

Multigrid Inversion Algorithms with Applications to Optical Diffusion Tomography *

Seungseok Oh, Adam B. Milstein, Charles A. Bouman, and Kevin J Webb
School of Electrical and Computer Engineering, Purdue University
West Lafayette, Indiana 47907-1285

Abstract

In this paper, we propose a general framework for nonlinear multigrid inversion applicable to any inverse problem in which the forward model can be naturally represented at differing resolutions. In multigrid inversion, the problem is adjusted to be solved at each resolution by using the solutions at both finer and coarser resolutions. To do this, we formulate a consistent set of cost functionals across resolutions. At each resolution, both the forward and inverse problems are discretized at the lower resolution; thus reducing computation. Our simulation results for the problem of optical diffusion tomography indicate that multigrid inversion can dramatically reduce computation in this application.

1 Introduction

Recently, some new imaging modalities, such as optical diffusion tomography (ODT) and electrical impedance tomography, have received great attention. For example, ODT holds great potential as a safe non-invasive medical diagnostic modality with chemical specificity. However, solving the required inverse problems for these new modalities is very computationally challenging because the associated forward models are described by the solution to a three-dimensional (3-D) partial differential equation (PDE).

To solve inverse problems, we need to choose a discretization grid spacing to represent both the forward model and its subsequent inverse operation. Although a fine grid is desirable because it reduces modeling error and increases the resolution of the final image, these improvements are obtained at the expense of a dramatic increase in computational cost. Solving problems at fine resolution also tends to slow convergence because the convergence speed of most fixed grid algorithms depends on the spectral characteristics of the error [1, 2]. Furthermore, fixed grid optimization methods essentially perform a local search of the cost

function, and therefore tend to become trapped in local minima.

In this paper we propose a method we call multigrid inversion. Multigrid inversion is a general approach to applying nonlinear multigrid optimization to the solution of inverse problems. A key innovation in our approach is that the resolution of both the forward and inverse models are varied at different grid resolutions. This makes our method particularly well suited to the solution of inverse problems with PDE forward models for a number of reasons: (1) The computation is dramatically reduced by using coarse grids to solve the forward model PDE; (2) the coarse grid forward model is computed using a discretized version of the true PDE, thereby preserving its nonlinear characteristics; (3) a wide variety of optimization methods can be used for solving the inverse problem at each grid.

The multigrid inversion method is formulated in an optimization framework by defining a sequence of optimization functionals at varying resolutions. In order for the method to have well behaved convergence to the correct fine grid solution, it is essential that the cost functional at different scales be consistent. To achieve this, we formulate a recursive method for adapting the coarse grid functionals which guarantees that the fine grid solution is also a fixed point of the multigrid algorithm.

This paper will be organized as follows. Section 2 introduces general concept of the multigrid inversion algorithm. In Section 3, we illustrate an application of the proposed multigrid inversion to the ODT problem, and its numerical results will be provided in Section 4. Section 5 makes concluding remarks.

2 Multigrid Inversion Framework

2.1 Inverse Problems

Let Y be a random vector of (real or complex) measurements, and let x be a finite dimensional vector representing the unknown quantity, in our case an image, to be reconstructed. For any inverse problem, there is

* Appeared in 36th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 3-6, 2002.

then a forward model $f(x)$ given by

$$E[Y|x] = f(x), \quad (1)$$

which represents the computed means of the measurements given the image x .

We will assume that the measurements Y are conditionally Gaussian given x , so that

$$\log p(y|x) = -\frac{1}{2\alpha} \|y - f(x)\|_{\Lambda}^2 - \frac{P}{2} \log(2\pi\alpha|\Lambda|^{-1}), \quad (2)$$

where Λ is a positive definite weight matrix, P is the dimensionality of the measurement (i.e. the length of Y in real-valued data, or twice its length in complex-valued data), α is a parameter proportional to the noise variance, and $\|w\|_{\Lambda}^2 = w^H \Lambda w$.

Our objective is to invert the forward model of (1) and thereby estimate x from a particular measurement vector y . All these methods work by computing the value of x which minimizes a cost functional of the form

$$\frac{1}{2\alpha} \|y - f(x)\|_{\Lambda}^2 + \frac{P}{2} \log(2\pi\alpha|\Lambda|^{-1}) + S(x), \quad (3)$$

where $S(x)$ is a stabilizing functional used to regularize the inverse. Since the noise variance parameter α is usually unknown in practice, we will estimate both α and x by simultaneously maximizing over both quantities. Minimization of (3) with respect to α and dropping constants yields the final cost functional to be optimized

$$c(x) = \frac{P}{2} \log \|y - f(x)\|_{\Lambda}^2 + S(x). \quad (4)$$

2.2 Multigrid Inversion Algorithm

Once the cost function of (4) is formulated, the inverse is computed by minimizing the cost function with respect to x . In this section, we derive the basic multigrid inversion algorithm for solving the minimization of (4).

Let $x^{(0)}$ denote the finest grid image, and let $x^{(q)}$ be a coarse resolution representation of $x^{(0)}$ with a grid sampling period of 2^q times the finest grid sampling period. To obtain a coarse resolution image $x^{(q+1)}$ from a fine resolution one $x^{(q)}$, we use the relation $x^{(q+1)} = I_{(q)}^{(q+1)} x^{(q)}$, where $I_{(q)}^{(q+1)}$ is a linear decimation matrix. We use $I_{(q+1)}^{(q)}$ to denote the corresponding linear interpolation matrix.

We first define a cost functional, $\tilde{c}^{(q)}(x^{(q)})$, with a form analogous to that of (4), but with quantities indexed by the scale q

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_{\Lambda}^2 + S^{(q)}(x^{(q)}) \quad (5)$$

Notice that the forward model $f^{(q)}(\cdot)$ and the stabilizing functional $S^{(q)}(\cdot)$ are both evaluated at scale q . This is important because evaluation of the forward model at low resolution substantially reduces computation. The specific form of $f^{(q)}(\cdot)$ generally results from the physical problem being solved with an appropriate grid spacing. In Section 3, we will give a typical example for ODT where $f^{(q)}(\cdot)$ is computed by discretizing a 3-D PDE using a grid spacing proportional to 2^q . The reduced number of variables in the forward model dramatically decreases computation. The quantity $y^{(q)}$ in (5) denotes an adjusted measurement vector at scale q . The stabilizing functional at each scale is fixed and chosen to best approximate the fine scale functional. We give an example of such a stabilizing functional later in Section 2.3.

We will now explain how the cost functions at each scale can be matched to produce a consistent solution. To do this, we define an adjusted cost functional

$$\begin{aligned} c^{(q)}(x^{(q)}) &= \tilde{c}^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} \\ &= \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_{\Lambda}^2 \\ &\quad + S^{(q)}(x^{(q)}) - r^{(q)} x^{(q)}, \end{aligned}$$

where $r^{(q)}$ is a row vector used to adjust the functionals gradient. At the finest scale, all quantities take on their fine scale values and $r^{(q)} = 0$, so that $\tilde{c}^{(0)}(x^{(0)}) = c^{(0)}(x^{(0)}) = c(x)$. Our objective is then to derive recursions for the quantities $y^{(q)}$ and $r^{(q)}$ that match the cost functions at fine and coarse scales.

Let $x^{(q)}$ be the current solution at grid q . We would like to improve this solution by performing an iteration of fixed grid optimization at the coarser grid $q+1$ and using this solution to correct the finer grid solution. This coarse grid update is expressed as

$$\tilde{x}^{(q+1)} \leftarrow \text{Fixed_Grid_Update}(I_{(q)}^{(q+1)} x^{(q)}, c^{(q+1)}(\cdot)), \quad (6)$$

where the operator $\text{Fixed_Grid_Update}(x_{init}, c(\cdot))$ is any fixed grid update algorithm (e.g. conjugate gradient, steepest descent, or iterative coordinate descent) which is designed to reduce $c(\cdot)$ with the initial value x_{init} , $I_{(q)}^{(q+1)} x^{(q)}$ is the initial condition formed by decimating $x^{(q)}$, and $\tilde{x}^{(q+1)}$ is the updated value. We may now use this result to update the finer grid solution. We do this by interpolating the *change* in the coarser scale solution

$$\tilde{x}^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)} (\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)}). \quad (7)$$

Ideally, the new solutions $\tilde{x}^{(q)}$ should be at least as good as the old solution $x^{(q)}$. Specifically, we would

like $c^{(q)}(\tilde{x}^{(q)}) \leq c^{(q)}(x^{(q)})$ when the fixed grid algorithm monotonically reduces the cost function. However, this may not be the case if the cost functionals are not consistent. In fact, for a naively chosen set of cost functionals, the coarse scale correction could easily move the solution away from the optimum.

This problem of inconsistent cost functions is eliminated if the fine and coarse scale cost functions are equal within an additive constant. This means we would like the following approximate equality to hold for all values of $\tilde{x}^{(q+1)}$.

$$\begin{aligned} c^{(q+1)}(\tilde{x}^{(q+1)}) &\simeq \\ c^{(q)}(x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)})) &+ \text{constant} \end{aligned} \quad (8)$$

Our objective is then to choose a coarse scale cost functional which matches the fine cost functional as described in (9). We do this by the proper selection of the $y^{(q+1)}$ and $r^{(q+1)}$. First, we enforce the condition that the initial error between the forward model and measurements be the same at coarse scale and fine scales

$$y^{(q+1)} - f^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) = y^{(q)} - f^{(q)}(x^{(q)}) . \quad (9)$$

This yields the update for $y^{(q+1)}$

$$y^{(q+1)} \leftarrow y^{(q)} - \left[f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) \right] . \quad (10)$$

Intuitively, the term in the bracket compensates for the forward model mismatch between resolutions.

Next, we use the condition introduced in [3] to enforce the condition that the gradients of the coarse and fine cost functionals be equal at the current values of $x^{(q)}$ and $x^{(q+1)} = I_{(q)}^{(q+1)}x^{(q)}$. More precisely, we enforce the condition that

$$\nabla c^{(q+1)}(x^{(q+1)}) \Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} = \nabla c^{(q)}(x^{(q)}) I_{(q+1)}^{(q)} .$$

This condition is essential to assuring that the optimum solution is a fixed point of the multigrid inversion algorithm [3]. The equality of (11) can be enforced at the current value $x^{(q)}$ by choosing

$$\begin{aligned} r^{(q+1)} \leftarrow \nabla \tilde{c}^{(q+1)}(x^{(q+1)}) \Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} \\ - \left(\nabla c^{(q)}(x^{(q)}) \right) I_{(q+1)}^{(q)} \end{aligned} \quad (11)$$

where $\tilde{c}^{(q)}(\cdot)$ is the unadjusted cost function defined in (5).

The multigrid V algorithm [2] is obtained by applying this two-grid algorithm recursively in resolution,

as shown in the pseudocode in Fig. 1. After initialization of $r^{(0)} \leftarrow 0$ and $y^{(0)} \leftarrow y$, we can then minimize (4) through iterative application of the multigridV(\cdot) subroutine at resolution 0. In this figure, we use the notation $c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)})$ to make the dependency on $y^{(q+1)}$ and $r^{(q+1)}$ explicit.

2.3 Stabilizing Functionals

The stabilizing functionals at coarser scales q must be selected so that

$$S^{(q)}(x^{(q)}) \simeq S(x) . \quad (12)$$

To illustrate how to achieve this relation, consider an example of the generalized Gaussian Markov random field (GGMRF) image prior model. For the GGMRF prior in MAP estimation, the stabilizing functional is usually of the form

$$S(x) = \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p , \quad (13)$$

where σ is a normalization parameter, and $1 \leq p \leq 2$ controls the degree of edge smoothness [4, 5, 3]. The corresponding coarse scale stabilizing functionals are derived to be

$$S^{(q)}(x^{(q)}) = \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i^{(q)} - x_j^{(q)}|^p . \quad (14)$$

For a d dimensional problem, (12) can be satisfied by selecting the normalization parameter $\sigma^{(q)}$ as

$$\sigma^{(q)} = 2^{(q)(1-\frac{d}{p})} \cdot \sigma^{(0)} , \quad (15)$$

where we assume $x_i - x_j \simeq (x_i^{(q)} - x_j^{(q)})/2^q$.

3 Application to Optical Diffusion Tomography

Optical diffusion tomography (ODT) is a method for measuring cross-sections of optical properties from the measured data of light transmitted through highly scattering medium. In frequency domain ODT, the measured modulation envelope of the optical flux density is used to reconstruct the absorption coefficient μ_a and diffusion coefficient D at each discretized grid point. However, for simplicity, we will only consider reconstruction of the absorption coefficient. The 3-D domain is discretized, and the set of unknown μ_a at each discrete grid point forms an image x . The complex amplitude $\phi_k(r)$ of the modulation envelope due to a point source at position a_k and angular frequency ω satisfies the frequency domain diffusion equation

$$\nabla \cdot [D(r)\nabla\phi_k(r)] + [-\mu_a(r) - j\omega/c]\phi_k(r) = -\delta(r - a_k)$$

```

 $x^{(q)} \leftarrow \text{MultigridV}(q, x^{(q)}, y^{(q)}, r^{(q)}) \{$ 
  Repeat  $\nu_1^{(q)}$  times
     $x^{(q)} \leftarrow \text{Fixed\_Grid\_Update}(x^{(q)}, c^{(q)}(\cdot; y^{(q)}, r^{(q)}))$  //Fine grid update
  If  $q = Q - 1$ , return  $x^{(q)}$  //If coarsest scale, return result
   $x^{(q+1)} \leftarrow I_{(q)}^{(q+1)} x^{(q)}$  //Decimation
  Compute  $y^{(q+1)}$  using (10)
  Compute  $r^{(q+1)}$  using (11)
   $x^{(q+1)} \leftarrow \text{MultigridV}(q + 1, x^{(q+1)}, y^{(q+1)}, r^{(q+1)})$  //Coarse grid update
   $x^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)} (x^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)})$  //Coarse grid correction
  Repeat  $\nu_2^{(q)}$  times
     $x^{(q)} \leftarrow \text{Fixed\_Grid\_Update}(x^{(q)}, c^{(q)}(\cdot; y^{(q)}, r^{(q)}))$  //Fine grid update
  Return  $x^{(q)}$  //Return result
 $\}$ 

```

Figure 1: Pseudo-code specification of subroutine for multigrid V-cycle inversion.

where r is position and c is the speed of light in the medium. Then the measurement of a detector at location b_m resulting from a source at location a_k is modeled by the complex value $\phi_k(b_m)$. Note that $\phi_k(b_m)$ depends on the unknown image x . The complete forward model function $f(x)$ is then given by the set of $\phi_k(b_m)$'s, and the measurement vector y is also organized in the corresponding order. Note that $f(x)$ is a highly nonlinear function since it is given by the solution to a PDE using coefficients x .

Our objective is to estimate the unknown image x from the measurements y . Using an independent Gaussian shot noise model [3] and the GGMRF image prior model, the MAP estimate of x in a Bayesian framework is reduced to the minimization [3] of

$$c(x) = \frac{P}{2} \log \|y - f(x)\|_{\Lambda}^2 + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p, \quad (16)$$

where Λ is diagonal with $\Lambda_{(i,i)} = 1/|y_i|$. This cost functional has the same form as (4) with the stabilizing functional given by (13). We use multigrid inversion to solve the required optimization problem using the set of coarse grid cost functionals developed in Section 2.2 and 2.3. For fixed grid updates at each scale q , we use the ICD algorithm which also incorporates sequential updates of α and x . See [5, 3] for details of the fixed grid ICD algorithm.

4 Numerical Results

The performance of the multigrid inversion algorithm was investigated for the ODT problem. A cubic tissue phantom of dimension $10 \times 10 \times 10$ cm on an edge shown in Fig. 2(a) was used for reconstructions. The μ_a background was linearly varied from 0.01 cm^{-1} to 0.04 cm^{-1} in the x-direction, except in

the outermost boundary of width 1.25 cm which was set to a constant value of $\mu_a = 0.025 \text{ cm}^{-1}$. Two spherical μ_a inhomogeneities were added with diameters of 1.85 cm and the values of $\mu_a = 0.1 \text{ cm}^{-1}$ (left-top) and $\mu_a = 0.12 \text{ cm}^{-1}$ (right-bottom). The spheres were both centered along the z axis of the phantom. The diffusion coefficient D was constant with $D = 0.03$ cm. Eight sources with a modulation frequency of 100 MHz and nine detectors were located on each face. All source detector pairs were used, except those on the same face of the cube. To simulate the measurements, we solved the diffusion equations with $129 \times 129 \times 129$ grid points. Gaussian shot noise was added to the data, and the average signal-to-noise ratio for sources and detectors on opposite faces was 35 dB. Since the finer discretization more accurately approximates the continuous domain PDEs and thus the real measurements, we used this fine grid forward solution as the simulated measurement data for all reconstructions.

Reconstructions with $65 \times 65 \times 65$ grid points were constructed using Fixed-grid algorithms and the multigrid-V optimization algorithms with 2, 3, and 4 levels of resolution. The eight outermost layers of grid points are fixed to their true values in order to avoid singularities near the sources and detectors. These regions were excluded from all cross-section figures and the evaluation of RMS reconstruction error. In order to make fair comparisons of computational speed, all iteration measurements were scaled to units of a single fixed grid iteration at the finest scale. For each algorithm the reconstruction was initialized to the average value of the true phantom, which is $\mu_a = 0.026 \text{ cm}^{-1}$ and $D = 0.03$ cm. The image prior model used $p = 1.2$ and $\sigma = 0.018 \text{ cm}^{-1}$.

The multigrid algorithms speed the convergence both in the sense of cost function and root-mean-

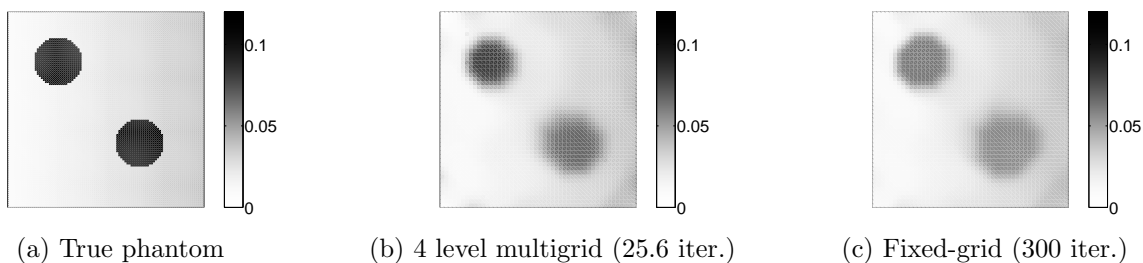


Figure 2: Cross-sections of the $z = 0$ plane

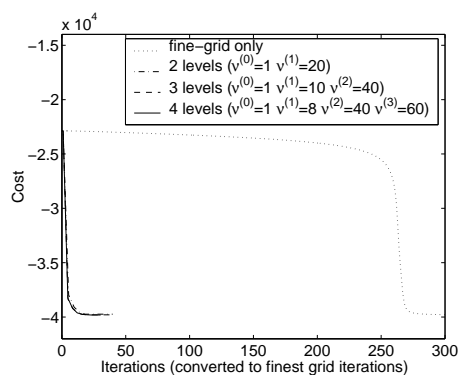


Figure 3: Cost function versus iteration numbers

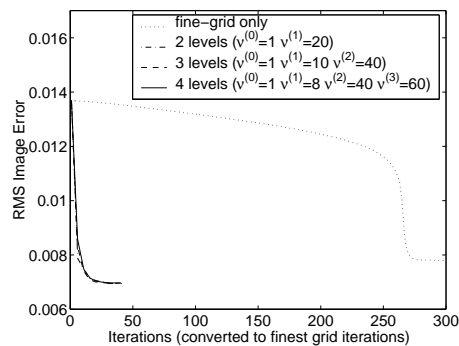


Figure 4: RMS image error versus iteration numbers

square (RMS) reconstruction error, as shown in Fig. 3 and Fig. 4. The multigrid algorithms converge in only 25 iterations while the fixed algorithm required in 280 iterations. Convergence speed and reconstruction image quality were similar for all multigrid algorithms but were significantly worse for the fixed grid algorithm. In fact, the multigrid algorithms converged to a slightly lower value of the cost functional (i.e. -39810 versus -39795) which appears to have a significant impact on reconstruction quality. The RMS image error for the 4 level multigrid algorithm converged to 0.0069 while the fixed algorithm converged to a slightly higher value of 0.0078. The reconstructions in Fig. 2(b) also shows that the multigrid algorithms reconstructed images are visually more accurate than the fixed grid algorithm in Fig. 2(c).

5 Conclusions

We proposed a multigrid inversion algorithm which is particularly well suited for nonlinear inverse problems with forward models given by the solution to a PDE. Multigrid inversion works by dynamically computing consistent cost functionals across different scales. This guarantees that the solution to the multigrid algorithm is also a fixed point to the fine grid algorithm. Experimental results for the ODT applica-

tion demonstrate that multigrid inversion can dramatically speed convergence while producing robust and stable convergence.

Acknowledgements

This work was supported by the National Science Foundation under contract CCR-0073357.

References

- [1] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Process.*, vol. 41, no. 2, pp. 534–548, February 1993.
- [2] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A multigrid tutorial, 2nd Ed.*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.
- [3] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, "Nonlinear multigrid algorithms for Bayesian optical diffusion tomography," *IEEE Trans. on Image Process.*, vol. 10, no. 6, pp. 909–922, June 2001.
- [4] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. on Image Process.*, vol. 2, no. 3, pp. 296–310, July 1993.
- [5] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography using iterative coordinate descent optimization in a Bayesian framework,"

J. Opt. Soc. Am. A, vol. 16, no. 10, pp. 2400–2412,
October 1999.