Multilevel approximations in sample-based inversion from the Dirichlet-to-Neumann map

J D Moulton¹, C Fox² and D Svyatskiy¹

¹Mathematical Modeling and Analysis, MS B284, Los Alamos National Laboratory,
Los Alamos, NM 87544
²Department of Physics, University of Otago, Dunedin, New Zealand

E-mail: moulton@lanl.gov, fox@physics.otago.ac.nz, dasvyat@lanl.gov

Abstract.
In 2005, Christen and Fox introduced a delayed acceptance Metropolis-Hastings (DAMH) algorithm that improved computational efficiency in sample-based imaging of electrical conductivity (EIT). That work used a linear approximation to the forward map in the first step of the algorithm. In this paper, we develop an alternative approximation for use in DAMH, namely a multilevel approximation developed from the hierarchy of coarse-scale models obtained by variational coarsening. This approach builds on two important strengths of robust multigrid solvers. First, the cost of a fine-scale solution of the forward map scales linearly with the degrees of freedom, and hence, it is provides better efficiency for algorithms performing sample-based inference. Second, the homogenization implicit in robust variational multigrid methods gives better solutions at coarse scales than homogenization by averaging of coefficients. We report results from a stylized example in electrical impedance imaging where data is a noisy and incomplete measurement of the Dirichlet-to-Neumann map.

1. Introduction
Calculating inferential solutions to inverse problems is currently a specialized and compute-intensive job, particularly in those cases where simulation of the forward map requires solution of a partial differential equation. A canonical example of such an inverse problem is the recovery of the conductivity within an object from measurements of potential and flux at the boundary. This problem arises in a number of settings, including the mapping of underground water transmissivity, mapping of acoustic or optical scattering from steady-state measurements, or the recovery of electrical conductivity or permittivity from stationary electrical measurements.

In this paper we consider the inverse problem of recovering the spatially-dependent electrical conductivity within a region from boundary measurements of current and voltage, often called electrical impedance tomography (EIT).

Then the forward map is simulated by solving an elliptic partial differential equation subject to boundary conditions determined by the measurement procedure. Equivalently, measurements correspond to incomplete and noisy measurements of the Dirichlet to Neumann map associated with the elliptic operator. In the deterministic setting the inverse problem may be viewed as inverting the implicit dependence of the Dirichlet to Neumann map on the internal conductivity.

In the statistical framework, solving the inverse problem corresponds to quantifying statistics of the posterior distribution of conductivities conditioned on measurements. We calculate
statistics over the posterior distribution using a Markov chain Monte Carlo (MCMC) procedure. In a standard Metropolis Hastings sampling algorithm computational cost is dominated by repeated numerical simulation of the forward map, and hence repeated solution of the large linear system arising from discretization of the boundary value problem. In this paper we explore the computational speedup available from using multi-level approximations to the forward map in the delayed acceptance Metropolis Hastings algorithm.

2. Electrical Impedance Tomography

Figure 1 shows a typical measurement setup for EIT (at the Physics Department, University of Kuopio). Instrumentation is used to assert a current flow between pairs of electrodes and the resulting potential generated at electrodes is simultaneously measured.

A somewhat idealized model for these measurements is the Neumann boundary-value problem (BVP)

\[- \nabla \cdot \sigma(x) \nabla \phi(x) = 0 \quad x \in \Omega \]
\[\sigma(x) \frac{\partial \phi(x)}{\partial n(x)} = j(x) \quad x \in \partial \Omega\]  

(1)

along with a reference for \( \phi \) such as \( \int_{\partial \Omega} c d\ell = 0 \) for some function \( c \). Here \( \sigma(x) \) denotes the unknown internal conductivity throughout the region \( \Omega \), \( j(x) \) is the current density asserted on the boundary \( \partial \Omega \), and \( \phi(x) \) is the resulting potential throughout \( \Omega \). Measurements are made of the potential on the boundary \( \phi|_{\partial \Omega} \) at the location of electrodes.

We consider the idealized case where point electrodes are placed around the boundary of the object, \( \partial \Omega \), at positions \( x_1, x_2, \ldots, x_E \). The current density asserted at the boundary is defined by the vector of currents \( j = (j(x_1), j(x_2), \ldots, j(x_E))^T \), and is explicitly

\[j(x) = \sum_{n=1}^{E} j_n \delta(x - x_n)\]  

Conservation of current requires \( \sum_{n=1}^{E} j_n = 0 \). The measured vector of voltages is denoted \( v = (\phi(x_1), \phi(x_2), \ldots, \phi(x_E))^T \). If \( N \) different current vectors \( j^{(1)}, j^{(2)}, \ldots, j^{(N)} \) are asserted, and the resulting voltages measured are \( v^{(1)}, v^{(2)}, \ldots, v^{(N)} \), then the measured data consists of the set of current-voltage pairs

\[d = \{ j^{(n)} , v^{(n)} \}_{n=1}^{N} \]
The forward map is denoted $A : \sigma \mapsto d$. Simulation of the forward map requires multiple solutions of the Neumann BVP in Equation 1, one for each boundary current asserted. The inverse problem is to find $\sigma(x)$ from $d$.

2.1. Dirichlet to Neumann map

The Dirichlet-to-Neumann (DtN) map $\Gamma_\sigma : \phi|_{\partial \Omega} \mapsto \sigma \frac{\partial \phi}{\partial n}|_{\partial \Omega}$ is the mapping from Dirichlet data, or boundary potential, to Neumann data or current density crossing the boundary. The conductivity $\sigma$ in the region implicitly defines the DtN map since the BVP with Dirichlet data uniquely determines the potential $\phi$ throughout $\Omega$ and hence $\sigma \frac{\partial \phi}{\partial n}$ at the boundary. Linearity of the DtN map follows since the solution of the BVP is a linear function of boundary data. However, the DtN map is not a linear function of $\sigma$. It is often useful to think of the measured data as consisting of points on the graph of the DtN map (restricted to electrodes). Then the inverse problem may be thought of as a regression problem for the true DtN map, parameterized in terms of the conductivity $\sigma$.

3. Bayesian Formulation and Solution

The presence of measurement errors, or noise, means that a practical measurement process is probabilistic, and the inverse problem is naturally a problem in statistical inference. We consider the simple case of additive noise $n$ with probability density function $\pi_N(n)$. The measurement process can then be written

$$d = A(\sigma) + n$$

and we see that the data is now a random variable that is dependent on $\sigma$. The (conditional) probability density for measuring $d$ given that $\sigma$ is the true conductivity is

$$\pi_L(d|x) = \pi_N(d - A(x)).$$

In a Bayesian formulation, inference about $\sigma$ is based on the posterior density

$$\pi(\sigma|d) = \frac{\pi_L(d|\sigma)\pi_P(\sigma)}{\pi_D(d)}$$

where $\pi_P(\sigma)$ denotes the prior probability density incorporating knowledge about the unknown conductivity independent of the data $d$. For our purposes it is sufficient to take the denominator $\pi_D(d)$ to be a finite constant once data $d$ is measured, thus ensuring that the posterior density is normalizable.

3.1. Sample-based solutions

Statistics that summarize the posterior density may be calculated as expectations over the posterior distribution. For any function $f$ of $\sigma$ the posterior expected value of $f$ may be calculated using Monte Carlo integration as follows. If \{\sigma_t, t = 1, 2, \ldots, N\} are distributed according to the posterior distribution, $\pi(\cdot|d)$, then

$$\mathbb{E}[f(\cdot)] \approx \frac{1}{N} \sum_{t=1}^{N} f(\sigma_t).$$

Hence evaluating expectations reduces to the task of drawing samples from the given posterior distribution $\pi(\cdot|d)$. The Markov chain Monte Carlo (MCMC) algorithm achieves sampling by generating \{\sigma_t\}_{t=0}^\infty as a Markov chain of random variables with a $t$-step distribution that tends to $\pi(\sigma|d)$, as $t \to \infty$. Thus the algorithm produces a random walk through the space of feasible conductivities with the long-term probability that the walk will visit a particular conductivity $\sigma$ tending to the desired posterior distribution. This ergodic property allows a sequence of states from the chain to used for evaluating the sum in Equation 5.
3.2. The Metropolis-Hastings Algorithm

In the following we write the abbreviated $\pi(\sigma)$ for $\pi(\sigma|d)$. We construct a Markov chain with the desired equilibrium properties by simulating a suitable transition kernel using the Metropolis-Hastings (MH) algorithm, given in algorithm 1. Note that this algorithm depends only on the ratio of densities, and hence the normalization constant is not required (but it is important that it exists). The stochastic update at step $t$ with the chain in state $\sigma_t$ is $\sigma_{t+1} = \text{MH}(\sigma_t)$.

\[ \sigma_{\text{new}} = \text{MH}(\sigma) \]

- Draw candidate state $\sigma'$ from proposal distribution $q(\cdot|\sigma)$
- Let $\alpha(\sigma, \sigma') = \min\left(1, \frac{q(\sigma|\sigma')\pi(\sigma')}{q(\sigma'|\sigma)\pi(\sigma)}\right)$
- With probability $\alpha(\sigma, \sigma')$ accept $\sigma'$ setting $\sigma_{\text{new}} = \sigma'$, otherwise reject setting $\sigma_{\text{new}} = \sigma$.

Algorithm 1: Metropolis-Hastings algorithm simulating operation by a suitable transition kernel. The proposal distribution $q(\sigma'|\sigma)$ may be any distribution that guarantees the chain is aperiodic and irreducible.

Typically many thousands or millions of iterations of the MH algorithm are required to achieve sufficiently accurate estimates via the approximation in Equation 5. Since the forward map must be simulated in each iteration, this sets the primary computational cost of implementing this algorithm in the context of inverse problems.

3.3. Delayed Acceptance MH

An algorithm that utilizes approximations to speed up the basic MH algorithm was given by Christen and Fox [1]. They considered the state-dependent approximation $\pi^*_{\sigma}(\cdot)$ to the posterior distribution calculated using a cheap approximation to the forward map, to give a modified Metropolis-Hastings MCMC. Once a proposal is generated from the proposal distribution $q(\sigma'|\sigma)$, to avoid calculating $\pi(\sigma')$ for proposals that are rejected, they first evaluate the proposal using the approximation $\pi^*_{\sigma}(\sigma')$ to create a second proposal distribution $q^*(\sigma'|\sigma)$ that is then used in a standard Metropolis-Hastings algorithm. The full definition is given in algorithm 2.

\[ \sigma_{\text{new}} = \text{DAMH}(\sigma) \]

- Draw $\sigma'$ from proposal distribution $q(\sigma'|\sigma)$
- Let $\alpha(\sigma, \sigma') = \min\left(1, \frac{q(\sigma|\sigma')\pi^*_{\sigma}(\sigma')}{q(\sigma'|\sigma)\pi^*_{\sigma}(\sigma)}\right)$
- With probability $\alpha(\sigma, \sigma')$ promote $\sigma'$ and go on, otherwise set $\sigma_{\text{new}} = \sigma$, i.e. reject $\sigma'$ and exit
- Define $q^*(\sigma'|\sigma) = \alpha(\sigma, \sigma')q(\sigma'|\sigma)$
- Let $\beta(\sigma, \sigma') = \min\left(1, \frac{q^*(\sigma'|\sigma')\pi(\sigma')}{q(\sigma'|\sigma)\pi(\sigma)}\right)$
- With probability $\beta(\sigma, \sigma')$ accept $\sigma'$ and set $\sigma_{\text{new}} = \sigma'$, otherwise reject and set $\sigma_{\text{new}} = \sigma$.

Algorithm 2: Delayed acceptance Metropolis-Hastings algorithm that makes use of an approximation.
When the approximation used is good, the second acceptance probability \( \beta(\sigma, \sigma') \) will be close to 1. In that case the exact calculation is only performed for proposals that are accepted. If the approximate calculation is also substantially cheaper than the exact calculation, the speedup achieved is the inverse of the acceptance rate. However, reduction in statistical (and computational) efficiency necessarily occurs when the approximation incorrectly classifies proposals.

Christen and Fox [1] presented an example of imaging the resistors in a network, using a local linearization of the forward map to demonstrate an order of magnitude speedup over the standard Metropolis Hastings algorithm. In that case local moves were employed in the proposal step to ensure accuracy in the linear approximation, thereby avoiding a substantial loss in statistical efficiency.

However, there are many cases where the need to propose only local, or small, changes to the state is overly restrictive. Examples are where proposals are based on gradients, or where state space is discrete in such a way that no allowable change in state is small. In those cases it is desirable to use an approximation to the forward map that has a greater domain of accuracy than do the local linear approximations. A suitable choice is approximations based on coarse discretization of the forward map, such as the variational coarsening of the BVP that we employ.

As noted by Christen and Fox [1], DAMH reduces to a one-step implementation of the surrogate transition method [2] when the approximation is independent of the state. They used a local linear approximation to the forward map, which is clearly state-dependent, and hence required the more general DAMH algorithm. However the situation is not so clear when using incomplete iterative solvers to form an approximation, as we do here. Then state (in)dependence often relies on whether or not the iteration is initialized using the current solution (and hence state). Since we also employ a multi-level coarsening that depends on the solution, the approximation we use will depend to some degree on state, and hence also requires DAMH.

4. Multilevel Approximations of the Forward Map

4.1. Robust Multigrid Solvers

The solution of the forward map relies on the solution of a discrete, linear, system of equations. Moreover, this system is large and sparse, suggesting that the necessary efficiency is best provided by iterative methods. In this case robust multigrid methods, which are optimal in the sense that the solution cost grows linearly in the number of unknowns, \( N \), are a natural choice. Specifically, these methods achieve their efficiency through the recursive use of successively coarser discrete problems (i.e., a hierarchy of coarse-scale discrete models), in conjunction with a complementary smoothing iteration.

For readers unfamiliar with multigrid methods, an an introduction to multigrid methods is given in [3], and more comprehensive treatment is given in [4]. Here, we highlight the basic elements of the most common recursive cycle, the V-cycle, which is shown schematically in Figure 2 and is described as follows:

- the residual on a particular grid is smoothed
  (i.e., it must be well approximated on a coarser grid)
- the residual is then restricted to the coarser grid
- repeat these two steps until the coarsest grid is reached
- solve on the coarsest grid
- interpolate to provide a correction on the next finer grid
- smooth the new residual
- repeat these two steps until the fine-grid is reached
Thus, it is apparent that the overall efficiency of the algorithm depends on the performance of the smoother, as well as the accuracy of the coarse-grid and interpolation operators. In EIT we need to ensure that the algorithm is robust with respect to highly heterogeneous and discontinuous coefficients. This need focuses our attention on the development of coarse-grid and interpolation operators because of the natural connection that the construction of these operators has with homogenization. Specifically, simply averaging coefficients of the fine-scale model to define a coarse-scale model of the same mathematical form does not, in general, capture the influence of the fine-scale operator at the coarse-scale. Instead, the construction of the hierarchy of operators in robust multigrid methods is achieved through a variational principle. Specifically, once an interpolation operator is defined based on the fine-scale discrete operator, the coarse-grid operator that minimizes the error in the range of the interpolation can be constructed. This construction of the Galerkin or variational coarse-grid operator does not make any assumption about the form of the coarse-scale model, and as a result it is significantly more accurate than other approaches. In fact, this variational operator is at the heart of robust multigrid solvers on structured grids, such as Black Box Multigrid (BoxMG) [5, 6], and for unstructured grids, such as Algebraic Multigrid (AMG) [7, 8], and Smoothed Aggregation based AMG [9, 10].

4.2. A Framework for Approximation
Multigrid algorithms are most often regarded as iterative methods that accelerate the solution of a discrete system of equations. Although this view is consistent with the popular class of multigrid methods known as the Correction Scheme (CS), it overlooks the powerful approximation framework that is provided by the lesser known Full Approximation Scheme (FAS) multigrid algorithm. FAS was originally developed for solving nonlinear problems, and is a natural extension of the CS algorithm to this case. In fact, for a linear problem the CS and FAS iterates are identical. However, in contrast to CS, FAS not only generates corrections, but also an approximation to the solution, on all coarse levels. In this sense, FAS provides an interesting opportunity to view the fine-scale as a means to improve the accuracy of the coarse-scale solution [11].

Recent work in this area includes the Multilevel Upscaling (MLUPS) algorithm developed in [12] for single-phase Darcy flow in highly heterogeneous porous media. The MLUPS algorithm leverages the robust and accurate hierarchy of coarse-scale models generated through the variational coarsening algorithm in BoxMG. Moreover, MLUPS uses these components in a very inexpensive FAS style algorithm with only a single V-cycle to achieve accuracy comparable to the much more costly multiscale finite element method (MSFEM) [13]. In this preliminary study the accuracy requirement of the synthetic EIT data is quite high, and the samples are drawn at the fine-scale. Thus, we consider the approximate forward map to be given by a small number of V(1,1) cycles of BoxMG with Red-Black Gauss-Siedel relaxation.
Figure 3. A phantom (true) spatially varying conductivity used in the EIT example. Conductivity values are 4 (black) and 3 (white) in arbitrary units.

5. An Example in Electrical Impedance Tomography (EIT)

Figure 3 shows a phantom two-level conductivity over a square domain $\Omega = [0, 1] \times [0, 1]$. The conductivity takes one of the two values 4 (black) and 3 (white) in arbitrary units of conductivity, and is constant over pixels.

We place four point electrodes evenly spaced on each edge of the square domain, and simulate current injection and voltage measurement at these electrodes. For the purpose of this work we work with the Green functions in which unit current is injected into each electrode, in turn, and removed evenly from all electrodes. Potential is then measured at all electrodes with respect to the reference potential which is the mean value over all electrodes. Measurement error is simulated by adding iid zero mean Gaussian noise giving a signal to noise ratio of about 1000:3, which is typical for actual electrical measurements.

We specify a prior distribution by modelling $\sigma$ on the pixel lattice as an Ising Markov random field, with distribution

$$
\pi_P(\sigma) \propto \exp \left( \theta \sum_{i=1}^{N^2} \sum_{j \sim i} \delta_{\sigma_i, \sigma_j} \right)
$$

where the sum over $j \sim i$ is a sum over pixel neighbors, $\theta$ is a smoothing parameter, and $\delta_{a,b}$ is the indicator function for the event $a = b$.

We specified the proposal distribution by choosing one of the moves

(i) flip a pixel
(ii) swap a pair of pixels neighboring a random pixel

at random with relative probability 1:3. While these moves are naive and most proposals correspond to no change in $\sigma$, these null proposals are straightforward to identify and reject so do not affect efficiency.

5.1. Pixel FEM Simulation of Forward Map

The basic, fine scale, discretization of the forward map based on a finite-element method (FEM) in which the pixels are taken as elements. The potential within each pixel is given by bilinear interpolation of nodal values, i.e. the potential at the four corners. The discretized system gives the discrete variational statement

$$
\phi = \arg \min \phi^T K \phi - j^T \phi \quad \text{subject to} \quad \phi \cdot c = 0
$$
Figure 4. (a) Marginal Posterior Mode (b) Conductivity Variance

where $c$ is a vector that is 1 for nodes that are electrodes and is otherwise zero, and $K$ is the
stiffness matrix. An exact penalty method is used to assert the essential boundary condition, so
the minimum is found by solving the normal equations

$$2 \left( K + c^T c \right) \phi = j.$$

5.2. Numerical Results

The standard MH algorithm using an exact solution of the FEM equations was run, and also the
DAMH algorithm using approximate solutions from an incomplete multigrid iteration. Iteration
parameters used in the exact solution were

- Red-Black Gauss-Seidel point relaxation
- $V(1,1)$ cycles
- Convergence criteria: $L_2$-norm of the residual less than $10^{-14}$ (typically 13 V-cycles)

while the approximate solves used

- Red-Black Gauss-Seidel point relaxation
- $V(1,1)$ cycles
- Convergence criteria: $L_2$-norm of the residual less than $10^{-5}$ (typically 3-4 V-cycles)

Since both standard MH and DAMH converge to the same target distribution, output
statistics calculated from the two chains are identical. Figure 4 shows the marginal posterior
mode (MPM) and the variance in conductivity. The MPM gives the posterior mode for each
pixel, and hence is the value that each pixel took most frequently in the output state of the
Markov chain. For two-level images, as here, the MPM may be calculated by taking the mean
conductivity and then setting each pixel to the closest allowable value. The variance is depicted
as a grey-scale image with white showing large variance while black is zero variance. It is evident
that the greatest uncertainty in the reconstruction, i.e. the regions with greatest variance, are
the boundaries between conductivity values that are towards the center of the image. This is
typical for reconstructed images in EIT.

We also collected some output statistics over the chain such as the sequence of values of the
log likelihood $\log \pi_L (\sigma_t)$ and the log prior $\log \pi_P (\sigma_t)$. Traces of these statistics are shown in
Figure 5. As can be seen, convergence of both MH and DAMH algorithms appears satisfactory,
though in both cases longer output runs would be desirable.

We also evaluated the integrated autocorrelation time (IAT) for the two sequences of log
likelihood to evaluate the statistical efficiency of each chain. The IAT is, roughly speaking, the
length of the output chain that has the same variance reducing power as a single independent
Figure 5. Trace of log likelihood and log prior of the MH algorithm (top row) and the DAMH algorithm (bottom row).

Figure 6. Integrated autocorrelation for log likelihood is shown for the single stage MH algorithm (left) and two-stage algorithm (right). The integrated autocorrelation time for the two-stage algorithm is 165, versus 83.7 for the single stage algorithm.

sample, and hence is a measure of the effective time between independence of output states of a chain. A smaller IAT is desirable since it indicates a more statistically efficient chain. Evaluation of the IAT for the two chains is shown in Figure 6. The standard MH algorithm achieved an IAT of 83.7 (based on the log likelihood) while the DAMH algorithm using the approximate calculation had an IAT of 165, i.e. almost twice as long. Hence the DAMH algorithm will need
to be run for twice as many iterations as the MH algorithm to produce sample-based statistics with the same accuracy. This reduction in statistical efficiency is consistent with the first step in the DAMH algorithm falsely rejecting 47% of proposals, i.e. rejecting 47% of proposals that would have been accepted by the exact calculation. However, faster compute time of the approximate solution meant that overall the DAMH algorithm produced effectively independent samples in about half the compute time of standard MH.

6. Conclusions
In this paper we gave an example in which coarse-scale solutions available in a multilevel solution to the forward map is used in the first step of a DAMH algorithm. For proposals that are ‘promoted’ in that approximate accept/reject step, the solution is then evaluated exactly ensuring that samples are correctly distributed. A numerical example in EIT with two-level conductivity was presented to demonstrate application of the algorithm.

The marginal posterior mode was presented as an example of a posterior statistic that could be used as an estimate of true parameters. The posterior variance was also shown, and confirmed the standard intuition that resolution in EIT is greatest near the boundary, and is least at the boundary of conductivity changes that are away from the boundary.

In the computed example statistical efficiency was roughly halved by use of DAMH. In the near future we expect to further investigate the parameters determining exact and approximate multigrid iterations to give greatest computationally efficiency.

Acknowledgments
The work of JDM and DS was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396, and was partially supported by the DOE Office of Science Advanced Scientific Computing Research (ASCR) Program in Applied Mathematics Research and the Laboratory Directed Research and Development Program. CF was supported by an ISAT Linkages Fund grant from the Royal Society of New Zealand.

References
[1] Christen J A and Fox C 2005 Journal of Computational and Graphical Statistics 14 795 – 810
[2] Liu J S 2001 Monte Carlo Strategies in Scientific Computing (Springer) ISBN 0387952306
[3] Briggs W L, Henson V E and McCormick S F 2000 A Multigrid Tutorial (Philadelphia: SIAM Books) second edition
[4] Trottenberg U, Oosterlee C and Schuller A 2001 Multigrid (Academic Press)
[5] Dendy J E 1982 J. Comput. Phys. 48 366–386
[6] Dendy J E 1987 SIAM J. Sci. Stat. Comput. 8 673–685
[7] Stüben K 1983 Appl. Math. Comput. 13 419–451
[8] Stüben K 2001 J. Comp. Appl. Math. 128 281 – 309
[9] Vanek P, Mandel J and Brezina M 1996 Computing 56 179 – 96
[10] Brezina M, Falgout R, MacLachlan S, Mantuflle T, McCormick S and Ruge J 2005 SIAM Rev. 47 317 – 346
[11] Brandt A 1984 Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamics The Weizmann Institute of Applied Science Rehovot, Israel
[12] MacLachlan S P and Moulton J D 2006 Water Resour. Res. 42 W02418
[13] Hou T and Wu X 1997 J. Comput. Phys. 134 169–189
[14] Hastings W K 1970 Biometrika 57 97 – 109