithm that ensures that the steps in a Gibbs sampler requiring solution of the BVP can be performed rapidly. As far as we are aware, correctly evaluated Gibbs sampling has not been previously developed for EIT. While Gibbs sampling does not necessarily give a fast method for computational inference, the sampler is now amenable to acceleration techniques that we expect will produce an fast computational route to performing inference in the EIT inverse problem. The analysis we have given shows that the same computational structure will also lead to efficient algorithms for other inverse problems where the forward map requires solution of a PDE.

Experiments using synthetic data have performed as expected, and shown that this algorithm can be used for solving the inverse problem in that case. Whilst this is cause for satisfaction, it remains to test this algorithm using real data where modeling complexities also produce more computational complexity in terms of the evaluation of the forward map and the prior model. However, given the optimal way in which the Woodbury formula evaluates the likelihood, we expect that the algorithm we have given will play some part that solution, at least when using a low-level representation and prior. The desire to use mid-level and high-level priors brings further difficulties since then coordinate directions, in the sense of the resistor network or a FEM discretization, do not correspond to allowable changes in the trial image. Sampling algorithms for these higher-level representations remains a challenging area of research.

ACKNOWLEDGMENTS

This research was supported by Marsden grant 10-UOO-221.

REFERENCES

- [1] Parker, A. and Fox, C., "Sampling Gaussian distributions in Krylov spaces with conjugate gradients," *SIAM Journal on Scientific Computing (SISC)* **34**(3), B312–B334 (2012).
- [2] Fox, C., "Polynomial accelerated MCMC, and other sampling algorithms inspired by computational optimization," in [*Monte Carlo and Quasi-Monte Carlo Methods 2012*], Dick, J., Kuo, F. Y., Peters, G. W., and Sloan, I. H., eds., submitted, Springer-Verlag (2012).
- [3] Strang, G., [*Introduction to Applied Mathematics*], Wellesley-Cambridge Press (1986).
- [4] Glauber, R., "Time dependent statistics of the Ising model," *J. Math. Phys.* **4**, 294307 (1963).
- [5] Turcin, V., "On the computation of multidimensional integrals by the Monte Carlo method," *Probab. Appl.* **16**, 720724 (1971).
- [6] Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence* **6**, 721–741 (1984).
- [7] Gilks, W. R., Roberts, G. O., and George, E. I., "Adaptive direction sampling," *The Statistician* **43**(1) (1994).
- [8] Meyer, R., Cai, B., and Perron, F., "Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2," *Computational Statistics and Data Analysis* **52**(7), 3408–3423 (2008).
- [9] Fox, C., Nicholls, G., and Palm, M., "Efficient solution of boundary-value problems for image reconstruction via sampling," *Journal of Electronic Imaging* **9**(3), 251–259 (2000).