

Multigrid Algorithms for Optimization and Inverse Problems

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

A variety of new imaging modalities, such as optical diffusion tomography, require the inversion of a forward problem that is modeled by the solution to a three-dimensional partial differential equation. For these applications, image reconstruction can be formulated as the solution to a non-quadratic optimization problem.

In this paper, we discuss the use of nonlinear multigrid methods as both tools for optimization and algorithms for the solution of difficult inverse problems. In particular, we review some existing methods for directly formulating optimization algorithm in a multigrid framework, and we introduce a new method for the solution of general inverse problems which we call multigrid inversion. These methods work by dynamically adjusting the cost functionals at different scales so that they are consistent with, and ultimately reduce, the finest scale cost functional. In this way, the multigrid optimization methods can efficiently compute the solution to a desired fine scale optimization problem. Importantly, the multigrid inversion algorithm can greatly reduce computation because both the forward and inverse problems are more coarsely discretized at lower resolutions. An application of our method to optical diffusion tomography shows the potential for very large computational savings.

Keywords: multigrid algorithms, inverse problems, optimization, optical diffusion tomography, multiresolution

1. INTRODUCTION

A large class of image processing problems, such as deblurring, high-resolution rendering, image recovery, image segmentation, motion analysis, and tomography, require the solution of inverse problems. Often, the numerical solution of these inverse problems can be computationally demanding, particularly when the problem must be formulated in three dimensions.

Recently, some new imaging modalities, such as optical diffusion tomography (ODT)¹ and electrical impedance tomography (EIT),² have received great attention. For example, ODT holds great potential as a safe non-invasive medical diagnostic modality with chemical specificity.³ However, the inverse problems associated with these new modalities present a number of difficult challenges. First, the forward models are described by the solution of a partial differential equation (PDE) which is computationally demanding to solve. Second, the unknown image is formed by the coefficients of the PDE, so the forward model is highly nonlinear, even when the PDE is itself linear. Thus, the image reconstruction requires the solution to a non-quadratic optimization problem. Finally, these problems typically are inherently 3-D due to the 3-D propagation of energy in the scattering media being modeled. Since many phenomena in nature are mathematically described by PDEs, many other inverse problems have similar computational difficulties, including microwave tomography, thermal wave tomography, and inverse scattering.

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

Further author information: Send correspondence to Charles A. Bouman.

Charles A. Bouman : E-mail: bouman@ecn.purdue.edu, Telephone: 1 765-494-0340, Fax: 1 765-494-3358

In many inverse problems, it is necessary 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. It is well-known that the convergence speed of most fixed grid algorithms depends on spectral characteristics of error. For example, many fixed grid algorithms such as iterative coordinate descent (ICD) * effectively eliminate error at high spatial frequencies, but low frequency errors are damped slowly.^{4,5} Furthermore, fixed grid optimization methods essentially perform a local search of the cost function, and therefore tend to become trapped in local minima. Thus, optimization at a fixed fine resolution may produce poor results when the cost functional being minimized is not convex.

Multigrid algorithms have been investigated to reduce computations for inverse problems. While multigrid algorithms are closely related to general multiresolution techniques such as wavelet transforms and coarse-to-fine approaches, they provides several unique advantages over the other multiresolution approaches. First, since some form of iteration is usually imperative in nonlinear problems, multigrid algorithms are well adapted for solving nonlinear problems because of their iterative nature.⁵ Second, they provide with a systematic method to go from fine to coarse as well as from coarse to fine. Thus, unlike coarse-to-fine approaches, coarse scale updates can be applied whenever they are expected to be effective. Finally, since multigrid algorithms perform operations at each resolution on the space domain image using results obtained at both the finer and coarser resolutions, they can be naturally used to enforce nonnegativity constraints that are essential to obtain physically meaningful image in many imaging systems, unlike wavelet transforms.

In this paper, we discuss the use of nonlinear multigrid methods for the solution of inverse problems. We review prior approaches in which a multigrid solver was used to solve optimization problems, and we introduce our approach based on direct formulation of inverse problems in an optimization framework. In particular, we introduce a new method for the solution of general inverse problems which we call multigrid inversion. In multigrid inversion, the optimization problems resulting from inverse problems are solved by dynamically adjusting the cost functionals at different scales so that they are consistent with, and ultimately reduce, the finest scale cost functional. Since the coarse scale problems involves reduced number of variables, the multigrid optimization methods can efficiently compute the solution to a desired fine scale optimization problem. Importantly, the multigrid inversion algorithm can greatly reduce computation because both the forward and inverse problems are more coarsely discretized at lower resolutions.

This paper is organized as follows. Section 2 formulates inverse problems directly in an optimization framework. In section 3, we review some existing methods to apply multigrid solver methods to solve the optimization problem resulting from inverse problems. Section 4.1 introduces the general concept of the multigrid inversion algorithm, and Section 4.2 discusses relation between multigrid inversion and other approaches, such as functional substitution and multigrid solver methods. In Section 5, we illustrate the application of the multigrid inversion method to the ODT problem, and its numerical results are provided. Finally, Section 6 makes concluding remarks.

2. OPTIMIZATION FRAMEWORK FOR 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 then a forward

*ICD is generally referred to as Gauss-Seidel in the PDE literature, and sometimes referred to as iterative conditional mode (ICM) in the MAP estimation literature.

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

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

This cost functional involves a logarithm in the data term. However, the discussion in this paper are equally applicable to the cost functional without the logarithm, i.e. $c(x) = \frac{P}{2} \|y - f(x)\|_{\Lambda}^2 + S(x)$.

3. MULTIGRID SOLVER APPROACHES FOR OPTIMIZATION

Once the cost functional of (4) is formulated, the inverse is computed by minimizing the cost functional with respect to x . A possible approach to minimize (4) is to solve the equation arising from the first order optimality condition

$$\nabla c(x) = 0 , \quad (5)$$

where $\nabla c(\cdot)$ denotes the row vector formed by the gradient of the functional $c(\cdot)$. Since multigrid algorithms were originally developed to facilitate the computation of PDE solvers,⁵ multigrid equation solvers can be naturally used to numerically solve the optimality condition equation (5).

In linear inverse problems, if we use a quadratic stabilizing functional, the equation (5) is given by a linear equation. In this case we can simply apply linear multigrid methods, which have been developed for solving linear PDEs. However, nonlinear inverse problems result in nonlinear optimality condition equations. In addition, even in case of linear inverse problems, the equation (5) may be nonlinear if a non-quadratic stabilizing functional is used, for example, if we use a non-Gaussian image prior model in MAP estimation.^{6,7} Therefore, linear multigrid solver approaches cannot be applied directly in these cases.

One approach is to circumvent the nonlinearity by local approximation of original cost functional with a quadratic functional. Then, we can use linear multigrid methods to solve the optimality condition for this quadratic functional, which is a linear equation. For example, McCormick and Wade⁸ applied this approach to the nonlinear inverse problems of EIT by linearization of the nonlinear forward model. In related work, Dreyer

et al. also applied linear multigrid solvers after the original Lagrangian was approximated into a quadratic programming problem.⁹

Using nonlinear multigrid solvers, such as the full approximation scheme (FAS), directly accounts for nonlinearity of inverse problems. Bouman and Sauer showed that nonlinear multigrid algorithms could be applied to inversion of Bayesian tomography problems.⁶ This work used nonlinear multigrid techniques to compute maximum *a posteriori* (MAP) reconstructions with non-Gaussian prior distributions and a non-negativity constraint. Borcea used a nonlinear multigrid approach for EIT problems.¹⁰

4. NONLINEAR MULTIGRID INVERSION

Our approach to the minimization of (4) is to apply multigrid methods directly in an optimization framework by formulating cost functionals at different scales. Since this approach works entirely in an optimization framework, it has the following advantages. First, problems that are not easily transformed to systems of equations, such as problems involving nonlinear inequality constraints, can be more straightforwardly dealt with.¹¹ Second, generally, it is known that the convergence theory for optimization algorithms is stronger than that for nonlinear equation algorithms.¹¹ Third, for a traditional finite-dimensional optimization problem, applying optimization algorithms is more flexible than trying to solve the optimality conditions using an algorithm for solving nonlinear equations.¹¹ Finally, the change of the cost serves as a measure of the quality of the solution.

Ye *et al.*^{7, 12} formulated the multigrid approach directly in an optimization framework, and used the method to solve ODT problems. Their multigrid optimization algorithm works by formulating coarse scale cost functionals using iterative decimation of the Fréchet derivative matrix in the data term. Importantly, this approach is based on the matching of cost functional derivatives at different scales. In related work, Nash and Lewis formulated multigrid algorithms for the solution of a broad class of optimization problems.^{11, 13}

In this section, we propose a multigrid inversion algorithm for solving the optimization of (4). 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. First, the computation can be dramatically reduced by using coarser grids to solve the forward model PDE. In previous approaches, the forward model PDE was solved only at the finest grid. This means that coarse grid updates were either computationally costly, or a linearization approximation was made for the coarse grid forward model.^{7, 8} Second, the coarse grid forward model can be modeled by the correct discretized PDE, preserving the nonlinear characteristics of the forward model. Finally, a wide variety of optimization methods can be used for solving the inverse problem at each grid. For example, common methods such as pre-conditioned conjugate gradient and/or adjoint differentiation can be employed at each grid resolution.

4.1. Nonlinear Multigrid Inversion

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 coarser resolution image $x^{(q+1)}$ from a finer resolution image $x^{(q)}$, 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 coarse grid cost functional, $\tilde{c}^{(q)}(x^{(q)})$, with a form analogous to that of (4), but with quantities indexed by the scale q

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_{\Lambda}^2 + S^{(q)}(x^{(q)}) . \quad (6)$$

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 due to the reduced number of variables. The specific form of $f^{(q)}(\cdot)$ generally results from the physical problem being solved with an appropriate grid spacing. In Section 5, we will give a typical example for ODT where $f^{(q)}(\cdot)$ is computed by discretizing the 3-D PDE using a grid spacing proportional to 2^q . The quantity $y^{(q)}$ in (6) 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 do not describe in this paper how to adjust the stabilizing functional. For an example of the choice of coarse scale stabilizing functional, see our previous works,^{7,14} where the generalized Gaussian Markov Random field model is used as the image prior model in MAP estimation.

In the remainder of this section, we explain how the cost functionals at each scale can be matched to produce a consistent solution. To do this, we define an adjusted cost functional

$$\begin{aligned} c^{(q)}(x^{(q)}) &= \tilde{c}^{(q)}(x^{(q)}) - r^{(q)}x^{(q)} \\ &= \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_{\Lambda}^2 + S^{(q)}(x^{(q)}) - r^{(q)}x^{(q)}, \end{aligned} \quad (7)$$

where $r^{(q)}$ is a row vector used to adjust the functional's 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 recursive expressions for the quantities $y^{(q)}$ and $r^{(q)}$ that match the cost functionals at fine and coarse scales.

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

$$\tilde{x}^{(q+1)} \leftarrow \text{Fixed_Grid_Update}(I_{(q)}^{(q+1)}x^{(q)}, c^{(q+1)}(\cdot)), \quad (8)$$

where $\tilde{x}^{(q+1)}$ is the updated value, and the operator $\text{Fixed_Grid_Update}(x_{init}, c(\cdot))$ is any fixed grid update algorithm designed to reduce the cost functional $c(\cdot)$ starting with the initial value x_{init} . In (8), the initial condition $I_{(q)}^{(q+1)}x^{(q)}$ is formed by decimating $x^{(q)}$. We may now use this result to update the finer grid solution. We do this by interpolating the *change* in the coarser scale solution.

$$\tilde{x}^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}). \quad (9)$$

Ideally, the new solutions $\tilde{x}^{(q)}$ should be at least as good as the old solution $x^{(q)}$. Specifically, we would like $c^{(q)}(\tilde{x}^{(q)}) \leq c^{(q)}(x^{(q)})$. 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 functionals is eliminated if the fine and coarse scale cost functionals are equal within an additive constant.[†] This means we would like

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \cong c^{(q)}(x^{(q)} + I_{(q+1)}^{(q)}(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)})) + \text{constant} \quad (10)$$

to hold for all values of $\tilde{x}^{(q+1)}$. Our objective is then to choose a coarse scale cost functional which matches the fine cost functional as described in (10). We do this by the proper selection of $y^{(q+1)}$ and $r^{(q+1)}$. First, we enforce the condition that the initial error between the forward model and measurements be the same at the coarse and fine scales, giving

$$y^{(q+1)} - f^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) = y^{(q)} - f^{(q)}(x^{(q)}). \quad (11)$$

[†]A constant offset has no effect on the value of x which minimizes the cost functional.

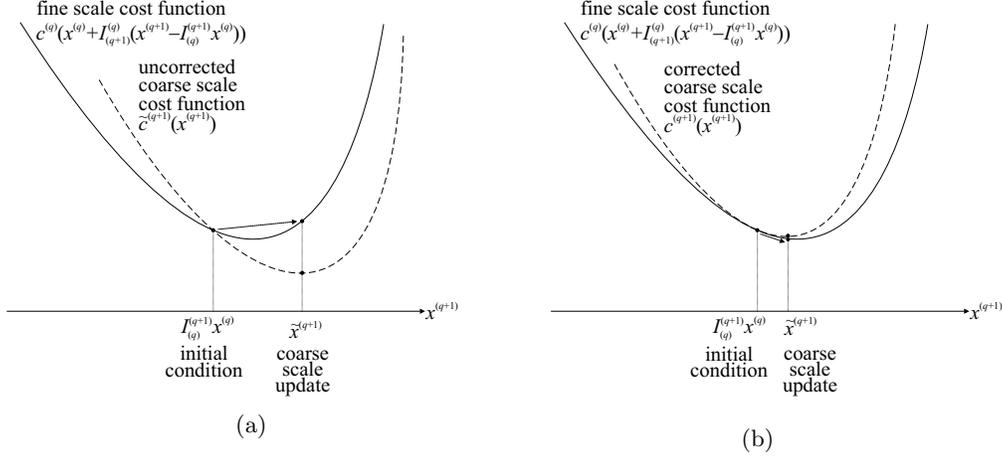


Figure 1. The role of adjustment term $r^{(q+1)}x^{(q+1)}$: (a) When the gradients of the fine scale and coarse scale cost functionals are different at the initial value, the updated value may increase the fine grid cost functional's value. (b) When the gradients of the two functionals are matched, a properly chosen coarse scale functional can guarantee that the coarse scale update reduces the fine scale cost.

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

$$y^{(q+1)} \leftarrow y^{(q)} - \left[f^{(q)}(x^{(q)}) - f^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) \right]. \quad (12)$$

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

Next, we use the condition introduced in^{7, 11–13} 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 match the derivatives of both sides in (10) with respect to $x^{(q+1)}$ at the current values, giving

$$\nabla^{(q+1)}c^{(q+1)}(x^{(q+1)}) = \nabla^{(q+1)}c^{(q)}(x^{(q)} + I_{(q+1)}^{(q)}(x^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)})) \quad (13)$$

at $x^{(q+1)} = I_{(q)}^{(q+1)}x^{(q)}$, where $\nabla^{(q+1)}$ is the spatial gradient operator with respect to $x^{(q+1)}$. The condition (13) is reduced to the condition that

$$\nabla^{(q+1)}c^{(q+1)}(x^{(q+1)}) \Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} = \nabla^{(q)}c^{(q)}(x^{(q)})I_{(q+1)}^{(q)}, \quad (14)$$

where the interpolation operator $I_{(q+1)}^{(q)}$ actually functions as a decimation operator because it is being multiplied by the gradient vector from the left. This condition is essential to assuring that the optimum solution is a fixed point of the multigrid inversion algorithm,⁷ and is illustrated graphically in Fig. 1. Figure 1(a) shows the case in which the gradients of the fine scale and coarse scale functions are different at the initial value. In this case, the surrogate function can not upper bound the value of the fine scale functional, and the updated value may actually increase the fine grid cost functionals value. Figure 1(b) illustrates the case in which the gradients of the two functionals are matched. In this case, a properly chosen coarse scale functional can upper bound the fine scale functional, and the coarse scale update is guaranteed to reduce the fine scale cost. The equality of (14) can be enforced at the current value $x^{(q)}$ by choosing

$$r^{(q+1)} \leftarrow \nabla^{(q+1)}\tilde{c}^{(q+1)}(x^{(q+1)}) \Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} - \left(\nabla^{(q)}\tilde{c}^{(q)}(x^{(q)}) - r^{(q)} \right) I_{(q+1)}^{(q)}, \quad (15)$$

```

 $x^{(q)} \leftarrow \text{MultigridV}(q, x^{(q)}, y^{(q)}, r^{(q)}) \{$ 
  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 (12)
  Compute  $r^{(q+1)}$  using (15)
   $x^{(q+1)} \leftarrow \text{MultigridV}(q + 1, x^{(q+1)}, y^{(q+1)}, r^{(q+1)})$  //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 2: Pseudo-code specification of the subroutine for the Multigrid-V inversion.

where $\tilde{c}^{(q)}(\cdot)$ is the unadjusted cost functional defined in (6).

The multigrid V algorithm⁵ is obtained by applying this two-grid algorithm recursively in resolution, as shown in the pseudocode in Fig. 2. After initialization of $r^{(0)} \leftarrow 0$ and $y^{(0)} \leftarrow y$, we can then minimize (4) through iterative application of the $\text{multigridV}(\cdot)$ subroutine at resolution 0. The Multigrid-V algorithm then moves from fine to coarse to fine resolutions with each iteration. 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. Notice that $\nu_1^{(q)}$ fixed grid iterations are done before the coarse grid correction, and that $\nu_2^{(q)}$ iterations are done afterwards. The convergence speed of the algorithm can be tuned through the choice of $\nu_1^{(q)}$ and $\nu_2^{(q)}$ at each scale.

4.2. Relations with Other Approaches

4.2.1. Surrogate Functional Approaches

Multigrid inversion proposed in Section 4.1 can be viewed as a method to simplify a potentially expensive optimization by temporarily replacing the original cost functional by a lower resolution one, which is computationally more tractable and preserves its local properties such as the first derivative. In fact, there is a large class of optimization methods which depend on the use of so-called surrogate functionals, or functional substitution methods to speed or simplify optimization. A classic example of a surrogate functional is the Q-function used in the EM algorithm.^{15,16} More recently, De Pierro discovered that this same basic method could be applied to tomography problems in a manner that allowed parallel updates of pixels in the computation of penalized ML reconstructions.^{17,18} De Pierro's method has since been exploited to both prove convergence and allow parallel updates for ICD methods in tomography.^{19,20}

However, the application of surrogate functionals to multigrid inversion is unique in that the substituting functional is at a coarser scale and therefore has an argument of lower dimension. As with traditional approaches, the surrogate functional should be designed to guarantee monotone convergence of the original cost functional. In the case of the multigrid algorithm, a sequence of optimization functionals at varying resolutions should be designed so that the entire multigrid update decreases the finest resolution cost function.

4.2.2. Multigrid Solver Approaches

It is interesting to relate the multigrid inversion to more traditional multigrid solver approaches. In solver approaches, we may use the approximate coarse scale cost functional of the same form as the finest scale

original one. For example, for the case of the cost functional $c(x)$ in (4), the functional $\tilde{c}^{(q)}(x^{(q)})$ in (6) can be used for all resolutions, without the linear correction term $r^{(q)}x^{(q)}$ in (7).

Generally, the cost functional $\tilde{c}^{(q)}(x^{(q)})$ is the integral of some physical quantity over some spatial domain, and the unknown $x_s^{(q)}$ is sampled on a uniform grid with dimension d and sampling period of 2^q . Therefore, one would expect the partial derivative of

$$\frac{\partial \tilde{c}^{(q)}(x^{(q)})}{\partial x_s^{(q)}}$$

to scale linearly with the size of the voxel indexed by s . Using this assumption, we define the scaled gradient as

$$N^{(q)}(x^{(q)}) = 2^{-qd} \nabla^{(q)} \tilde{c}^{(q)}(x^{(q)}) , \quad (16)$$

where 2^{qd} is the size of a voxel for scale q and voxel dimension d . The factor 2^{-qd} compensates for the increase in the gradient as a function of voxel volume. If we further assume that the interpolation operator $I_{(q)}^{(q+1)}$ is chosen to preserve average values (i.e. its rows sum to 1), then we would expect

$$N^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) \simeq N^{(q)}(x^{(q)})[I_{(q)}^{(q+1)}]^T , \quad (17)$$

where $N^{(q)}(\cdot)$ is assumed to be a row vector, so we multiply by the transpose of the interpolation operator.

Suppose that we would like to solve an optimality condition at scale q

$$N^{(q)}(x^{(q)}) = h^{(q)} , \quad (18)$$

where $h^{(q)}$ is a row vector used to compensate for equation mismatch between resolutions, which we must derive. We set $h^{(0)} = 0$, so that (18) is the same as (5) at the finest scale. This section will show that (18) corresponds to the optimality condition for $c^{(q)}(\cdot)$ in multigrid inversion, by showing that $h^{(q)} = 2^{-qd} \cdot r^{(q)}$ for reasonable choices of interpolation and decimation operators.

To solve (18), nonlinear multigrid solvers correct fine scale solutions by coarse grid correction (9) with the coarse scale solution to the optimality condition at scale $q + 1$. The coarse scale optimality condition is given by

$$N^{(q+1)}(x^{(q+1)}) = h^{(q+1)}[I_{(q)}^{(q+1)}]^T + \left[N^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) - N^{(q)}(x^{(q)})[I_{(q)}^{(q+1)}]^T \right] , \quad (19)$$

where the term in the bracket guarantees that the true solution of the fine scale equation (18) is a fixed point of (19). The equation (19) can be expressed as

$$N^{(q+1)}(x^{(q+1)}) = h^{(q+1)} , \quad (20)$$

where

$$\begin{aligned} h^{(q+1)} &= N^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) - \left[N^{(q)}(x^{(q)}) - h^{(q)} \right] [I_{(q)}^{(q+1)}]^T \\ &= 2^{-(q+1)d} \left[\nabla^{(q+1)} \tilde{c}^{(q+1)}(I_{(q)}^{(q+1)}x^{(q)}) - \left[\nabla^{(q)} \tilde{c}^{(q)}(x^{(q)}) - 2^{qd} h^{(q)} \right] 2^d [I_{(q)}^{(q+1)}]^T \right] . \end{aligned} \quad (21)$$

Many multigrid methods^{5,7} typically choose linear decimation and interpolation operators so that they satisfy

$$I_{(q+1)}^{(q)} = 2^d [I_{(q)}^{(q+1)}]^T \quad (22)$$

except at boundaries. Using (22) we see that $h^{(q+1)} = 2^{-(q+1)d} \cdot r^{(q+1)}$, and thus (20) is the same as the optimality condition for $c^{(q+1)}$ in (7). Therefore, this solver approach can be interpreted as an application of multigrid optimization. However, it is interesting to note multigrid optimization has the advantages that it can be used for a wide variety of interpolation and decimation operators, and is not restricted to the constraint of (22).

5. APPLICATION TO OPTICAL DIFFUSION TOMOGRAPHY

5.1. Optical Diffusion Tomography

Optical diffusion tomography(ODT) is a method for measuring cross-sections of optical properties from the measurement 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 consider only reconstruction of the absorption coefficient. The 3-D domain is discretized, and the set of unknown μ_a at each discrete grid point forms a 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

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

where r is position and c is the speed of light in the medium. Then the measurement at a detector 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 model⁷ and the GGMRF image prior model, the MAP estimate of x in a Bayesian framework is reduced to the minimization⁷ of

$$c(x) = \frac{P}{2} \log \|y - f(x)\|_{\Lambda}^2 + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p , \quad (23)$$

where Λ is diagonal with $\Lambda_{(i,i)} = 1/|y_i|$, the set \mathcal{N} consists of all pairs of adjacent grid points, b_{i-j} represents the weighting assigned to the pair $\{i, j\}$, σ is a parameter that controls the overall weighting, and $1 \leq p \leq 2$ controls the degree of edge smoothness. We use multigrid inversion to solve the required optimization problem using the set of coarse grid cost functionals developed in Section 4. For fixed grid updates at each scale q , we use the ICD algorithm which also incorporates sequential updates of α and x . See Ye *et al.*^{7,21} for details of the fixed grid ICD algorithm.

5.2. Numerical Results

We examined the performance of the multigrid inversion for the ODT problem. A cubic tissue phantom of dimension $10 \times 10 \times 10$ cm on an edge was used for reconstructions. Fig. 3(a) shows its cross-section for the plane $z = 0$, where the origin is at the center of the phantom. The μ_a background was linearly varied from 0.01 cm^{-1} to 0.04 cm^{-1} in the x-direction, except for the outermost region with width 1.25 cm, which was homogeneous with $\mu_a = 0.025 \text{ cm}^{-1}$. Two spherical μ_a inhomogeneities have diameters of 1.85 cm and values of $\mu_a = 0.1 \text{ cm}^{-1}$ (left-top) and $\mu_a = 0.12 \text{ cm}^{-1}$ (right-bottom), and their centers are on the plane $z = 0$. The diffusion coefficient D was homogeneous with $D = 0.03 \text{ cm}$. To simulate the measurements, we solved the diffusion equations with $257 \times 257 \times 257$ grid points. Since 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. 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. 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.

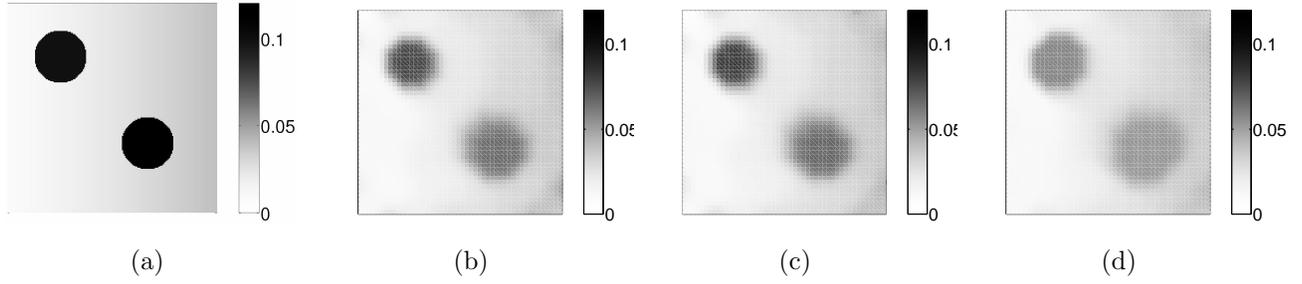


Figure 3. Cross-sections of the $z = 0$ plane : (a) true phantom (b) 4 level multigrid reconstructions with 19.35 iterations, (c) 3 level multigrid reconstructions with 19.95 iterations, (d) fixed-grid reconstructions with 270 iterations.

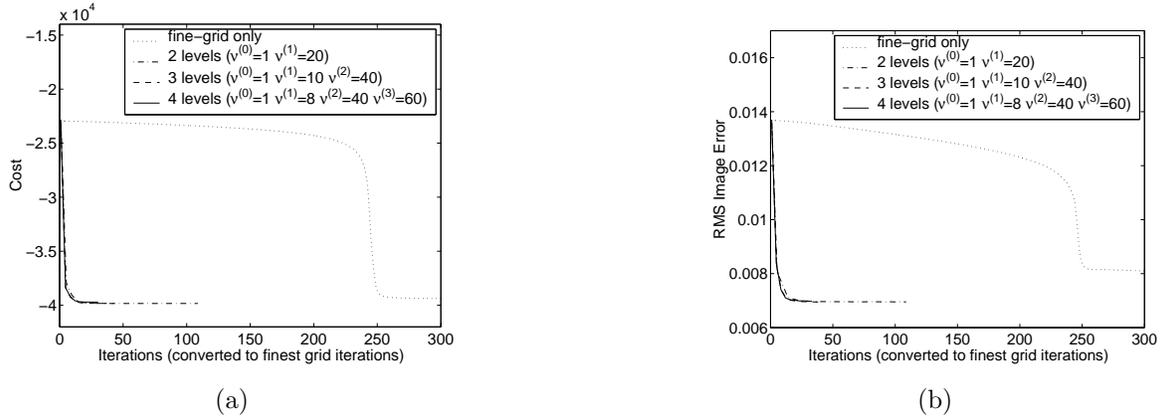


Figure 4. Convergence of (a) cost and (b) RMS image error when reconstructions were initialized with average values.

All comparisons of fixed-grid and multigrid inversion algorithms were made at $65 \times 65 \times 65$ resolution. The reconstruction was performed for all nodes except the eight outermost layers of grid points. These border regions were fixed to their true values in order to avoid singularities near the sources and detectors. These region have also been excluded from all cross-section figures and the evaluation of root-mean-square(RMS) reconstruction error. In order to make fair comparisons of computational speed, we scale the number of iterations for all methods into units of single fixed grid iterations at the finest scale. The image prior model used $p = 1.2$ and $\sigma = 0.018 \text{ cm}^{-1}$.

For the first experiment, all algorithms were initialized with the average values of the true phantom, which were $\mu_a = 0.026 \text{ cm}^{-1}$ and $D = 0.03 \text{ cm}$.[‡] Figure 4 shows that the multigrid algorithms converged much faster than the fixed grid algorithm, both in the sense of cost and RMS error. The multigrid algorithms converged in only 20 iterations, while the fixed algorithm required 270 iterations. Even after 200 iterations, the fixed grid algorithm still changed very little in the convergence plots.

Figure 3 shows reconstructions produced by the four algorithms. The reconstructed image quality for all three multigrid algorithms is nearly identical, but the reconstructed quality is significantly worse for the fixed grid algorithm. In fact, the multigrid algorithms converged to slightly lower values of the cost functional (-39833 to -39763) than the fixed-grid algorithm (-39392), and the RMS image error for the multigrid reconstructions

[‡]In practice, this is not possible since the average value is not known, but it was done because it favors the fixed-grid algorithm.

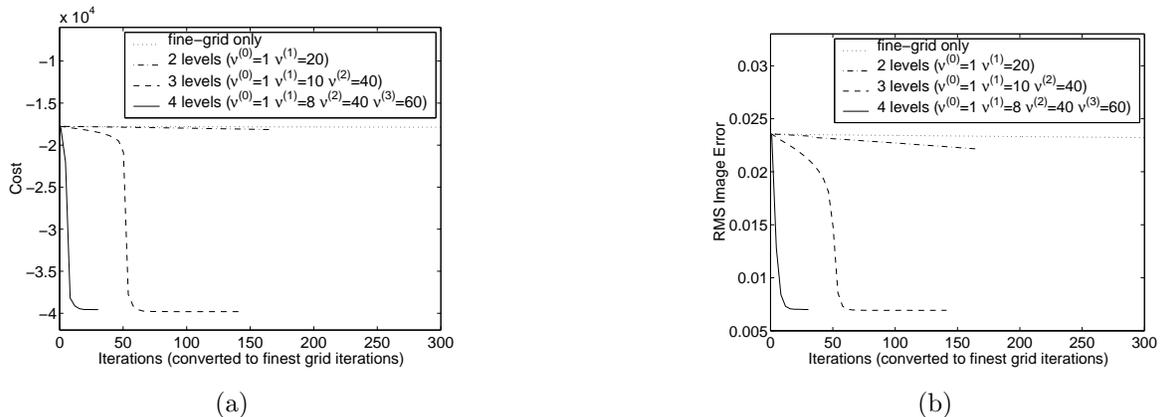


Figure 5: Convergence of (a) cost and (b) RMS image error with a poor initial guess.

ranged from 0.0069 to 0.007, while the fixed algorithm converged to the higher RMS error of 0.0081.

To investigate the sensitivity of convergence with respect to initialization, we performed reconstructions with a poor initial estimate. The initial image was homogeneous, with a value of 1.75 times the true phantom’s average value. The plots in Fig. 5 show that the three and four level multigrid algorithms converged rapidly. In particular, the four level multigrid algorithm converges almost as rapidly as it did when initialized with the true phantom’s average value. The fixed grid algorithm changed very little from the initial estimate even after 300 iterations, and the two grid algorithm progressed slowly. These results suggest that higher level multigrid algorithms are necessary to overcome the effects of a poor initial estimate.

6. CONCLUDING REMARKS

Multigrid methods are powerful to solve optimization problems resulting from inverse problems. We have formulated multigrid algorithms for nonlinear inverse problems directly in an optimization framework. The nonlinear multigrid inversion works by simultaneously varying the resolution of both the forward solution and inverse computation, preserving the nonlinearity of the forward model. Multigrid inversion is formulated in a general framework and is applicable to a wide variety of inverse problems, but it is particularly well suited for the inversion of nonlinear forward models such as those modeled by the solution of PDEs. The application of multigrid inversion to optical diffusion tomography (ODT) indicate that multigrid inversion can dramatically reduce computation, particularly if the reconstruction resolution is high, and the initial condition is inaccurate. Perhaps more importantly, multigrid inversion showed robust convergence under a variety of conditions and while solving an optimization problem that is subject to local minima.

REFERENCES

1. D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, “Imaging the body with diffuse optical tomography,” *IEEE Signal Proc. Magazine* **18**, pp. 57–75, Nov. 2001.
2. G. J. Saulnier, R. S. Blue, J. C. Newell, D. Isaacson, and P. M. Edic, “Electrical impedance tomography,” *IEEE Signal Proc. Magazine* **18**, pp. 31–43, Nov. 2001.
3. J. S. Reynolds, A. Przadka, S. Yeung, and K. J. Webb, “Optical diffusion imaging: a comparative numerical and experimental study,” *Appl. Optics* **35**, pp. 3671–3679, July 1996.

4. K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing* **41**, pp. 534–548, February 1993.
5. W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial, 2nd Ed.*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.
6. C. A. Bouman and K. Sauer, "Nonlinear multigrid methods of optimization in Bayesian tomographic image reconstruction," in *Proc. of SPIE Conf. on Neural and Stochastic Methods in Image and Signal Processing*, vol. 1766, pp. 296–306, (San Diego, CA), July 19–24 1992.
7. 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 Processing* **10**, pp. 909–922, June 2001.
8. S. F. McCormick and J. G. Wade, "Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography," *Inverse Problems* **9**, pp. 697–713, 1993.
9. T. Dreyer, B. Maar, and V. Schulz, "Multigrid optimization in applications," *J. Computational and Applied Math.* **120**, pp. 67–84, 2000.
10. L. Borcea, "Nonlinear multigrid for imaging electrical conductivity and permittivity at low frequency," *Inverse Problems* **17**, pp. 329–359, April 2001.
11. S. G. Nash, "A multigrid approach to discretized optimization problems," *J. of Optimization methods and software* **14**, pp. 99–116, 2000.
12. J. C. Ye, C. A. Bouman, R. P. Millane, and K. J. Webb, "Nonlinear multigrid optimization for Bayesian diffusion tomography," in *Proc. of IEEE Int'l Conf. on Image Proc.*, (Kobe, Japan), October 25–28 1999.
13. R. M. Lewis and S. G. Nash, "A multigrid approach to the optimization of systems governed by differential equations," in *8-th AIAA/USAF/ISSMO Symp. Multidisciplinary Analysis and Optimization*, (Long Beach, CA), 2000.
14. S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "Multigrid inversion algorithms with applications to optical diffusion tomography," in *Proc. of 36th Asilomar Conference on Signals, Systems, and Computers*, (Monterey, CA), Nov. 2002.
15. L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statistics* **37**, pp. 1554–1563, 1966.
16. L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statistics* **41**(1), pp. 164–171, 1970.
17. A. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. on Medical Imaging* **14**, pp. 132–137, March 1995.
18. J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. on Medical Imaging* **15**, pp. 687–699, October 1996.
19. J. Fessler, E. Ficareo, N. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. on Medical Imaging* **16**, pp. 166–175, April 1997.
20. J. Zheng, S. S. Saquib, K. Sauer, and C. A. Bouman, "Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Trans. on Image Processing* **9**, pp. 1745–1759, Oct. 2000.
21. 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. Optical Society America A* **16**, pp. 2400–2412, October 1999.