A General Framework for Nonlinear Multigrid Inversion

Seungseok Oh, *Student Member, IEEE*, Adam B. Milstein, *Student Member, IEEE*, Charles A. Bouman, *Fellow, IEEE*, and Kevin J. Webb, *Fellow, IEEE*

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 is particularly difficult because the forward problem is both nonlinear and computationally expensive to evaluate. In this paper, we propose a general framework for nonlinear multigrid inversion that is applicable to a wide variety of inverse problems. The multigrid inversion algorithm results from the application of recursive multigrid techniques to the solution of optimization problems arising from inverse problems. The method works 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 inversion algorithm efficiently computes the solution to the desired fine-scale inversion problem. Importantly, the new 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 Bayesian optical diffusion tomography with a generalized Gaussian Markov random-field image prior model shows the potential for very large computational savings. Numerical data also indicates robust convergence with a range of initialization conditions for this nonconvex optimization problem.

Index Terms—Inverse problems, multigrid algorithms, multiresolution, optical diffusion tomography (ODT).

I. 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]–[4] and electrical impedance tomography (EIT) [5], have received much attention. For example, ODT holds great potential as a safe, noninvasive medical diagnostic modality with chemical specificity [6]. 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. Finally, these problems typically are inherently three-dimensional (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, numerous other inverse problems have similar computational difficulties, including microwave tomography [7], thermal wave tomography [8], and inverse scattering [9].

To solve inverse problems, most algorithms, such as conjugate gradient (CG), steepest descent (SD), and iterative coordinate descent (ICD) [10] work by performing all computations using a fixed discretization grid. While tremendous progress has been made in reducing the computational complexity of these fixed-grid methods, computational cost is still of great concern. Perhaps, more importantly, fixed-grid optimization methods are essentially performing a local search of the cost function and are, therefore, more susceptible to being trapped in local minima that can result in poorer quality reconstructions.

Multiresolution techniques have been widely investigated to reduce computation for inverse problems. Even simple multiresolution approaches, such as initializing fine resolution iterations with coarse solutions [11]–[15], have been shown to be effective in many imaging problems. Wavelets have been studied for Bayesian tomography [16]–[20], and both wavelet and multiresolution models have been applied in Bayesian formulations of emission tomography [21]–[24] and thermal wave tomography [25]. For ODT, a two resolution wavelet decomposition was used to speed inversion of a problem linearized with a Born approximation [26].

Multigrid methods are a special class of multiresolution algorithms which work by recursively operating on the data at different resolutions, using the ideas of nested iterations and coarse grid correction [27]–[32]. Multigrid algorithms originally attracted interest as a method for solving PDEs by effectively removing smooth error components, which are not always damped in fixed-grid relaxation schemes. In particular, the full approximation scheme (FAS) of Brandt [27] can be used to solve nonlinear PDEs. Multigrid methods have been used to expedite convergence in various image processing problems, for example, lightness computation [33], shape-from-X [33], [34], optical flow estimation [33], [35]–[38], signal/image smoothing [39], [40], image segmentation [40], [41], image matching [42], image restoration [43], anisotropic diffusion [44], sparse-data surface representation [45], interpolation of missing image data [40], [46], and image binarization [34].

More recently, multigrid algorithms have been used to solve image reconstruction problems. Bouman and Sauer showed that nonlinear multigrid algorithms could be applied to inversion of Bayesian tomography problems [47]. This work used nonlinear multigrid techniques to compute maximum *a posteriori* (MAP) reconstructions with non-Gaussian prior distributions and a non-negativity constraint. McCormick and Wade [48] applied multigrid methods to a linearized EIT problem, and Borcea [49] used a nonlinear multigrid approach to EIT based on a direct nonlinear formulation analogous to FAS in nonlinear multigrid PDE solvers. Brandt *et al.* developed multigrid methods for EIT [50] and atmospheric data assimilation [51] and applied multigrid or multiscale methods to various numerical computation problems including inverse problems [52], [53]. Johnson *et al.* [54] applied an algebraic multigrid algorithm to inverse bioelectric field problems formulated with the finite-element method. In [55], [56], Ye *et al.* formulated the multigrid approach directly in an optimization framework, and used the method to solve ODT problems. In related work, Nash and Lewis formulated multigrid algorithms for the solution of a broad class of optimization problems [57], [58]. Importantly, both the approaches of Ye and Nash are based on the matching of cost functional derivatives at different scales.

In this paper, we propose a method we call multigrid inversion [59]–[62]. Multigrid inversion is a general approach for 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. This makes our method particularly well suited to the solution of inverse problems with PDE forward models for a number of reasons, as follows.

1) 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 [48], [55], [56].
2) The coarse grid forward model can be modeled by a correctly discretized PDE, preserving the nonlinear characteristics of the forward model.
3) A wide variety of optimization methods can be used for solving the inverse problem at each grid. Hence, common methods, such as pre-conditioned conjugate gradient and/or adjoint differentiation [63], [64] can be employed at each grid resolution.

While the multigrid inversion method is motivated by the solution of inverse problems, such as ODT and EIT, it is generally applicable to any inverse problem in which the forward model can be naturally represented at differing grid resolutions.

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

dates will not change an exact solution to the fine grid problem, i.e., that the exact fine grid solution is always a fixed point of the multigrid algorithm. In addition, we show that under certain conditions, the nonlinear multigrid inverse algorithm is guaranteed to produce monotone convergence of the fine grid cost functional. We present experimental results for the ODT application which show that the multigrid inversion algorithm can provide dramatic reductions in computation when the inversion problem is solved at the resolution necessary to achieve a high-quality reconstruction.

This paper is organized as follows. Section II introduces the general concept of the multigrid inversion algorithm and Section II-D discusses its convergence. In Section III, we illustrate the application of the multigrid inversion method to the ODT problem and its numerical results are provided in Section IV. Finally, Section V makes concluding remarks.

II. MULTIGRID INVERSION FRAMEWORK

In this section, we overview regularized inverse methods and then formulate the general multigrid inversion approach.

A. 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 a forward model $f(x)$ given by

$$E[Y|x] = f(x) \tag{1}$$

which represents the computed means of the measurements given the image x. For many inverse problems, such as ODT, the forward model $f(x)$ is given by the solution of a PDE where x determines the coefficients of the discretized PDE. We will assume that the measurements Y are conditionally Gaussian given x, so that

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

where Λ is a positive definite weight matrix, P is the dimensionality of the measurement, α is a parameter proportional to the noise variance, and $\|w\|_\Lambda^2 = w^H \Lambda w$. Note that the measurement noise covariance matrix is equal to $\alpha\Lambda^{-1}$. When the data values are real valued, P is equal to the length of the vector Y, but when the measurements are complex, then P is equal to twice the dimension of Y.

Our objective is to invert the forward model of (1) and thereby estimate x from a particular measurement vector y. There are a variety of methods for performing this estimation, including maximum *a posteriori* (MAP) estimation, penalized maximum likelihood, and regularized inversion. All of these methods work by computing the value of x which minimizes a cost functional of the form

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

where $S(x)$ is a stabilizing functional used to regularize the inverse. Note that in the MAP approach, $S(x) = -\log p(x)$,

where $p(x)$ is the prior distribution assumed for x. We will estimate both the noise variance parameter α and x by jointly maximizing over both quantities [65]. Minimization of (3) with respect to α yields the condition $\hat{\alpha} = (1/P)\|y - f(x)\|_\Lambda^2$. Substitution of $\hat{\alpha}$ into (3) and dropping constants yields the cost functional to be optimized as

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

where we will generally assume $c(x)$ is a continuously differentiable function of x.

We have found that joint optimization over α and x has a number of important advantages. First, in many applications, the absolute magnitude of the measurement noise is not known in advance, while the relative noise magnitude may be known. In such a scenario, it is useful to simultaneously estimate the value of α along with the value of x [55], [56], [66]. More importantly, we have found that the logarithm in the expression of (4) makes optimization less susceptible to being trapped in local minima. In any case, the multigrid methods we describe are equally applicable to the case when α is fixed. In this case, the cost functional is given by $c(x) = (1/2\alpha)\|y - f(x)\|_\Lambda^2 + S(x)$, instead of (4).

B. Fixed-Grid Inversion

Once the cost functional of (4) is formulated, the inverse is computed by solving the associated optimization problem

$$\hat{x} = \arg \min_x \left\{ \frac{P}{2} \log \|y - f(x)\|_\Lambda^2 + S(x) \right\}. \qquad (5)$$

Most optimization algorithms, such as CG, SD, and ICD, work by iteratively minimizing the cost functional. We express a single iteration of such a fixed-grid optimizer as

$$x_{\text{update}} \leftarrow \text{Fixed\_Grid\_Update}(x_{\text{init}}, c(\cdot)) \qquad (6)$$

where $c(\cdot)$ is the cost functional being minimized, x_{init} is the initial value of x, and x_{update} is the updated value.[1] We will generally assume that the fixed-grid algorithm reduces the cost functional with each iteration, unless the initial value of x is at a local minimum of the cost functional. Therefore, we say that an update algorithm is *monotone* if $c(x_{\text{update}}) \leq c(x_{\text{init}})$, with strict inequality when $\nabla c(x_{\text{init}}) \neq 0$ or $x_{\text{update}} \neq x_{\text{init}}$. Repeated application of a monotone fixed-grid optimizer will produce a sequence of estimates with monotonically decreasing cost. Thus, we may approximately solve (5) through iterative application of (6).

In many inverse problems, such as ODT, the forward model computation requires the solution of a 3-D PDE which must be discretized for numerical solution on a computer. Although a fine discretization 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. For a 3-D problem, the computational cost

typically increases by a factor of 8 each time the resolution is doubled. Solving problems at fine resolution also tends to slow convergence. For example, many fixed-grid algorithms such as ICD[2] effectively eliminate error at high spatial frequencies, but low-frequency errors are damped slowly [10], [29].

C. Multigrid Inversion Algorithm

In this section, we derive the basic multigrid inversion algorithm for solving the optimization of (5). 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, as

$$\tilde{c}^{(q)}(x^{(q)}) = \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_\Lambda^2 + S^{(q)}(x^{(q)}). \quad (7)$$

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 III, 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 (7) denotes an adjusted measurement vector at scale q. Note that, in this work, we assume that $y^{(q)}$ and $f^{(q)}(\cdot)$ are of the same length at every scale q, so that the data resolution is not a function of 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 II-E.

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

$$\begin{aligned} c^{(q)}(x^{(q)}) &= \tilde{c}^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} \\ &= \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_\Lambda^2 \\ &\quad + S^{(q)}(x^{(q)}) - r^{(q)} x^{(q)} \end{aligned} \qquad (8)$$

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

---

[1]We use the $\leftarrow$ symbol to denote assignment of a value to a variable, thereby eliminating the need for time indexing in update equations.

[2]ICD is generally referred to as Gauss-Seidel in the PDE literature literature.

result to correct the finer grid solution. This coarse grid update is

$$\tilde{x}^{(q+1)} \leftarrow \text{Fixed\_Grid\_Update}\left(I_{(q)}^{(q+1)}x^{(q)}, c^{(q+1)}(\cdot)\right) \quad (9)$$

where $I_{(q)}^{(q+1)}x^{(q)}$ 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 by

$$\tilde{x}^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}\left(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}\right). \quad (10)$$

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 is monotone. 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.[3] This means we would like

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \cong c^{(q)}\left(x^{(q)}\right.$$
$$\left. + I_{(q+1)}^{(q)}\left(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}\right)\right) + \text{constant} \quad (11)$$

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 (11). 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

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

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

$$y^{(q+1)} \leftarrow y^{(q)} - \left[f^{(q)}(x^{(q)}) - f^{(q+1)}\left(I_{(q)}^{(q+1)}x^{(q)}\right)\right]. \quad (13)$$

The term in the square brackets in (13) compensates for the forward model mismatch between resolutions.

Next, we use the condition introduced in [55]–[58] 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

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

where $\nabla c(x)$ is the row vector formed by the gradient of the functional $c(\cdot)$. This condition is essential to assure that the optimum solution is a fixed point of the multigrid inversion algorithm [56] and is illustrated graphically in Fig. 1. In Section II-D, we will also show how this condition can be used

---

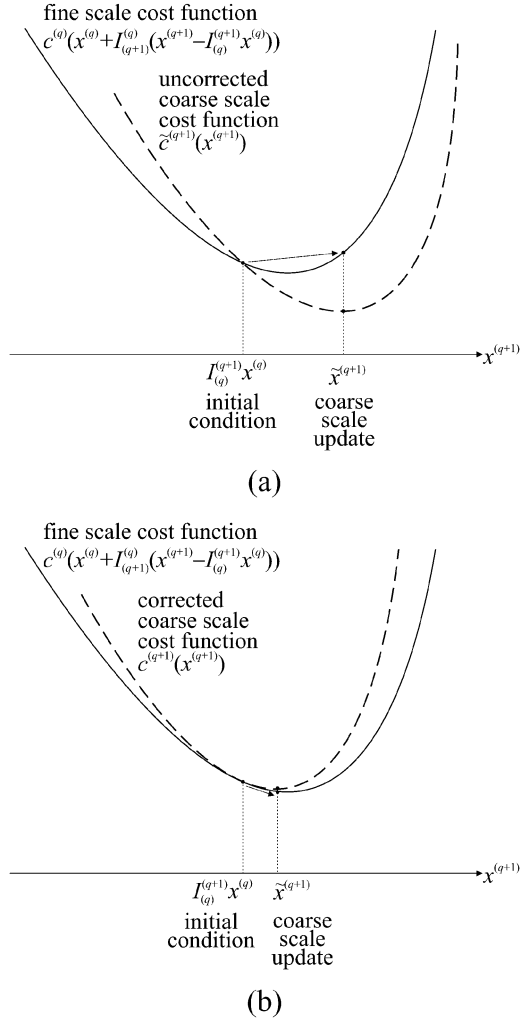[3] A constant offset has no effect on the value of x which minimizes the cost functional.

Fig. 1. 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.

along with other assumptions to ensure monotone convergence of the multigrid inversion algorithm. Note that in (14), the interpolation matrix $I_{(q+1)}^{(q)}$, which comes from the chain rule of differentiation, actually functions like a decimation operator because it multiplies the gradient vector on the right. Importantly, the condition (14) holds for any choice of decimation and interpolation matrices.

The equality of (14) can be enforced at the current value $x^{(q)}$ by choosing

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

where $\tilde{c}^{(q)}(\cdot)$ is the unadjusted cost functional defined in (7). By evaluating the gradients and using the update relation of (15), we obtain

$$r^{(q+1)} \leftarrow g^{(q+1)} - \left(g^{(q)} - r^{(q)}\right)I_{(q+1)}^{(q)} \quad (16)$$

$x^{(q)} \leftarrow$ Twogrid_Update$(q, x^{(q)}, y^{(q)}, r^{(q)})$ {

    Repeat $\nu_1^{(q)}$ times

        $x^{(q)} \leftarrow$ Fixed_Grid_Update$(x^{(q)}, c^{(q)}(\,\cdot\,; y^{(q)}, r^{(q)}))$ //Fine grid update

    $x^{(q+1)} \leftarrow I_{(q)}^{(q+1)} x^{(q)}$ //Decimation

    Compute $y^{(q+1)}$ using (13)

    Compute $r^{(q+1)}$ using (16)

    Repeat $\nu_1^{(q+1)}$ times

        $x^{(q+1)} \leftarrow$ Fixed_Grid_Update$(x^{(q+1)}, c^{(q+1)}(\,\cdot\,; y^{(q+1)}, r^{(q+1)}))$ //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

        $x^{(q)} \leftarrow$ Fixed_Grid_Update$(x^{(q)}, c^{(q)}(\,\cdot\,; y^{(q)}, r^{(q)}))$ //Fine grid update

    Return $x^{(q)}$ //Return result

}

Fig. 2. Pseudocode specification of a two-grid inversion algorithm. The notation $c^{(q+1)}(x^{(q+1)}; y^{(q+1)}, r^{(q+1)})$ is used to make the cost functional's dependency on $y^{(q+1)}$ and $r^{(q+1)}$ explicit.

main( ) {

    Initialize $x^{(0)}$ with a background estimate

    $r^{(0)} \leftarrow 0$

    $y^{(0)} \leftarrow y$

    Choose number of fixed grid iterations $\nu_1^{(0)}, \ldots, \nu_1^{(Q-1)}$ and $\nu_2^{(0)}, \ldots, \nu_2^{(Q-1)}$

    Repeat until converged:

        $x^{(0)} \leftarrow$ MultigridV$(q, x^{(0)}, c^{(0)}(\,\cdot\,; y^{(0)}, r^{(0)}))$

}

(a)

$x^{(q)} \leftarrow$ MultigridV$(q, x^{(q)}, y^{(q)}, r^{(q)})$ {

    Repeat $\nu_1^{(q)}$ times

        $x^{(q)} \leftarrow$ 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 (13)

    Compute $r^{(q+1)}$ using (15)

    $x^{(q+1)} \leftarrow$ 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

        $x^{(q)} \leftarrow$ Fixed_Grid_Update$(x^{(q)}, c^{(q)}(\,\cdot\,; y^{(q)}, r^{(q)}))$ //Fine grid update

    Return $x^{(q)}$ //Return result

}

(b)

Fig. 3. Pseudocode specification of (a) the main routine for multigrid inversion and (b) the subroutine for the Multigrid-V inversion. The Multigrid-V algorithm is similar to the two-grid algorithm, but recursively calls itself to perform the coarse grid update.

where $g^{(q)}$ and $g^{(q+1)}$ are the gradients of the unadjusted cost functional at the fine and coarse scales, respectively, given by

$$g^{(q)} = -\frac{P}{\|y^{(q)} - f^{(q)}(x^{(q)})\|_\Lambda^2}$$
$$\times \mathrm{Re}\left\{\left(y^{(q)} - f^{(q)}(x^{(q)})\right)^H \Lambda A^{(q)}\right\}$$
$$+ \nabla S^{(q)}(x^{(q)}) \tag{17}$$

$$g^{(q+1)} = -\frac{P}{\|y^{(q)} - f^{(q)}(x^{(q)})\|_\Lambda^2}$$
$$\times \mathrm{Re}\left\{\left(y^{(q)} - f^{(q)}(x^{(q)})\right)^H \Lambda A^{(q+1)}\right\}$$
$$+ \nabla S^{(q+1)}\left(I_{(q)}^{(q+1)} x^{(q)}\right) \tag{18}$$

where H is the conjugate transpose (Hermitian) operator and

$A^{(q)}$ denotes the gradient of the forward model or Fréchet derivative given by

$$A^{(q)} = \nabla f^{(q)}(x^{(q)}) \tag{19}$$

$$A^{(q+1)} = \left. \nabla f^{(q+1)}(x^{(q+1)}) \right|_{x^{(q+1)} = I_{(q)}^{(q+1)} x^{(q)}}. \tag{20}$$

As a summary of this section, Fig. 2 shows the pseudocode for implementing the two-grid algorithm. 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.

The Multigrid-V algorithm [29] is obtained by simply replacing the fixed-grid update at resolution $q + 1$ of the two-grid algorithm with a recursive subroutine call, as shown in the pseudocode in Fig. 3(b). We can then solve (5) through iterative application of the Multigrid-V algorithm, as shown in Fig. 3(a). The Multigrid-V algorithm then moves from fine to coarse to fine resolutions with each iteration.

D. Convergence of Multigrid Inversion

Multigrid inversion can be viewed as a method to simplify a potentially expensive optimization by temporarily replacing the original cost functional by a lower resolution one. 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 [68], [69]. 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 [70], [71]. De Pierro's method has since been exploited to both prove convergence and allow parallel updates for ICD methods in tomography [72], [73].

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.

Fig. 1 graphically illustrates the use of surrogate functionals in multigrid inversion. Fig. 1(a) shows the case in which the gradients of the fine scale and coarse scale (i.e., surrogate) 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 functional's value. Fig. 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 concepts illustrated in Fig. 1 can be formalized into conditions that guarantee the monotone convergence of the multigrid algorithms. The following theorem, proved in Appendix I, gives a set of sufficient conditions for monotone convergence of the multigrid inversion algorithm.

Theorem: (Multigrid Monotone Convergence): For $0 \le q < Q - 1$, define the functional $\xi^{(q+1)} : \mathbb{R}^{N^{(q+1)}} \to \mathbb{R}$

$$\xi^{(q+1)}(x^{(q+1)}) = \tilde{c}^{(q+1)}(x^{(q+1)})$$
$$- \tilde{c}^{(q)}\left(x^{(q)} + I_{(q+1)}^{(q)}\left(x^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)} \right) \right) \tag{21}$$

where $N^{(q+1)}$ is the number of voxels in $x^{(q+1)}$, $\mathbb{R}$ is the set of real numbers and the functions $\tilde{c}^{(q)}(\cdot)$ and $\tilde{c}^{(q+1)}(\cdot)$ are continuously differentiable. Assume that the following conditions are satisfied.

1) The fixed-grid update is monotone for $0 \le q < Q$.
2) $\xi^{(q)}(\cdot)$ is convex on $\mathbb{R}^{N^{(q)}}$ for $0 < q < Q$.
3) The adjustment vector $r^{(q+1)}$ is given by (15) for $0 \le q < Q$.
4) $\nu_1^{(q)} + \nu_2^{(q)} \ge 1$ for $0 \le q < Q$.

Then, the multigrid algorithm of Fig. 3 is monotone for $c^{(0)}(\cdot)$.

The conditions 1, 3, and 4 of the Theorem are easily satisfied for most problems. However, the difficulty lies in satisfying condition 2, convexity of $\xi^{(q)}(\cdot)$ for $q > 0$. If the eigenvalues of the Hessian of $\xi^{(q)}(\cdot)$ are lower-bounded, the convexity condition can be satisfied by adding a convex term, such as $\gamma \|x^{(q)}\|^2$, to $\tilde{c}^{(q)}(\cdot)$ for $q > 0$, where γ is a sufficiently large constant. However, addition of such a term tends to slow convergence by making the coarse scale corrections too conservative.

When the forward model is given by a PDE, it can be difficult or impossible to verify or guarantee the convexity condition of 2. Nonetheless, the theorem still gives insight into the convergence behavior of the algorithm, and, in Section IV, we will show that empirically, for the difficult problem of ODT, the convergence of the multigrid algorithm is monotone in all cases, even without the addition of any convex terms.

E. Stabilizing Functionals

The coarse scale stabilizing functionals $S^{(q)}(x^{(q)})$ may be derived through appropriate scaling of $S(x)$. A general class of stabilizing functional has the form

$$S(x) = \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \rho\left(\frac{|x_i - x_j|}{\sigma} \right) \tag{22}$$

where 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 $\rho(\cdot)$ is a symmetric function that penalizes the differences in adjacent pixel values. Such a stabilizing functional results from the selection of a prior density $p(x)$ corresponding to a Markov random field (MRF) [74]. A wide variety of functionals $\rho(\cdot)$ have been suggested for this purpose [75]–[77]. Generally, these methods attempt to select these functionals so that large differences in pixel value are not excessively penalized, thereby allowing the accurate formation of sharp edge discontinuities.

The stabilizing functional at scale q must be selected so that

$$S^{(q)}(x^{(q)}) \cong S(x). \qquad (23)$$

This can be done by using a form similar to (22) and applying scaling factors to result in

$$S^{(q)}(x^{(q)}) = 2^{qd} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \rho \left(\frac{\left| x_i^{(q)} - x_j^{(q)} \right|}{2^q \sigma} \right) \qquad (24)$$

where d is the dimension of the problem. Here we assume that $x_i - x_j \cong \left(x_i^{(q)} - x_j^{(q)} \right)/2^q$, and we use the constant 2^{qd} to compensate for the reduction in the number of terms as the sampling grid is coarsened.

In our experiments, we use the generalized Gaussian Markov random field (GGMRF) image prior model [13], [14], [56], [67], [77] given by

$$p(x) = \frac{1}{\sigma^N z(p)} \exp \left[-\frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right] \qquad (25)$$

where σ is a normalization parameter, $1 \leq p \leq 2$ controls the degree of edge smoothness, and $z(p)$ is a partition function. For the GGMRF prior, the stabilizing functional is given by

$$S(x) = \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \qquad (26)$$

and the corresponding coarse scale stabilizing functionals are derived using (24) to be

$$S^{(q)}(x^{(q)}) = \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i^{(q)} - x_j^{(q)} \right|^P \qquad (27)$$

where $\sigma^{(q)}$ is given by

$$\sigma^{(q)} = 2^{q(1-(d/p))} \cdot \sigma^{(0)}. \qquad (28)$$

III. APPLICATION TO ODT

ODT is a method for determining spatial maps of optical absorption and scattering properties from measurements of light intensity transmitted through a highly scattering medium. In frequency domain ODT, the measured modulation envelope of the optical flux density is used to reconstruct the absorption coefficient and diffusion coefficient at each discretized grid point. However, for simplicity, we will only consider reconstruction of the absorption coefficient.

The complex amplitude $\phi_k(r)$ of the modulation envelope due to a point source at position s_k and angular frequency ω satisfies the frequency domain diffusion equation

$$\nabla \cdot [D(r)\nabla \phi_k(r)] + \left[-\mu_a(r) - \frac{j\omega}{c} \right] \phi_k(r) = -\delta(r - s_k) \qquad (29)$$

where r is position, c is the speed of light in the medium, $\mu_a(r)$ is the absorption coefficient, and $D(r)$ is the diffusion coefficient. The 3-D domain is discretized into N grid points, denoted by $r_1, r_2 \ldots, r_N$. The unknown image is then represented by an N-dimensional column vector $x = [\mu_a(r_1), \mu_a(r_2), \ldots, \mu_a(r_N)]^T$ containing the absorption coefficients at each discrete grid point, where T is the transpose operator. We will use the notation $\phi_k(r; x)$ in place of $\phi_k(r)$, in order to emphasize the dependence of the solution on the unknown image x. Then, the measurement of a detector at location d_m resulting from a source at location s_k can be modeled by the complex value $\phi_k(d_m; x)$. The complete forward model function is then given by[4]

$$f(x) = \Big[\phi_1(d_1; x), \phi_1(d_2; x), \ldots, \phi_1(d_M; x),$$
$$\phi_2(d_1; x), \ldots, \phi_K(d_M; x) \Big]^T. \qquad (30)$$

Note that $f(x)$ is a highly nonlinear function because it is given by the solution to a PDE using coefficients x. The measurement vector is also organized similarly as $y = [y_{11}, y_{12}, \ldots, y_{1m}, y_{21}, \ldots, y_{KM}]^T$, where y_{km} is the measurement with the source at s_k and the detector at d_m.

Our objective is to estimate the unknown image x from the measurements y. In a Bayesian framework, the MAP estimate of x is given by

$$\hat{x}_{MAP} = \arg\max_{x \geq 0} \{ \log p(y|x) + \log p(x) \} \qquad (31)$$

where $p(y|x)$ is the data likelihood and $p(x)$ is the prior model for image x, which is assumed to be strictly positive in value. We use an independent Gaussian shot noise model (see [67] for details of this noise model) with the form given in (2), where the weight matrix Λ is given by

$$\Lambda = \text{diag} \left(\frac{1}{|y_{11}|}, \ldots, \frac{1}{|y_{1M}|}, \frac{1}{|y_{21}|}, \ldots, \frac{1}{|y_{KM}|} \right). \qquad (32)$$

For the prior model, we use the GGMRF density of (25) for $p(x)$. Using the formulation of Section II-A, the ODT imaging problem is reduced to the optimization

$$(\hat{x}_{MAP}, \hat{\alpha}) = \arg\max_{x \geq 0} \max_{\alpha} \left\{ -\frac{1}{2\alpha} \| y - f(x) \|_\Lambda^2 \right.$$
$$\left. -\frac{P}{2} \log \alpha - \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p \right\} \qquad (33)$$

where constant terms are neglected. Minimizing (33) with respect to α reduces the cost functional to

$$c(x) = \frac{P}{2} \log \| y - f(x) \|_\Lambda^2 + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p. \qquad (34)$$

---

[4]For simplicity of notation, we assume that all source-detector pairs are used. However, in our experimental simulations, we use only a subset of all possible measurements. In fact, practical limitations can often limit the available measurements to a subset so that $P \neq 2KM$.

This cost functional has the same form as (4) with the stabilizing functional given by (26). The gradient terms of the stabilizing functional used in (17) and (18) are given by

$$\nabla S(x) = \frac{1}{\sigma^p} \sum_{j \in \mathcal{N}_n} b_{n-j} |x_n - x_j|^{p-1} \mathrm{sgn}(x_n - x_j). \quad (35)$$

We use multigrid inversion to solve the required optimization problem with coarse grid cost functionals of the form

$$c^{(q)}(x^{(q)}) = \frac{P}{2} \log \|y^{(q)} - f^{(q)}(x^{(q)})\|_\Lambda^2$$
$$+ \frac{1}{p(\sigma^{(q)})^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i^{(q)} - x_j^{(q)} \right|^p - r^{(q)} x^{(q)} \quad (36)$$

where $\sigma^{(q)}$ is given by (28) with $d = 3$.

At each scale q, we must also select a fixed-grid optimization algorithm. For simplicity, we minimize (36) by alternatively minimizing with respect to α and x using the update formulas

$$\alpha \leftarrow \frac{1}{P} \|y - f(x)\|_\Lambda^2 \quad (37)$$

$$x \leftarrow \approx \arg \min_{x \geq 0} \left\{ \frac{1}{2\alpha} \|y - f(x)\|_\Lambda^2 \right.$$
$$\left. + \frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} |x_i - x_j|^p - rx \right\} \quad (38)$$

where all expressions are interpreted as their corresponding scale q quantities. The fixed-scale optimization (38) is performed using ICD optimization, as described in [67]. ICD requires the evaluation of the Fréchet derivative matrix of (19). For the ODT problem, it can be shown that the Fréchet derivative is given by [78]

$$A_{(k-1)M+m,n} = \frac{\partial [f(x)]_{(k-1)M+m}}{\partial x_n}$$
$$= \frac{\partial \phi_k(d_m; x)}{\partial x_n}$$
$$= -G(s_k, r_n; x) G(d_m, r_n; x) V \quad (39)$$

where V is the voxel volume, $G(r_s, r_o; x)$ is the diffusion equation Green's function for the problem domain computed using the image x, with r_s as the source location and r_o as the observation point, and domain discretization errors are ignored [14], [78]. Since the ODT problem is inherently 3-D, the Fréchet derivative matrix is usually very large. Fortunately, the separable structure of the Fréchet derivative can be use to substantially reduce memory requirements by storing the two quantities

$$\phi = [G(s_1, r_1; x), \dots, G(s_1, r_N; x)$$
$$G(s_2, r_1; x), \dots, G(s_K, r_N; x)] \quad (40)$$
$$\psi = [G(d_1, r_1; x), \dots, G(d_1, r_N; x)$$

$$G(d_2, r_1; x), \dots, G(d_M, r_N; x)] \quad (41)$$

and computing A on the fly [14].

The ICD algorithm is initialized by setting a state vector $\hat{y}$ equal to the forward model output for the current value of x, giving

$$\hat{y} \leftarrow f(x). \quad (42)$$

Each ICD iteration is then computed by visiting each voxel n once using a random order, and updating each pixel value x_n and the state $\hat{y}$ using the following expressions:

$$x_{\mathrm{old,n}} \leftarrow x_n \quad (43)$$
$$x_n \leftarrow \arg \min_{u \geq 0} \left\{ \frac{1}{2\alpha} \|y - \hat{y} - A_{*n}(u - x_n)\|_\Lambda^2 \right.$$
$$\left. + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_n} b_{n-j} |u - x_j|^p - r_n u \right\} \quad (44)$$
$$\hat{y} \leftarrow \hat{y} + A_{*n}(x_n - x_{\mathrm{old,n}}) \quad (45)$$

where A_{*n} is the nth column of the matrix A. Note that the state $\hat{y}$ keeps a running estimate of the forward model output by (45), so that subsequent state updates can be computed efficiently.

Fig. 4 shows a detailed pseudocode specification for the fixed-grid and multigrid algorithms for the ODT application. In particular, it explicitly shows the computation of the quantities $\phi^{(q)}$ and $\psi^{(q)}$ used in the computation of the Fréchet derivative.

IV. NUMERICAL RESULTS

This section contains the results of numerical experiments using simulated data sets. In all cases, our simulated physical measurements were generated using a $257 \times 257 \times 257$ grid discretization of the domain and the MUDPACK [79] PDE solver. We used the highest practical resolution for the forward model simulation, so as to achieve the best possible accuracy of the simulated measurements. Since the sources and detectors are not located exactly on the grid points, a 3-D linear interpolation of the nearest grid points was also used.

Our experiments used two tissue phantoms, which we refer to as the homogeneous and nonhomogeneous phantoms. Both phantoms had dimensions of $10 \times 10 \times 10$ cm, and each face contained eight sources and nine detectors with a single modulation frequency of 100 MHz, as shown in Fig. 5. So, the number of sources was $K = 48$, and the number of detectors was $M = 54$. Some experiments used all source/detector pairs ($P = 2KM = 5184$), while others only used source/detector pairs on different faces of the cube ($P = 2K(M/6) \times 5 = 4320$). A zero-flux boundary condition on the outer boundary was imposed to approximate the physical boundary condition [14], [67], [78].

The homogeneous phantom had the constant values $\mu_a = 0.02$ cm^{-1} and $D = 0.03$ cm. For the inhomogeneous phantom of Fig. 6(a), the μ_a background was linearly varied from 0.01 cm^{-1} to 0.04 cm^{-1} in a direction perpendicular to a surface of the cubic phantom, except for the outermost region of

main( ) {

    Initialize $x^{(0)}$ with a background estimate

    For $q = 1, 2, \ldots, Q - 1, \quad x^{(q)} \leftarrow I_{(q-1)}^{(q)} x^{(q-1)}$

    For $q = 0, 1, \ldots, Q - 1, \quad r^{(q)} \leftarrow 0 \quad$ and $\quad y^{(q)} \leftarrow y$

    Repeat until converged: {

        Compute $\phi^{(0)}, \psi^{(0)}$ and $\hat{y} \leftarrow f^{(0)}(x^{(0)})$

        If Multigrid Inversion :

            Choose $\nu_1^{(0)}, \ldots, \nu_1^{(Q-1)}$ and $\nu_2^{(0)}, \ldots, \nu_2^{(Q-1)}$

            $x^{(0)} \leftarrow \text{MultigridV}(0, x^{(0)}, y^{(0)}, r^{(0)}, \phi^{(0)}, \psi^{(0)}, \hat{y})$

        If Fixed Grid Inversion :

            $x^{(0)} \leftarrow \text{Fixed\_Grid\_Update}(x^{(0)}, y^{(0)}, r^{(0)}, \phi^{(0)}, \psi^{(0)}, \hat{y})$

    }

}

(a)

$x \leftarrow \text{Fixed\_Grid\_Update}(x, y, r, \phi, \psi, \hat{y})$ {

    Compute $\alpha \leftarrow \frac{1}{P} \|y - \hat{y}\|_\Lambda^2$

    For $n = 0, \ldots, N - 1$ (in random order), {

        Compute column vector A_{*n} with (39)

        Update x_n, as described by Ye, *et al.*[67]:

$$x_{old,n} \leftarrow x_n$$

$$x_n \leftarrow \arg\min_{u \geq 0} \left\{ \frac{1}{2\alpha} \|y - \hat{y} - A_{*n}(u - x_n)\|_\Lambda^2 + \frac{1}{p\sigma^p} \sum_{j \in \mathcal{N}_n} b_{n-j} |u - x_j|^p - r_n u \right\}$$

$$\hat{y} \leftarrow \hat{y} + A_{*n}(x_n - x_{old,n})$$

    }

}

(b)

$x^{(q)} \leftarrow \text{MultigridV}(q, x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ {

    For $\nu = 1, \ldots, \nu_1^{(q)}$

        $x^{(q)} \leftarrow \text{Fixed\_Grid\_Update}(x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ //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 $\phi^{(q+1)}, \psi^{(q+1)}$ and $\hat{y} \leftarrow f^{(q+1)}(x^{(q+1)})$

    Compute $y^{(q+1)}$ using (13)

    Compute $r^{(q+1)}$ using (16)

    $x^{(q+1)} \leftarrow \text{MultigridV}(q + 1, x^{(q+1)}, y^{(q+1)}, r^{(q+1)}, \phi^{(q+1)}, \psi^{(q+1)}, \hat{y})$ //Coarse grid update

    $x^{(q)} \leftarrow x^{(q)} + I_{(q+1)}^{(q)}(x^{(q+1)} - I_{(q)}^{(q+1)} x^{(q)})$ //Coarse grid correction

    For $\nu = 1, \ldots, \nu_2^{(q)}$

        $x^{(q)} \leftarrow \text{Fixed\_Grid\_Update}(x^{(q)}, y^{(q)}, r^{(q)}, \phi^{(q)}, \psi^{(q)}, \hat{y})$ //Fine grid update

    Return $x^{(q)}$ //Return result

}

(c)

Fig. 4. Pseudocode specification of fixed-grid and multigrid inversion methods for the ODT problem showing (a) the main routine for ODT problems, (b) the fixed-grid update, and (c) the Multigrid-V inversion.

width 1.25 cm, which was homogeneous with $\mu_a = 0.02 \, \text{cm}^{-1}$. Two spherical μ_a inhomogeneities with values of $\mu_a = 0.1 \, \text{cm}^{-1}$ (left-top) and $\mu_a = 0.12 \, \text{cm}^{-1}$ (right-bottom) were centered on the bisecting plane, which is parallel to the cubic phantom surfaces parallel to the background variation direction. The diffusion coefficient D was homogeneous with $D = 0.03$ cm. For both phantoms, the reconstruction was performed for all voxels except the eight, four, and two outer-
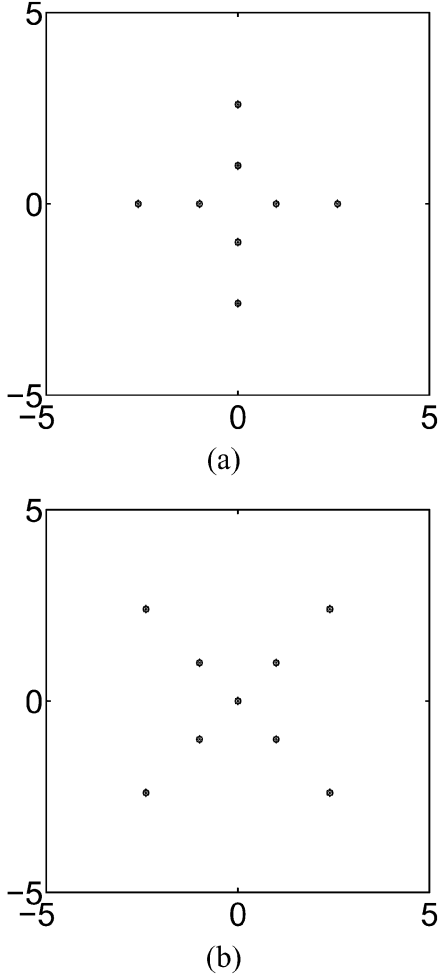
Fig. 5. (a) Source and (b) detector pattern on each face of the cube geometry. Two data set scenarios were considered: one containing all source/detector pairs and a second containing only source/detector pairs on different faces.

most layers of grid points for $65 \times 65 \times 65$, $33 \times 33 \times 33$, and $17 \times 17 \times 17$ reconstruction resolutions, respectively. These border regions were fixed to their true values in order to avoid singularities near the sources and detectors. These regions have also been excluded from all cross-section figures and the evaluation of root-mean-square (RMS) reconstruction error.

A. Evaluation of Required Forward Model Resolution

The objective of this section is to experimentally determine the forward model resolution required to produce a high-quality reconstruction. To do this, we first evaluated the accuracy of the forward model as a function of resolution using the homogeneous phantom. The forward model PDE was first solved as resolutions corresponding to $129 \times 129 \times 129$, $65 \times 65 \times 65$, $33 \times 33 \times 33$, and $17 \times 17 \times 17$ grid points. We then computed the distortion-to-noise ratio (DNR) for two scenarios. The first scenario included all source/detector pairs, and the second only included source/detector pairs on different faces. This was done because the close proximity of source/detector pairs on the same

TABLE I
DISTORTION-TO-NOISE (DNR) RATIO FOR VARIOUS FORWARD MODEL RESOLUTIONS. COARSE DISCRETIZATION INCREASED FORWARD MODEL ERROR, AND SOURCE/DETECTOR PAIRS ON THE SAME FACE HAD MUCH HIGHER DNR

Forward Model Resolution	Distortion-to-noise ratio	
	All measurements	Source/detector pairs on different faces
$17 \times 17 \times 17$	6.74×10^{-4}	9.96×10^{-7}
$33 \times 33 \times 33$	9.66×10^{-5}	2.85×10^{-8}
$65 \times 65 \times 65$	2.44×10^{-6}	3.35×10^{-9}
$129 \times 129 \times 129$	1.74×10^{-6}	1.04×10^{-10}

face can result in susceptibility to discretization errors in the forward model. The DNR for the forward solution with l grid points on each side was computed as

$$\text{DNR} = \frac{2}{P} \sum_{i=1}^{P/2} \frac{\left| y_{(257)i} - y_{(l)i} \right|^2}{\left| y_{(257)i} \right|} \tag{46}$$

where i is the index of source-detector pairs, $y_{(l)i}$ is the ith forward solution with l grid points on each side, $y_{(257)i}$ is the ith simulated measurement, which was computed wiht 257 grid points on each side, and $P/2$ is the number of complex measurements. Since $\left| y_{(257)i} \right|$ is proportional to the noise variance defined in (2) and (32), the DNR is proportional to the average ratio of discretization distortion and measurement noise.

Table I lists the DNR as a function of resolution for the two scenarios. Notice that for all resolutions the DNR is uniformly higher when source/detector pairs on the same face are included. As expected, the DNR also monotonically decreases as the resolution of the forward model is increased.

Next, we examined the reconstruction quality as a function of resolution using the inhomogeneous phantom. Gaussian shot noise was added to the data using Λ as given in (32) and [67], so that the average signal-to-noise ratio for sources and detectors on opposite faces was 35 dB. Fig. 6 shows a a cross section through the center of the inhomogeneities of the original phantom and the corresponding reconstructions for a variety of resolutions and data set scenarios.[5] Each reconstruction used $p = 1.2$, but the value of $\sigma^{(0)}$ was chosen from the range of 0.002 to 0.12, in order to minimize the RMS image error between the reconstructions and the decimation of the true phantom. The parameters and the resulting RMS errors are summarized in Table II.

Fig. 6 is consistent with the DNR measurement. The $65 \times 65 \times 65$ reconstruction from source/detector pairs on different faces has the best quality. Reconstructions at lower resolutions degrade rapidly, with very poor quality at $17 \times 17 \times 17$ resolution. Perhaps it is surprising that even the $65 \times 65 \times 65$

[5]These reconstructions were all produced using the multigrid algorithm with the mean phantom value as the initial condition because, in each case, this method converged to lowest cost among the tested algorithms.
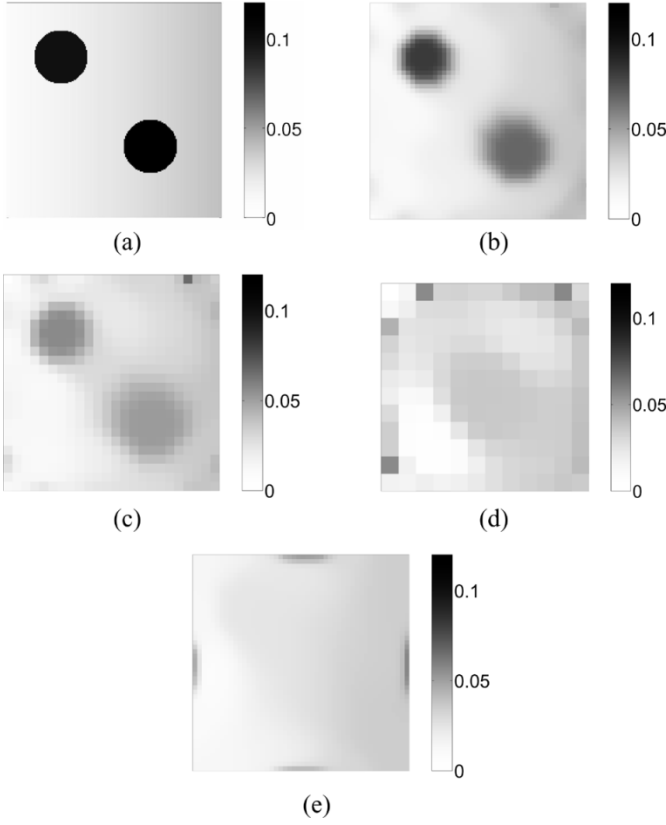
Fig. 6. Cross section through (a) the inhomogeneous phantom and the best reconstructions obtained using source detector pairs on different faces with (b) $65 \times 65 \times 65$ resolution, (c) $33 \times 33 \times 33$ resolution, (d) $17 \times 17 \times 17$ resolution, and (e) all source detector pairs with $65 \times 65 \times 65$ resolution. The $65 \times 65 \times 65$ reconstruction with different face source/detector pairs produced substantially better quality reconstruction. Reconstructions using all source/detector pairs failed even at $65 \times 65 \times 65$ reconstruction.

TABLE II
NORMALIZATION PARAMETER $\sigma^{(0)}$ THAT YIELDS THE BEST RECONSTRUCTION AND THE RESULTING RMS IMAGE ERROR BETWEEN THE RECONSTRUCTIONS AND THE DECIMATION OF THE TRUE PHANTOM

Resolution/Data Set	σ^0	RMS image error
$65 \times 65 \times 65$/diff. faces	0.018	0.0069
$33 \times 33 \times 33$/diff. faces	0.008	0.0079
$17 \times 17 \times 17$/diff. faces	0.004	0.0093
$65 \times 65 \times 65$/all	0.03	0.0099

resolution reconstruction fails when all source/detector pairs are used. This result emphasizes the importance of using sufficiently high resolution, particularly when source/detector pairs are closely spaced.

B. Multigrid Performance Evaluation

The performance of the fixed-grid and multigrid algorithms was evaluated using the measurements of the inhomogeneous phantom of Section IV-A. Based on the results of Section IV-A, all comparisons of fixed-grid and multigrid inversion algorithms were performed for the $65 \times 65 \times 65$ resolution using only source/detector pairs on different faces.

Our simulations compared fixed-grid inversion with multigrid inversion using 2, 3, and 4 levels of resolution. Table III lists these four cases together with our choice for the ν parameters at each scale. We selected the parameters ν to achieve robust convergence for a variety of problems. However, in other work [61], we have shown that these parameters can be adaptively chosen. The adaptive approach can further improve convergence speed and eliminates the need to select these parameters *a priori*. 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. To do this, we use the approximate theoretical number of multiplies and the corresponding relative complexity shown in Table III. However, we note that Table III indicates that the theoretical complexity of the multigrid iterations was somewhat lower then the experimentally measured complexity. See Appendix II for details of this conversion.

All reconstructions were done using the inhomogeneous phantom and a prior model with $p = 1.2$ and $\sigma^{(0)} = 0.018$ cm^{-1}. We chose $I_{(q)}^{(q+1)}$ to be the separable 3-D extensions of the one-dimensional (1-D) decimation matrix

$$
\begin{pmatrix}
\frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\
0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \dots & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \dots & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \frac{1}{4} & \frac{3}{4}
\end{pmatrix}
\tag{47}
$$

and $I_{(q+1)}^{(q)}$ to be the separable 3-D extension of the 1-D interpolation matrix

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} & 0 & \dots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \dots & 0 & \frac{1}{2} & \frac{1}{2} \\
0 & 0 & 0 & 0 & \dots & 0 & 0 & 1
\end{pmatrix}
\tag{48}
$$

respectively.

For the first experiment, all algorithms were initialized with the average values of the true phantom, which were $\mu_a = 0.026$ cm^{-1} and $D = 0.03$ cm. [6] Fig. 7 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.

Fig. 8 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 (-3.9833×10^4 to -3.9763×10^4) than the fixed-

---

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

TABLE III

COMPLEXITY COMPARISON FOR EACH ALGORITHM. THEORETICAL COMPLEX MULTIPLICATIONS ARE ESTIMATED WITH (55), AND THEORETICAL RELATIVE COMPLEXITY IS THE RATIO OF THE REQUIRED NUMBER OF MULTIPLICATIONS FOR ONE ITERATION TO THAT FOR ONE FIXED-GRID ITERATION. EXPERIMENTAL RELATIVE COMPLEXITY IS THE RATIO OF USER TIME REQUIRED FOR ONE ITERATION TO THAT FOR ONE FIXED-GRID ITERATION

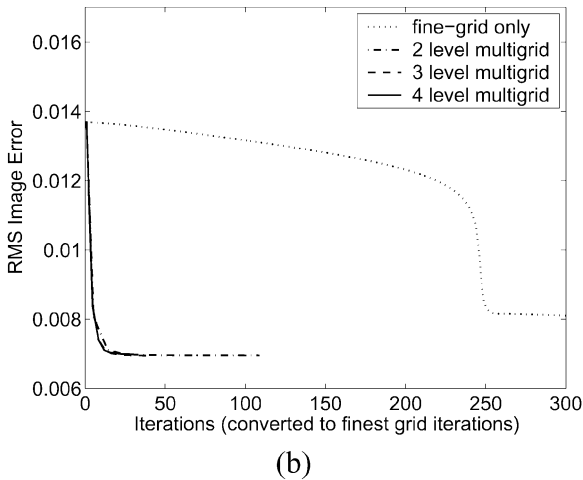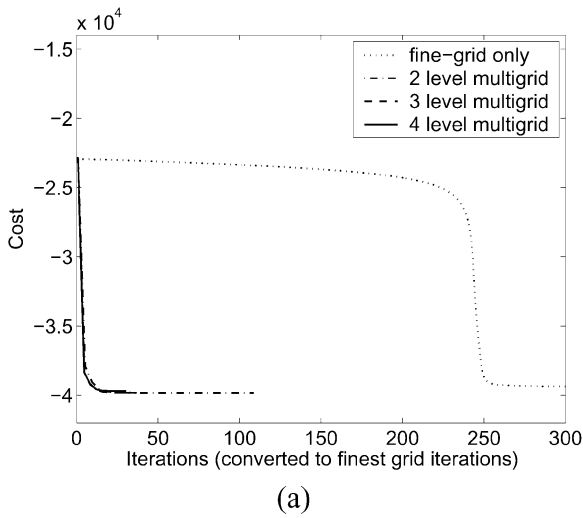| Algorithm		Parameters							Theoretical		Experimental
		$\nu_1^{(0)}$	$\nu_2^{(0)}$	$\nu_1^{(1)}$	$\nu_2^{(1)}$	$\nu_1^{(2)}$	$\nu_2^{(2)}$	$\nu_1^{(3)}$	Multiplications ($\times 10^6$)	Relative Complexity	Relative Complexity
Fixed-grid		1	·	·	·	·	·	·	5,799	1	1
Multigrid-V	2 levels	1	·	20	·	·	·	·	24,569	4.23	4.96
	3 levels	1	·	5	5	40	·	·	21,479	3.70	4.56
	4 levels	1	·	4	4	20	20	60	20,775	3.58	4.60

(a)

(b)

Fig. 7. Convergence of (a) cost function and (b) RMS image error when reconstructions were initialized with average values of true phantom. All multigrid algorithms converge about 13 times faster than the fixed-grid algorithm.
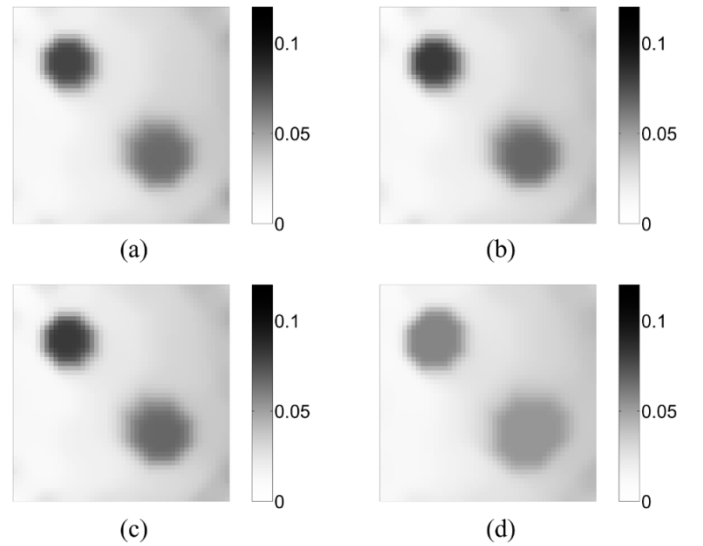

Fig. 8. Cross sections of reconstructions on the plane through the centers of the inhomogeneities using (a) a four-level multigrid with 19.35 iterations, (b) a three-level multigrid with 19.95 iterations, (c) a two-level multigrid with 18.24 iterations, and (d) 270 fixed-grid iterations. All the multigrid reconstructions have better image quality the the fixed-grid reconstruction.

grid algorithm (-3.9392×10^4), and the RMS image error for the multigrid reconstructions 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. 9 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.

V. CONCLUSION

We have proposed a nonlinear multigrid inversion algorithm which works by simultaneously varying the resolution of both the forward model and inverse computation. 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 problems such as those modeled by the solution of PDEs.
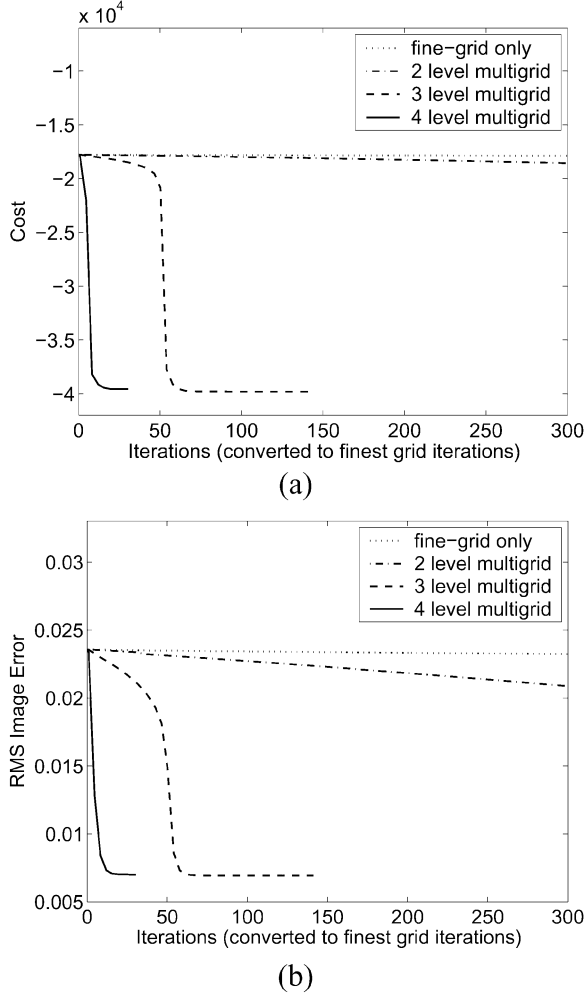
(a)

(b)

Fig. 9. Convergence of (a) cost function and (b) RMS image error with a poor initial guess. For higher level multigrid algorithms, the convergence was faster. In particular, the four-level multigrid algorithm converged almost as fast as when the reconstruction was initialized with the true phantom's average value.

We performed experimental simulations for the application of multigrid inversion to optical diffusion tomography using an ICD (Gauss–Seidel) fixed-grid optimizer. These simulations 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. Future investigation could also make these comparisons using other fixed-grid optimizers, such as conjugate gradient. Our experiments also indicated the importance of adequate resolution in the forward model.

APPENDIX I
PROOF OF MULTIGRID MONOTONE CONVERGENCE

We begin with two lemmas which give sufficient conditions to guarantee monotone decrease of the finer grid cost functional in the two-grid algorithm. All lemmas assume that the functions $\tilde{c}^{(q)}(\cdot)$ and $\tilde{c}^{(q+1)}(\cdot)$ are continuously differentiable.

Lemma 1: Assume that the following conditions are satisfied for a resolution $q \geq 0$:

1) The fixed-grid update is monotone at resolutions q and $q+1$.
2) A functional $\eta^{(q+1)} : \mathbb{R}^{N^{(q+1)}} \to \mathbb{R}$, defined by

$$\eta^{(q+1)}(x^{(q+1)}) = c^{(q+1)}(x^{(q+1)}) \\ - c^{(q)}\left(x^{(q)} + I_{(q+1)}^{(q)}\left(x^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}\right)\right) \quad (49)$$

has a global minimum at $x^{(q+1)} = I_{(q)}^{(q+1)}x^{(q)}$, where $x^{(q)}$ is the value resulting after the initial $\nu_1^{(q)}$ fixed-grid iterations.
3) $\nu_1^{(q)} + \nu_2^{(q)} \geq 1$.

Then, the two-grid inversion algorithm of Fig. 2 is monotone for the functional $c^{(q)}(\cdot)$.

Proof of Lemma 1: By the definition of monotonicity, the updated value $\tilde{x}^{(q+1)}$ of (9) satisfies

$$c^{(q+1)}(\tilde{x}^{(q+1)}) \leq c^{(q+1)}\left(I_{(q)}^{(q+1)}x^{(q)}\right). \quad (50)$$

Applying the definition of $\eta^{(q+1)}(\cdot)$ and the second condition, we have

$$\eta^{(q+1)}(\tilde{x}^{(q+1)}) \geq \eta^{(q+1)}\left(I_{(q)}^{(q+1)}x^{(q)}\right)$$

or equivalently

$$c^{(q+1)}(\tilde{x}^{(q+1)}) - c^{(q)}\left(x^{(q)} + I_{(q+1)}^{(q)}\left(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}\right)\right) \\ \geq c^{(q+1)}\left(I_{(q)}^{(q+1)}x^{(q)}\right) - c^{(q)}(x^{(q)}). \quad (51)$$

From the inequalities (50) and (51), it follows that

$$c^{(q)}\left(x^{(q)} + I_{(q+1)}^{(q)}\left(\tilde{x}^{(q+1)} - I_{(q)}^{(q+1)}x^{(q)}\right)\right) \leq c^{(q)}(x^{(q)}). \quad (52)$$

This inequality means that the coarse grid update and its subsequent coarse grid correction decreases the cost functional $c^{(q)}(\cdot)$. With the first condition, this guarantees the inequality in the definition of monotone convergence for $c^{(q)}(\cdot)$. Furthermore, by the first and fourth conditions, if $\nabla c^{(q)}(x^{(q)}) \neq 0$, the update at resolution q either before or after the coarse grid update strictly decreases $c^{(q)}(\cdot)$. Therefore, the two-grid algorithm is monotone under these assumptions.

Lemma 2. (Two-Grid Monotone Convergence): Assume that the following conditions are satisfied for a resolution $q \geq 0$:

1) The fixed-grid update is monotone at resolutions q and $q+1$.
2) $\xi^{(q+1)}(\cdot)$ is convex on $\mathbb{R}^{N^{(q+1)}}$.
3) The adjustment vector $r^{(q+1)}$ is given by (15).
4) $\nu_1^{(q)} + \nu_2^{(q)} \geq 1$.

Then, the two-grid inversion algorithm of Fig. 2 is monotone for the functional $c^{(q)}(\cdot)$.

Proof of Lemma 2: It is enough to show that the third and second conditions of this lemma imply the second condition in Lemma 1. By condition three, we know that

$$\eta^{(q+1)}(x^{(q+1)}) = \xi^{(q+1)}(x^{(q+1)}) + vx^{(q+1)} + \text{constant} \quad (53)$$

for some row vector v of length $N^{(q+1)}$. In fact, we know that (15) selects the vector v so that the gradients of the coarse and fine scale cost functionals are matched, and, therefore

$$\nabla\eta^{(q+1)}(x^{(q+1)})\Big|_{x^{(q+1)}=I_{(q)}^{(q+1)}x^{(q)}} = 0. \tag{54}$$

By (53), we also know that $\eta^{(q+1)}(\cdot)$ is a continuously differentiable convex function. Therefore, we know that $\eta^{(q+1)}(\cdot)$ must take on its global minimum value at $x^{(q+1)} = I_{(q)}^{(q+1)}x^{(q)}$.

Proof of Multigrid Monotone Convergence Theorem: Our proof is by induction. Consider the case when $q = Q - 2$, then we have the two-grid case, and the proof is trivial by Lemma 2.

Now consider $q < Q - 2$. By induction, we assume that the Multigrid-V algorithm applied at resolution $q + 1$ is monotone for the function $c^{(q+1)}(\cdot)$. This then meets condition 1 of Lemma 2, since the multigrid algorithm serves as the coarse grid optimizer in a two-grid algorithm. Therefore, the Multigrid-V algorithm applied at resolution q is monotone for the function $c^{(q)}(\cdot)$, and the induction is complete.

APPENDIX II
COMPUTATIONAL COMPLEXITY

In this appendix, we compare the computational cost of the proposed multigrid inversion algorithm for ODT problems with that of the fixed-grid ICD algorithm [67]. We use the number of complex multiplications required for one iteration of the V-cycle algorithm as a measure of computational complexity.

First, let us consider the computation required for one iteration of Fixed_Grid_Update(). Here, we use the analysis from [67]. Assuming F iterations are used for the linear PDE solver, the computation of Green's functions of (40) and (41) needs $5(K+M)FN_0$ multiplications, where N_0 is the number of grid points in the PDE domain. Then, we need PN and $(5/2)PN$ multiplications to compute (39) and (44), respectively, where N is the updated image size.[7] Thus, the total computational cost for one iteration of the ICD fixed-grid update is $5(K+M)FN_0 + (7/2)PN$ multiplications.

Now, let us estimate the computation required for one iteration of MultigridV() which operates at resolutions $0, \ldots, Q - 1$. For simplicity, we neglect the computational cost required for decimation and interpolation of images and the correction vector. In other words, we assume that the main computational cost at resolution q consists of the fixed-grid update on $x^{(q)}$ and the computation of $r^{(q)}$. To update $x^{(q)}$, one iteration of MultigridV() involves $\nu^{(q)} = \nu_1^{(q)} + \nu_2^{(q)}$ iterations of Fixed_Grid_Update(), which requires $\left[5(K+M)FN_0^{(q)} + (7/2)PN^{(q)}\right]\nu^{(q)}$ multiplications. Since $N_0^{(q)} = 8^{-q}N_0$ and $N^{(q)} = 8^{-q}N$ in 3-D problems, this is equal to $8^{-q} \times \left[5(K+M)FN_0 + (7/2)PN\right]\nu^{(q)}$ multiplications.

The correction vector $r^{(q)}$ is computed only once when the inversion proceeds from resolution q to $q + 1$. Since $g^{(q+1)}$ is computed in the optimization for the update of $x^{(q+1)}$, the only

additional computation for $r^{(q)}$ is computation of $g^{(q)}$ given by (17). To compute $g^{(q)}$, we first compute the Green's functions of (40) and (41) and then use them to compute Fréchet derivatives by (39), which requires $5(K+M)FN_0^{(q)}$ and $PN^{(q)}$ multiplications, respectively [67]. Then, $PN^{(q)}$ multiplications are required to evaluate the expression in the braces of (17). The resulting total complexity for computation of $r^{(q)}$ is $8^{-q} \times [5(K+M)FN_0 + 2PN]$ multiplications.

Thus, for resolutions $q = 0, \ldots, Q - 2$, the total complexity of the Multigrid-V algorithm is $8^{-q} \times [\{5(K+M)FN_0 + (7/2)PN\}\nu^{(q)} + \{5(K+M)FN_0 + 2PN\}]$ multiplications. At the coarsest resolution $q = Q - 1$, we do not need $r^{(Q-1)}$, so the complexity is $8^{-(Q-1)} \times \{5(K+M)FN_0 + (7/2)PN\}\nu^{(Q-1)}$ multiplications. Therefore, the total complexity for one Multigrid-V is

$$\sum_{q=0}^{Q-2}\left[8^{-q} \times \left\{\left\{5(K+M)FN_0 + \frac{7}{2}PN\right\}\nu^{(q)} \right.\right.$$
$$\left.\left. + \{5(K+M)FN_0 + 2PN\}\right\}\right]$$
$$+ 8^{-(Q-1)} \times \left\{5(K+M)FN_0 + \frac{7}{2}PN\right\}\nu^{(Q-1)} \tag{55}$$

where K is the number of sources, M is the number of detectors, P is the number of measurements, N_0 is the PDE image size, N is the updated image size, F is the number of iterations required for the linear forward solver, and $\nu^{(q)}$ is the number of iterations of fixed-grid update at resolution q.

Table III lists the estimated number of complex multiplications required for each iteration of the fixed-grid and Multigrid-V algorithms for typical values of parameters which we use in the simulations of Section IV-B. The values of the parameters are $K = 48$, $M = 54$, $P = 4320$, $N_0 = 65 \times 65 \times 65$, $N = 49 \times 49 \times 49$, and $F = 16$. We also provide the experimental computation time. One fixed-grid iteration took 55.5 minutes of user time on a Pentium-III 697 MHz Linux machine, and the complexity per iteration is 4.56–4.96 times larger for the multigrid algorithm. However, one multigrid iteration involves many coarser grid iterations, and the simulation results show that the number of iterations required for the multigrid algorithms to converge is substantially less than is required using the fixed-grid algorithm.

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 Process. Mag.*, vol. 18, no. 6, pp. 57–75, Nov. 2001.

[2] S. R. Arridge, "Optical tomography in medical imaging," *Inv. Probl.*, vol. 15, pp. R41–R93, 1999.

[3] J. C. Hebden, S. R. Arridge, and D. T. Delpy, "Optical imaging in medicine: I. Experimental techniques," *Phys. Med. Biol.*, vol. 42, pp. 825–840, 1997.

[4] S. R. Arridge and J. C. Hebden, "Optical imaging in medicine: II. Modeling and reconstruction," *Phys. Med. Biol.*, vol. 42, pp. 825–840, 1997.

[5] G. J. Saulnier, R. S. Blue, J. C. Newell, D. Isaacson, and P. M. Edic, "Electrical impedance tomography," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 31–43, Nov. 2001.

[6] J. S. Reynolds, A. Przadka, S. Yeung, and K. J. Webb, "Optical diffusion imaging: A comparative numerical and experimental study," *Appl. Opt.*, vol. 35, no. 19, pp. 3671–3679, Jul. 1996.

---

[7]In Section IV, we do not update the outermost region to avoid singularity problems, so N and N_0 are different in this case.

[7] E. C. Fear, S. C. Hagness, P. M. Meaney, M. Okoniewski, and M. A. Stuchly, "Enhancing breast tumor detection with near field imaging," *IEEE Microwave Mag.*, no. 3, pp. 48–56, Mar. 2002.

[8] R. L. Thomas, "Reflections of a thermal wave imager: Tow decades of research in photoacoustics and photothermal phenomena," *Anal. Sci.*, vol. 17, pp. S1–S4, Apr. 2001.

[9] R. Pike and P. Sabatier, *Scattering: Scattering and Inverse Scattering in Pure and Applied Science*. San Diego, CA: Academic, 2002.

[10] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.

[11] M. V. Ranganath, A. P. Dhawan, and N. Mullani, "A multigrid expectation maximization reconstruction algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 7, no. 4, pp. 273–278, Apr. 1988.

[12] T. Pan and A. E. Yagle, "Numerical study of multigrid implementations of some iterative image reconstruction algorithms," *IEEE Trans. Med. Imag.*, vol. 10, no. 4, pp. 572–588, Apr. 1991.

[13] A. B. Milstein, S. Oh, J. S. Reynolds, K. J. Webb, C. A. Bouman, and R. P. Millane, "Three-dimensional Bayesian optical diffusion tomography using experimental data," *Opt. Lett.*, vol. 27, pp. 95–97, Jan. 2002.

[14] S. Oh, A. B. Milstein, R. P. Millane, C. A. Bouman, and K. J. Webb, "Source-detector calibration in three-dimensional Bayesian optical diffusion tomography," *J. Opt. Soc. Amer. A*, vol. 19, no. 10, pp. 1983–1993, Oct. 2002.

[15] A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, and R. P. Millane, "Fluorescence optical diffusion tomography," *Appl. Opt.*, vol. 42, pp. 3081–3094, Jun. 2003.

[16] B. Sahiner and A. Yagle, "Image reconstruction from projections under wavelet constraints," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3579–3584, Dec. 1993.

[17] M. Bhatia, W. C. Karl, and A. S. Willsky, "Wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imag.*, vol. 15, no. 1, pp. 92–101, Jan. 1996.

[18] ——, "Tomographic reconstruction and estimation based on multiscale natural -pixel bases," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 463–478, Mar. 1997.

[19] N. Lee, "Wavelet-Vaguelette decompositions and homogenous equations," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 1998.

[20] A. Delaney and Y. Bresler, "Multiresolution tomographic reconstruction using wavelets," *IEEE Trans. Image Process.*, vol. 4, no. 6, pp. 799–813, June 1995.

[21] Z. Wu, G. T. Herman, and J. A. Browne, "Edge preserving reconstruction using adaptive smoothing in wavelet domain," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 3, San Francisco, CA, Oct. 31–Nov. 6 31–6, 1993, pp. 1917–1921.

[22] S. S. Saquib, C. A. Bouman, and K. Sauer, "A nonhomogeneous MRF model for multiresolution Bayesian estimation," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Lausanne, Switzerland, Sep. 16–19, 1996, pp. 445–448.

[23] R. Nowak and E. D. Kolaczyk, "A multiscale MAP estimation method for Poisson inverse problems," in *Proc. 32nd Asilomar Conf. Signals, Systems, Computers*, vol. 2, Pacific Grove, CA, Nov. 1–4, 1998, pp. 1682–1686.

[24] T. Frese, C. A. Bouman, and K. Sauer, "Adaptive wavelet graph model for Bayesian tomographic reconstruction," *IEEE Trans. Image Process.*, vol. 11, no. 7, pp. 756–770, Jul. 2002.

[25] E. L. Miller, L. Nicolaides, and A. Mandelis, "Nonlinear inverse scattering methods for thermal wave slice tomography: A wavelet domain approach," *J. Opt. Soc. Amer. A*, vol. 15, no. 6, pp. 1545–1556, Jun. 1998.

[26] W. Zhu, Y. Wang, Y. Deng, Y. Yao, and R. Barbour, "A wavelet-based multiresolution regularization least squares reconstruction approach for optical tomography," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 210–217, Feb. 1997.

[27] A. Brandt, "Multi-level adaptive solutions to boundary value problems," *Math. Comput.*, vol. 31, no. 138, pp. 333–390, Apr. 1977.

[28] U. Trottenberg, C. W. Oosterlee, and A. Schueller, *Multigrid*. London, U.K.: Academic, 2000.

[29] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd ed. Philadelphia, PA: SIAM, 2000.

[30] S. McCormick, Ed., *Multigrid Methods*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1987.

[31] W. Hackbusch, *Multigrid Methods and Applications*, ser. Springer. Berlin, Germany: Springer-Verlag, 1985.

[32] P. Wesseling, *An Introduction to Multigrid Methods*. Chichester, U.K.: Wiley, 1992.

[33] D. Terzopoulos, "Image analysis using multigrid relaxation methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 3, pp. 129–139, Mar. 1986.

[34] R. Kimmel and I. Yavneh, "An algebraic multigrid approach for image analysis," *SIAM J. Sci. Comput.*, vol. 24, no. 4, pp. 1218–1231, 2003.

[35] E. Enkelmann, "Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences," *Comput. Vis. Graph. Image Process.*, vol. 43, pp. 150–177, 1988.

[36] E. Memin and P. Perez, "Dense estimation and object-based segmentation of the optical flow with robust techniques," *IEEE Trans. Image Process.*, vol. 7, no. 5, pp. 703–719, May 1998.

[37] P. Hellier, C. Barillot, E. Mémin, and P. Pérez, "Hierarchical estimation of a dense deformation field for 3-d robust registration," *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 388–402, May 2001.

[38] S. Ghosal and P. Vanek, "Fast Algebraic Multigrid for discontinuous optical flow estimation," Center Comput. Math., Univ. Colorado, Denver, Tech. Rep. UCD-CCM-025, 1994.

[39] P. Saint-Marc, J. Chen, and G. Medioni, "Adaptive smoothing: A general tool for early vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 6, pp. 514–529, Jun. 1991.

[40] M. Unser, "Multigrid adaptive image processing," in *Proc. IEEE Int. Conf. Image Processing*, vol. I, Washington, DC, Oct. 1995, pp. 49–52.

[41] D. L. Pham and J. L. Prince, "Adaptive fuzzy segmentation of magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 737–752, Sep. 1999.

[42] S. Henn and K. Witsch, "A multigrid approach for minimizing a nonlinear functional for digitial image matching," *Computing*, vol. 64, pp. 339–348, 2000.

[43] K. Zhou and C. K. Rushforth, "Image restoration using multigrid methods," *Appl. Opt.*, vol. 30, no. 20, pp. 2906–2912, July 1991.

[44] S. T. Acton, "Multigrid anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 280–291, Mar. 1998.

[45] D. Terzopoulos, "The computation of visible-surface representations," *Proc. IEEE Int. Conf. Image Processing*, vol. 10, no. 4, pp. 417–438, Jul. 1988.

[46] M. Arigovindan, M. Sühling, P. Hunziker, and M. Unser, "Multigrid image reconstruction from arbitrarily spaced samples," in *Proc. IEEE Int. Conf. Image Processing*, vol. III, Rochester, NY, Sep. 22–25, 2002, pp. 381–384.

[47] C. A. Bouman and K. Sauer, "Nonlinear multigrid methods of optimization in Bayesian tomographic image reconstruction," in *Proc. SPIE Conf. Neural and Stochastic Methods in Image and Signal Processing*, vol. 1766, San Diego, CA, Jul. 19–24, 1992, pp. 296–306.

[48] S. F. McCormick and J. G. Wade, "Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography," *Inverse Problems*, vol. 9, pp. 697–713, 1993.

[49] L. Borcea, "Nonlinear multigrid for imaging electrical conductivity and permittivity at low frequency," *Inverse Problems*, vol. 17, pp. 329–359, Apr. 2001.

[50] R. Gandlin and A. Brandt, "Two multigrid algorithms for inverse problem in electrical impedance tomography," in *Proc. Copper Mountain Conf. Multigrid Methods*, Copper Mountain, CO, Mar. 30–Apr. 4 2003.

[51] A. Brandt and R. Gandlin, "Multigrid for atmospheric data assimilation: Analysis," in *Proc. Hyperbolic Problems: Theory, Numerics, and Applications*, Pasadena, CA, Mar. 2002, pp. 369–376.

[52] A. Brandt, "Multiscale and multiresolution methods: Theory and applications," in *Multiscale Scientific Computation: Review 2001*, T. J. Barth, T. F. Chan, and R. Haimes, Eds. Heidelberg, Germany: Springer Verlag, 2001, pp. 3–96.

[53] A. Brandt and D. Ron, "Multigrid solvers and multilevel optimization strategies," in *Multilevel Optimization and VLSICAD*, J. Cong and J. R. Shinnerl, Eds. Boston, MA: Kluwer, 2002, pp. 1–69.

[54] C. R. Johnson, M. Mohr, U. Ruede, A. Samsonov, and K. Zyp, "Multilevel methods for inverse bioelelectric field problems," in *Lecture Notes in Computational Science and Engineering—Multiscale and Multiresolution Methods: Theory and Applications*, T. J. Barth, T. F. Chan, and R. Haimes, Eds. Heidelberg, Germany: Springer-Verlag, 2001, vol. 20.

[55] J. C. Ye, C. A. Bouman, R. P. Millane, and K. J. Webb, "Nonlinear multigrid optimization for Bayesian diffusion tomography," in *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, Oct. 25–28, 1999.

[56] J. C. Ye, C. A. Bouman, K. J. Webb, and R. P. Millane, "Nonlinear multigrid algorithms for Bayesian optical diffusion tomography," *IEEE Trans. Image Process.*, vol. 10, no. 6, pp. 909–922, Jun. 2001.

[57] S. G. Nash, "A multigrid approach to discretized optimization problems," *J. Optimization Methods Software*, vol. 14, pp. 99–116, 2000.

[58] R. M. Lewis and S. G. Nash, "A multigrid approach to the optimization of systems governed by differential equations," in *Proc. 8th AIAA/USAF/ISSMO Symp. Multidisciplinary Analysis and Optimization*, Long Beach, CA, 2000.

[59] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb, "Multigrid inversion algorithms with applications to optical diffusion tomography," in *Proc. 36th Asilomar Conf. Signals, Systems, and Computers*, Monterey, CA, Nov. 2002, pp. 901–905.

[60] ——, "Multigrid algorithms for optimizations and inverse problems," in *Proc. SPIE/IS&T Conf. Computational Imaging 2003*, Santa Clara, CA, Jan. 20–25, 2003, pp. 59–70.

[61] ——, "Adaptive nonlinear multigrid inversion with applications to Bayesian optical diffusion tomography," in *Proc. IEEE Workshop Statistical Signal Processing*, St. Louis, MO, Sep. 2003.

[62] ——, "Nonlinear multigrid inversion," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003.

[63] S. S. Saquib, K. M. Hanson, and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data," in *Proc. SPIE Conf. Medical Imaging: Image Processing*, vol. 3034, Newport Beach, CA, Feb. 25–28, 1997, pp. 369–380.

[64] A. H. Hielscher, A. D. Klose, and K. M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 3, pp. 262–271, Mar. 1999.

[65] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems," in *Proc. IEEE Int. Conf. Image Processing*, vol. II, Lausanne, Switzerland, Sep. 16–19, 1996, pp. 473–476.

[66] ——, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Minneapolis, MN, Apr. 27–30, 1993, pp. 495–498.

[67] 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. Amer. A*, vol. 16, no. 10, pp. 2400–2412, Oct. 1999.

[68] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554–1563, 1966.

[69] 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. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.

[70] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.*, vol. 14, no. 3, pp. 132–137, Mar. 1995.

[71] J. Browne and A. R. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 687–699, May 1996.

[72] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 166–175, Feb. 1997.

[73] J. Zheng, S. S. Saquib, K. Sauer, and C. A. Bouman, "Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1745–1759, Oct. 2000.

[74] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Stat. Soc. B*, vol. 36, no. 2, pp. 192–236, 1974.

[75] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 194–202, Feb. 1989.

[76] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[77] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Mar. 1993.

[78] J. C. Ye, K. J. Webb, R. P. Millane, and T. J. Downar, "Modified distorted Born iterative method with an approximate Fréchet derivative for optical diffusion tomography," *J. Opt. Soc. Amer. A*, vol. 16, no. 7, pp. 1814–1826, July 1999.

[79] J. C. Adams, "MUDPACK: Multigrid portable FORTRAN software for the efficient solution of linear elliptic partial differential equations," *Appl. Math. Comput.*, vol. 34, pp. 113–146, 1989.
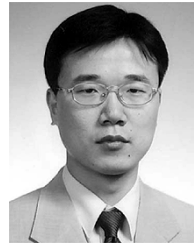
Seungseok Oh (S'00) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1997 and 1999, respectively. He is currently pursuing the Ph.D. degree at the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.

From 1999 to 2000, he was with Hanaro Telocom, Inc., as a Network Engineer. His current research interests include image processing, inverse problems, medical imaging, multimedia systems, and biomedical optics.

Adam B. Milstein (S'01) received the B.S.E.E. degree from Purdue University, West Lafayette, IN, in 2000. He is currently pursuing the Ph.D. degree in electrical engineering at Purdue University.

From 1997 to 2000, he was with NASA's Johnson Space Center through the Cooperative Education Program, doing several work terms in the Avionics and Robotics Divisions. Since 2000, he has been a Research Assistant at the School of Electrical and Computer Engineering, Purdue University.

Dr. Milstein is a recipient of the Steven C. Beering Fellowship and the IS&T's Raymond Davis Scholarship.

Charles A. Bouman (S'86–M'89–SM'97–F'01) received the B.S.E.E. degree from the University of Pennsylvania, Philadephia, in 1981, the M.S. degree in electrical engineering from the University of California, Berkeley, in 1982, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, under the support of an IBM graduate fellowship, in 1987 and 1989, respectively.

From 1982 to 1985, he was a staff member in the Analog Device Technology Group, Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge. In 1989, he joined the faculty of Purdue University, West Lafayette, IN, where he holds the rank of Professor with a primary appointment in the School of Electrical and Computer Engineering and a secondary appointment in the Department of Biomedical Engineering. His research focuses on the use of statistical image models, multiscale techniques, and fast algorithms in applications, including multiscale image segmentation, tomographic image reconstruction, image printing and rendering, and document segmentation and compression.

Prof. Bouman is a Fellow of the American Institute for Medical and Biomedical Engineering (AIMBE), a member of the SPIE and IS&T professional societies, a recipient of the IS&T's Raymond C. Bowman Award for outstanding contributions to digital imaging education and research, and a University Faculty Scholar of Purdue University. He has been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON PATTERN ANALYSIS and MACHINE INTELLIGENCE. He has also been the Awards Chair for the ICIP 1998 organizing committee, Co-Chair of the SPIE/IS&T conferences on Visual Communications and Image Processing 2000 (VCIP), and a member of the IEEE Image and Multidimensional Signal Processing Technical Committee. Currently, he is the Vice President of Publications and a member of the Board of Directors for the IS&T Society, and he is the Founder and Co-Chair of the SPIE/IS&T Conference on Computational Imaging.

Kevin J. Webb (S'81–M'84–SM'98–F'04) received the B.Eng. and M.Eng. degrees from the Royal Melbourne Institute of Technology, Australia, in 1978 and 1983, respectively, the M.S.E.E. degree rom the University of California, Santa Barbara, in 1981, and the Ph.D. degree from the University of Illinois, Urbana, in 1984.

He is a Professor at the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.