A MULTIGRID METHOD FOR DISTRIBUTED PARAMETER ESTIMATION PROBLEMS

U. M. ASCHER* AND E. HABER†

Abstract.
This paper considers problems of distributed parameter estimation from data measurements on solutions of partial differential equations (PDEs). A nonlinear least squares functional is minimized to approximately recover the sought parameter function (i.e., the model). This functional consists of a data fitting term, involving the solution of a finite volume or finite element discretization of the forward differential equation, and a Tikhonov-type regularization term, involving the discretization of a mix of model derivatives.

We develop a multigrid method for the resulting constrained optimization problem. The method directly addresses the discretized PDE system which defines a critical point of the Lagrangian. The discretization is cell-based. This system is strongly coupled when the regularization parameter is small. Moreover, the compactness of the discretization scheme does not necessarily follow from compact discretizations of the forward model and of the regularization term. We therefore employ a Marquardt-type modification on coarser grids. Alternatively, fewer grids are used and a preconditioned Krylov-space method is utilized on the coarsest grid. A collective point relaxation method (weighted Jacobi or a Gauss-Seidel variant) is used for smoothing. We demonstrate the efficiency of our method on a classical model problem from hydrology.

1. Introduction. This paper proposes an efficient multigrid method for the recovery of a coefficient function of an elliptic differential equation in 3D. Such problems arise in many applications, including DC resistivity [37], magnetotelluric inversion [34], diffraction tomography [15], impedance tomography [10], oil reservoir simulation [16] and aquifer calibration [20].

We write the forward problem as

$$\mathcal{A}(\mathrm{m})\,\mathrm{u} \;=\; \mathrm{q}, \tag{1.1}$$

where $\mathcal{A}$ refers to the differential operator defined on an appropriate domain $\Omega \subset \mathbb{R}^3$ and equipped with suitable boundary conditions. This operator depends on a model, $\mathrm{m}(\mathbf{x})$, which is to be approximately recovered based on measurement data b on the solution $\mathrm{u}(\mathbf{x})$ of (1.1) or its gradient $\nabla \mathrm{u}(\mathbf{x})$. We assume for simplicity that the forward problem for u is linear, as is often the case in practice.

An instance of (1.1) on which we have conducted our present experiments is given by

$$\nabla \cdot (e^{\mathrm{m}} \nabla \mathrm{u}) = \mathrm{q}, \qquad \mathbf{x} \in \Omega \tag{1.2a}$$

$$\nabla \mathrm{u} \cdot \mathbf{n} = 0, \qquad \mathbf{x} \in \partial\Omega. \tag{1.2b}$$

Such a forward problem arises in particular in DC resistivity, electrical impedance tomography (EIT) and hydrology.

The inverse problem is well-known to be ill-posed, and we follow the established Tikhonov-like regularization approach [39, 17]. Thus, we minimize the sum of a least squares data fitting term and a regularization functional which is a discretization of $\beta \|\mathrm{a}^{1/2} \nabla (\mathrm{m} - \mathrm{m}_{ref})\|^2$, where $\beta \geq 0$ is the regularization parameter, and $\mathrm{m}_{ref}(\mathbf{x})$ and $\mathrm{a}(\mathbf{x})$ are given functions representing some a priori information. In this paper we assume that $\mathrm{a}(\mathbf{x})$ is a diagonal, positive definite 3×3 matrix function. The norm is assumed by default to be that of $L_2(\Omega)$ or a corresponding discretization. (An extension to the Total Variation and Huber norms [41, 27, 18] will not be discussed here.)

In practice, both the forward problem (1.1) and the regularization functional must be discretized, and we assume for simplicity that the model m and the field u are discretized on the same grid, using a stable, accurate finite volume or finite element method, as developed for (1.2) in §2. This

*Department of Computer Science, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. (ascher@cs.ubc.ca). Supported in part under NSERC Research Grant 84306.

†Departments of Computer Science and of Earth and Oceanography Sciences, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada. (haber@cs.ubc.ca). Supported in part under NSERC CRD Grant 80357.

then leads, for a given (*finest*) grid, to a constrained minimization problem of the form

$$\begin{aligned} \text{minimize} \quad & \phi(u,m) \equiv \frac{1}{2}\|Qu-b\|^2 + \frac{\beta}{2}\|W(m-m_{ref})\|^2 \\ \text{subject to} \quad & A(m)u - q = 0, \end{aligned} \tag{1.3}$$

where u and m are grid functions ordered as vectors, A is the corresponding discretization of $\mathcal{A}$ assumed to be a nonsingular matrix[1], and W is likewise the discretization matrix of the weighted gradient operator $\mathrm{a}^{1/2}\nabla$. The vector m_{ref} corresponds to a discretized m_{ref}, and the matrix Q indicates the projection of u or of its gradient onto the locations in the grid to which the data b are associated. In the simplest case, which we assume here, Q arises from a trilinear interpolation from the finest grid to the measurement locations. The matrices A, W and Q are therefore all very large and sparse in a typical application.

REMARK 1.1. *There are two routes which may make the formulation easier to solve numerically or to understand mathematically, but which we specifically avoid, because the overall method becomes less practically useful:*

- *We* do not *pre-process and interpolate the data to all grid points, as this could introduce correlated (non-random) noise which cannot be subsequently removed. Thus, we deal with a singular Q^TQ, in general.*
- *We* do not *set $\beta = 0$, although using a sufficiently coarse grid for m introduces a regularizing effect. The essential problem with a grid regularization is that it may produce arbitrary results, as it ignores a priori information. A model which permits redundant detail is often better for practical problems [43, 42, 24]. In our experience, once the grid for m is fine enough to reconstruct a model in reasonable detail as a piecewise constant function, the regularizing effect achieved by varying the grid is neither sufficiently significant nor sufficiently smooth to be effectively controllable [4].*

□

Let us concentrate on the efficient solution of the constrained optimization problem (1.3). Typically in the literature, the constraints are eliminated and the resulting unconstrained formulation is solved by some variant of Newton's method, usually the Gauss-Newton method. A preconditioned conjugate gradient algorithm is applied at each iteration for the resulting reduced Hessian system.

On the other hand, computational approaches in optimal control usually address the constrained formulation corresponding to (1.3) directly. The reason for this traditional dichotomy is simply that in many optimal control problems there are additional inequality constraints, which makes the "clean" elimination of constraints impossible to carry out. In the latter case various nonlinear programming methods for constrained optimization have been employed, especially SQP; see, e.g., [38, 5, 35, 6, 33]. These methods seek a critical point for the Lagrangian

$$\mathcal{L} = \mathcal{L}(u,m,\lambda) = \frac{1}{2}\|Qu-b\|^2 + \frac{\beta}{2}\|W(m-m_{ref})\|^2 + \lambda^T\left(A(m)u - q\right), \tag{1.4}$$

i.e., a solution to the large system of nonlinear equations

$$\mathcal{L}_\lambda = Au - q = 0, \tag{1.5a}$$

$$\mathcal{L}_u = Q^T(Qu-b) + A^T\lambda = 0, \tag{1.5b}$$

$$\mathcal{L}_m = \beta W^TW(m-m_{ref}) + G^T\lambda = 0, \tag{1.5c}$$

where

$$G = G(u,m) = \frac{\partial(A(m)u)}{\partial m}.$$

[1] The solution of (1.2) is actually determined only up to a constant. To ensure that A is indeed nonsingular we may use some known value of u at a small part of $\partial\Omega$ to pin it down.

In such methods the elimination of u and of corresponding Lagrange multipliers λ, or of their respective update directions, is possible within each linearizing iteration (see [23]).

The problem (1.3) is nonlinear, hence an iterative method is employed for its solution. Within each such *outer*, nonlinear iteration for the optimization problem, iterative methods are employed whenever an inversion involving A, A^T or W^TW is to be executed. It seems wasteful if these *inner* iterations are to be applied to eliminate variables to a much higher accuracy (stricter tolerance) than the accuracy of the current iterate within the outer loop. Likewise, it is wasteful to eliminate some variables in terms of others to a high accuracy at each outer iteration. Yet, precisely such an imbalance of accuracies is implied by the constraint elimination approach (or the reduced Hessian method) [22, 19]. [2] In [22] we therefore advocated an *all-at-once* approach [36, 25]. Specifically, at each Newton-type iteration for (1.5) a variant of symmetric QMR is employed, and an effective preconditioner is obtained by solving the reduced Hessian system approximately. In addition, inexact Newton-type methods are utilized. See also [7, 8].

Indeed, a computational method for the optimization problem (1.3) should not be inferior to a method which simply solves the system (1.5) directly by viewing the latter as a discretization of a nonlinear PDE system: The additional information available in the optimization formulation should help in solving more difficult problems robustly, rather than slowing down the basic solution process. Thus we are led to consider in this paper a multigrid method for solving the discretized PDE system (1.5) in linearized form, without prior elimination; the outer iteration is handled using optimization techniques [33, 23].

Note that A, A^T and W^TW are all discretizations of elliptic differential operators which dominate their respective equations in (1.5) if β is large. In such a case devising a multigrid method (as well as other methods) is straightforward. But the actual value of β is determined for best recovery in the presence of noise in the data, and the corresponding values are usually small (unless the noise level is very high). For realistic values of β the term βW^TW does not dominate (1.5c) and hence the system (1.5) is strongly coupled.

A Newton linearization for solving the nonlinear equations (1.5) leads to the following permuted KKT system to be solved at each iteration:

$$H \begin{pmatrix} \delta u \\ \delta\lambda \\ \delta m \end{pmatrix} = - \begin{pmatrix} \mathcal{L}_\lambda \\ \mathcal{L}_u \\ \mathcal{L}_m \end{pmatrix}, \quad \text{where } H = \begin{pmatrix} A & 0 & G \\ Q^TQ & A^T & K \\ K^T & G^T & \beta W^TW + T \end{pmatrix}, \tag{1.6}$$

with

$$K = K(m,\lambda) = \frac{\partial(A^T\lambda)}{\partial m}, \quad T = T(u,m,\lambda) = \frac{\partial(G^T\lambda)}{\partial m}.$$

We use an outer iteration which constructs damped Newton-like iterates and controls convergence with a merit function which takes into account both progress towards optimality and reduction of infeasibilities. Furthermore, we construct finite volume discretizations which retain a good h-ellipticity measure, even for very small values of β.

In §2 we investigate the properties of the linearized, discretized PDE system, and we derive appropriate discretizations. Whereas the PDE system is elliptic, and even though A and W^TW correspond to discretizations with good h-ellipticity measures [12, 40], in case of a cell-centered, non-staggered discretization and when β is small the corresponding discrete h-ellipticity measure deteriorates for the discretized system of (1.6) as the grids get coarse. Thus, within the multigrid cycle, if the value of β is deemed too small for a given grid then it is increased in H but not in the right hand side of (1.6). This corresponds to a Levenberg-Marquardt approach [33].

[2] Note that this issue really arises only for problems in more than 1D, i.e., unlike those considered in [6, 35, 38, 5].

The components of our multigrid algorithm are described in §3. We use collective point relaxation. It is natural to embed this algorithm also in a nested iteration (aka FMG) [11, 40]. In the present context there is also an additional question of determining the regularization parameter. Unlike [2, 9] we have advocated in [4] to (roughly) determine a value for β on the coarsest grid first, and then keep it fixed when continuing to finer grids (with W in scaled form). The resulting algorithm typically requires only one or two Newton-type iterations on the finest grids. However, in the present article we have opted not to use a nested iteration in order to study the impact of our multigrid algorithm without the camouflaging effect of the continuation technique.

Numerical results for the problem (1.2) are presented in §4. The modified cell-centered multigrid method is seen to produce an efficient algorithm. Furthermore, as mentioned earlier, since there are an inner, multigrid iteration, and an outer, nonlinear optimization iteration, it is wasteful to solve the inner iteration to excessive accuracy (despite multigrid speed) while being far away from the solution of (1.5).

Moreover, it is useful following each nonlinear iteration to apply some iterations towards the solution of the forward problem (1.5a). This idea is reminiscent of post-stabilization [3, 1] and secondary correction in SQP methods [33], and it has advantages both in hastening convergence and in making the solution feasible before reaching optimality. Details of this idea and its performance are given in §4 as well.

In §5 we summarize our findings and then embark upon a discussion of several additional issues.

2. The discretized PDE system. The present section concentrates on the discretization of the large, sparse, linear system (1.6). We consider a modified cell-centered discretization and discuss ellipticity properties in the context of the inverse problem.

2.1. Ellipticity of the PDE system. We now proceed to investigate ellipticity of the linearized system. Let us assume for notational simplicity that the data are given everywhere, i.e. $Q \to I$ if the measurement is of u and $Q \to \nabla$ if the measurement is of ∇u values. Denote

$$\nabla u = \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T, \quad \nabla\lambda = \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \gamma_3)^T.$$

Furthermore, return to the differential equation (i.e. let the grid width tend to 0) and consider the problem (1.2). It is straightforward to show that the continuous operators which correspond to the discrete operators are

$$A \to \nabla \cdot (\sigma \nabla(\cdot)) \tag{2.1a}$$
$$W^T W \to -\nabla \cdot (\mathrm{a}\nabla(\cdot)) \tag{2.1b}$$
$$K(m,\lambda) \to \nabla \cdot [\sigma \boldsymbol{\gamma}^T (\cdot)] \tag{2.1c}$$
$$G(m,u) \to \nabla \cdot [\sigma \boldsymbol{\alpha}^T (\cdot)] \tag{2.1d}$$
$$T(m,u,\lambda) \to \sigma\, \boldsymbol{\alpha}^T \boldsymbol{\gamma}\, (\cdot) \tag{2.1e}$$
$$K^T(m,\lambda) \to -\sigma \boldsymbol{\gamma}^T (\nabla(\cdot)) \tag{2.1f}$$
$$G^T(m,u) \to -\sigma \boldsymbol{\alpha}^T (\nabla(\cdot)). \tag{2.1g}$$

Here $\sigma(\mathbf{x}) = e^{\mathrm{m}(\mathbf{x})} > 0$, and (2.1b) is defined with natural boundary conditions.

We consider a solution form involving only one wave number,

$$\begin{pmatrix} \delta\mathrm{u} \\ \delta\lambda \\ \delta\mathrm{m} \end{pmatrix} = \begin{pmatrix} \hat{\delta\mathrm{u}} \\ \hat{\delta\lambda} \\ \hat{\delta\mathrm{m}} \end{pmatrix} e^{\imath \boldsymbol{\theta}^T \mathbf{x}},$$

where $\hat{\delta\mathrm{u}}$, $\hat{\delta\lambda}$ and $\hat{\delta\mathrm{m}}$ are amplitudes, and attention is restricted to high wave numbers $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$. For the present purpose we also assume that $\sigma, \boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are constants and $\mathrm{a} = I$.

For the $Q \to I$ case, this yields the symbol

$$C = \begin{pmatrix} -\sigma\|\boldsymbol{\theta}\|^2 & 0 & \imath\sigma\boldsymbol{\alpha}^T\boldsymbol{\theta} \\ 1 & -\sigma\|\boldsymbol{\theta}\|^2 & \imath\sigma\boldsymbol{\gamma}^T\boldsymbol{\theta} \\ -\imath\sigma\boldsymbol{\gamma}^T\boldsymbol{\theta} & -\imath\sigma\boldsymbol{\alpha}^T\boldsymbol{\theta} & \beta\|\boldsymbol{\theta}\|^2 + \sigma\boldsymbol{\alpha}^T\boldsymbol{\gamma} \end{pmatrix}. \tag{2.2}$$

Thus,

$$\begin{aligned} \det(C) = {} & \beta\sigma^2\|\boldsymbol{\theta}\|^6 \\ & + \sigma^3\boldsymbol{\alpha}^T\boldsymbol{\gamma}\|\boldsymbol{\theta}\|^4 + 2\sigma^3[\boldsymbol{\alpha}^T\boldsymbol{\theta}][\boldsymbol{\gamma}^T\boldsymbol{\theta}]\|\boldsymbol{\theta}\|^2 \\ & + \sigma^2[\boldsymbol{\alpha}^T\boldsymbol{\theta}]^2. \end{aligned} \tag{2.3}$$

Clearly, for any $\beta > 0$ fixed, the system is elliptic, because ellipticity is determined by regularity of C for all "sufficiently large" $\|\boldsymbol{\theta}\|$. The system is elliptic also in the case that $Q \to \nabla$; indeed, the only change in (2.3) is that the last term gets multiplied by $\|\boldsymbol{\theta}\|^2$.

Furthermore, consider the Gauss-Newton method: We set $K = 0$ and $T = 0$ in (1.6) and call the resulting symbol C_{GN}. In this case we have

$$\det(C_{GN}) = \beta\sigma^2\|\boldsymbol{\theta}\|^6 + \sigma^2[\boldsymbol{\alpha}^T\boldsymbol{\theta}]^2$$

for $Q \to I$ and

$$\det(C_{GN}) = \beta\sigma^2\|\boldsymbol{\theta}\|^6 + \sigma^2\|\boldsymbol{\theta}\|^2[\boldsymbol{\alpha}^T\boldsymbol{\theta}]^2$$

for $Q \to \nabla$. Thus, $\det(C_{GN}) > 0$ for every $\boldsymbol{\theta} \neq \mathbf{0}$.

2.2. Cell-centered discretization. We proceed here as in [4]. Thus, we write (1.2) as

$$\nabla \cdot \mathbf{J} = \mathrm{q} \quad \text{in } \Omega, \tag{2.4a}$$

$$\mathbf{J} - e^{\mathrm{m}}\nabla \mathrm{u} = \mathbf{0} \quad \text{in } \Omega, \tag{2.4b}$$

$$\mathbf{J} \cdot \mathbf{n}\Big|_{\partial\Omega} = 0 \tag{2.4c}$$

where $\Omega = [-1, 1]^3$. Again, we set $\sigma = e^{\mathrm{m}}$.

With N a positive integer and $h = 2/N$, we consider Ω as the union of N^3 cubic cells of side h each. This is the *primary grid*. Inside the (i, j, k)th grid cell we approximate m by a constant $m_{i,j,k}$, $1 \le i, j, k \le N$. See Figure 2.1.

Next, we integrate (2.4a) over each cell of the primary grid and use the midpoint rule. This yields

$$\begin{aligned} & h^{-1}[J^x_{i+1/2,j,k} - J^x_{i-1/2,j,k} + J^y_{i,j+1/2,k} - J^y_{i,j-1/2,k} + \\ & J^z_{i,j,k+1/2} - J^z_{i,j,k-1/2}] = q_{i,j,k}, \quad 1 \le i, j, k \le N. \end{aligned} \tag{2.5a}$$

The boundary conditions (2.4c) are naturally used to eliminate values of J^x, J^y, or J^z at cells which are next to a boundary in (2.5a).

The components of u are placed, like m, at cell centers. The x-component, say, of (2.4b) is then discretized centered at the x-face of the cell, yielding

$$h^{-1}(u_{i+1,j,k} - u_{i,j,k}) = \sigma^{-1}_{i+1/2,j,k} J^x_{i+1/2,j,k}, \tag{2.5b}$$

where

$$\sigma_{i+1/2,j,k} = 1/(e^{-m_{i+1,j,k}} + e^{-m_{i,j,k}}) \tag{2.5c}$$

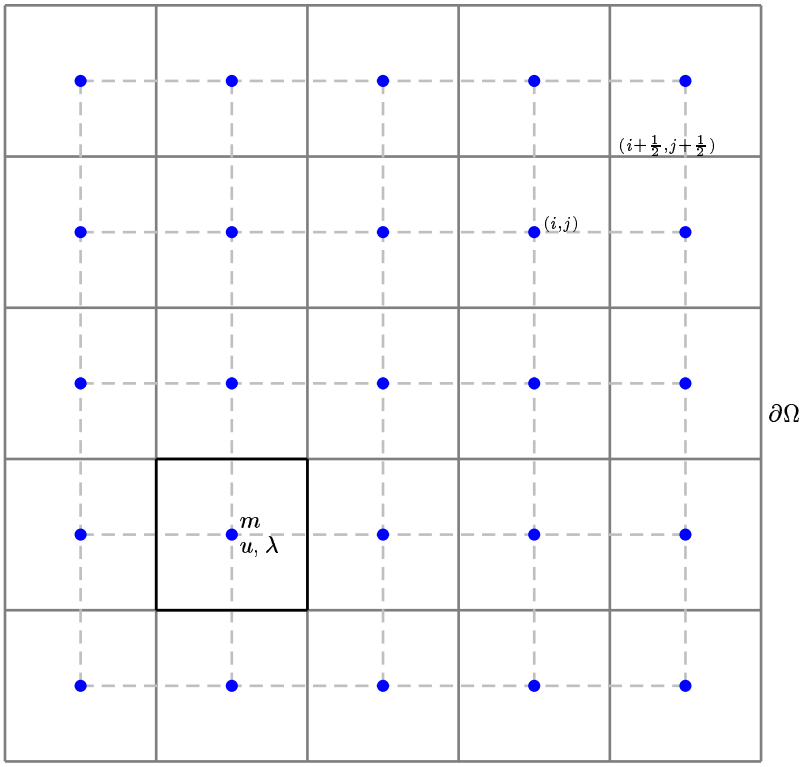

FIG. 2.1. *Cross-sections of the primary grid (solid) and dual grid (dashed), and the cell-centered discretization.*

is a harmonic average of two neighboring cell properties.

Using expressions similar to (2.5b) also in the y- and z-directions, the components of $\mathbf{J}$ in (2.5a) which are inside Ω can be eliminated. This yields a system of N^3 linear algebraic equations for u (for a given m) based on a 7-point discretization stencil. This system of equations has a constant null-space, but it becomes nonsingular upon setting u at a corner of the grid to 0.

For m we have a straightforward Poisson-like equation with natural boundary conditions, and we utilize the standard 7-point formula (in 3D). The resulting discrete system written as (1.6) is therefore cell-centered and non-staggered. The unknowns are all placed at the nodes of the dual grid. (See Figure 2.1, where these nodes correspond to the blue dots.)

This discretization appears to provide a natural, centered, compact scheme in (1.3). However, the resulting system (1.6) involves "long differences" in the terms $G^T\delta\lambda$ and $G\delta m$. Indeed, these terms correspond, according to (2.1g), to a discretization of $-\sigma(\nabla \mathrm{u})^T(\nabla(\delta\lambda))$ and $-\nabla\cdot[(\sigma\nabla \mathrm{u})\,\delta \mathrm{m}]$, respectively, and our discretization is non-staggered. By looking at the last row and the last column of H in (1.6) we would expect the latter fact to become important when $\beta > 0$ is very small.

To investigate the potential effect of this further, let us simplify notation by setting $\lambda \equiv 0$ in H of (1.6). This corresponds to either considering a Gauss-Newton iteration or considering a Newton iteration from such an initial guess (as we utilize in §4). Then $\boldsymbol{\gamma} = \mathbf{0}$ as well. Applying a local Fourier analysis [11, 40] for the case $Q \to I$, we consider the effect of the discretized, frozen differential operators on a mode

$$\begin{pmatrix} \hat{\delta u} \\ \hat{\delta\lambda} \\ \hat{\delta m} \end{pmatrix} e^{\imath \boldsymbol{\xi}^T \mathbf{x}/h}, \quad \boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T, -\pi \le \xi_i \le \pi.$$

This yields the discrete symbol matrix

$$C_h = \begin{pmatrix} -\sigma\hat{a}(\boldsymbol{\xi}) & 0 & \imath\sigma\hat{b}(\boldsymbol{\xi}) \\ 1 & -\sigma\hat{a}(\boldsymbol{\xi}) & 0 \\ 0 & -\imath\sigma\hat{b}(\boldsymbol{\xi}) & \beta\hat{a}(\boldsymbol{\xi}) \end{pmatrix} \quad \Rightarrow det(C_h) = \sigma^2(\beta\hat{a}^3 + \hat{b}^2)$$

where

$$\hat{a}(\boldsymbol{\xi}) = 2h^{-2}(3 - \cos\xi_1 - \cos\xi_2 - \cos\xi_3) > 0, \tag{2.6a}$$

$$\hat{b}(\boldsymbol{\xi}) = h^{-1}(\alpha_1 \sin\xi_1 + \alpha_2 \sin\xi_2 + \alpha_3 \sin\xi_3). \tag{2.6b}$$

Recall that a good measure of h-ellipticity is necessary for being able to design an effective relaxation method (e.g. see Chapter 8 of [40] and references therein). To assess this, we must consider high frequency modes $\boldsymbol{\xi} \in T_{high}$. Note that $\hat{b}(\pi,\pi,\pi) = 0$. For β small enough, therefore, the h-ellipticity measure of the system is approximately given by

$$E_h = \frac{\min\{det(C_h); \boldsymbol{\xi} \in T_{high}\}}{\max\{det(C_h)\}} \approx \frac{8\beta h^{-6}}{h^{-2}\|\boldsymbol{\alpha}\|^2} = \frac{8}{\|\boldsymbol{\alpha}\|^2}\beta h^{-4}. \tag{2.7}$$

In this case, the h-ellipticity measure E_h depends on the grid and, for a coarse enough grid and a small enough regularization parameter, it is clearly too small for practical purposes. However, this happens only when $\beta \ll h^4$.

For the case $Q \to \nabla$, a similar analysis yields

$$\det(C_h) = \sigma^2\hat{a}(\beta\hat{a}^2 + \hat{b}^2).$$

In this case the h-ellipticity measure becomes too small already when $\beta \ll h^2$.

Our remedy is simply to increase β inside H if necessary to ensure that, in the linear operator, $\beta \geq \eta h^4$ (or $\beta \geq \eta h^2$ for $Q \to \nabla$) for a sufficiently large constant η. (Thus, introducing this remedy depends on the values of both β and h.) This corresponds directly to a Levenberg-Marquardt correction for the reduced Hessian (see [22, 33]).

Alternatively, one can use a *cell-nodal* discretization, as described next. Designing a cheap, effective smoother for the resulting system is not simple, hence we end up not using the cell-nodal discretization. However, we develop it for the sake of completeness.

2.3. Cell-nodal discretization. The reason that E_h may become small in (2.7) as $\beta \to 0$ is that $\hat{b}(\boldsymbol{\xi})$ of (2.6b) vanishes if $\xi_i = \pi, i = 1,2,3$. This will be avoided if only short differences are used in discretizing ∇u and $\nabla\lambda$, essentially because terms involving $\sin\xi_i$ in (2.6b) will be replaced by terms involving $\sin\frac{\xi_i}{2}$. The latter does not vanish for high frequencies, i.e. when $\frac{\pi}{2} \leq |\xi_i| \leq \pi$ for some i.

A way to avoid long difference discretizations for ∇u and for $\nabla\lambda$ is to place the corresponding unknowns of u and λ at the nodes of the primary grid while keeping m cell-centered. See Figure 2.2.

In (1.3), and thus also in (1.5) and (1.6), the discretization for m remains unchanged and centered at cell centers. But the forward model discretization is now centered at the nodes of the primary grid, which are the cell centers of the dual grid.

Thus, (2.4a) is now integrated over dual grid cells. This yields

$$h^{-1}[J^x_{i+1,j+1/2,k+1/2} - J^x_{i,j+1/2,k+1/2} + J^y_{i+1/2,j+1,k+1/2} - J^y_{i+1/2,j,k+1/2} + J^z_{i+1/2,j+1/2,k+1} - J^z_{i+1/2,j+1/2,k}] = q_{i+1/2,j+1/2,k+1/2}, \quad 0 \leq i,j,k \leq N. \tag{2.8a}$$

Further, the x-component, say, of (2.4b) is now discretized along x-edges of the primary grid, yielding

$$\sigma_{i,j+1/2,k+1/2}(u_{i+1/2,j+1/2,k+1/2} - u_{i-1/2,j+1/2,k+1/2})/h = J^x_{i,j+1/2,k+1/2}, \tag{2.8b}$$
$$0 \leq i \leq N+1, \ 1 \leq j,k \leq N,$$

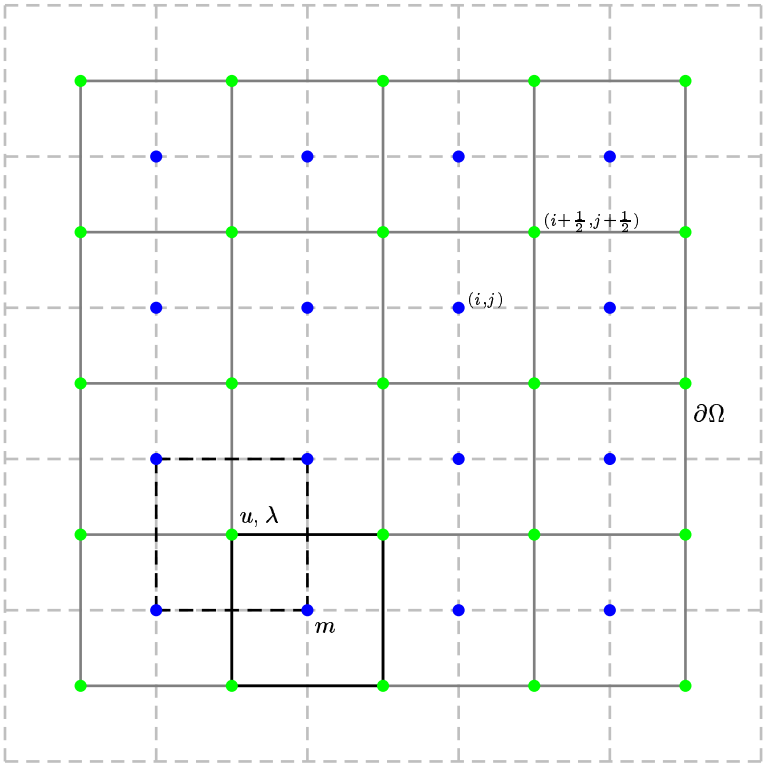

FIG. 2.2. *Cross-sections of the primary grid (solid) and dual grid (dashed), and the cell-nodal discretization.*

where

$$\sigma_{i,j+1/2,k+1/2} = \frac{1}{4}(e^{m_{i,j,k}} + e^{m_{i,j+1,k}} + e^{m_{i,j,k+1}} + e^{m_{i,j+1,k+1}}) \tag{2.8c}$$

is an arithmetic average of the four neighboring cell properties; see [21], where a similar procedure is developed to handle permeability with possible discontinuities in Maxwell's equations. The boundary conditions (2.4c) are then used to eliminate "ghost unknowns" which are outside Ω, e.g., $u_{-1/2,j,k} = u_{3/2,j,k}$, $u_{N+3/2,j,k} = u_{N-1/2,j,k}$, etc.

Using expressions similar to (2.8b) also in the y- and z-directions, the components of $\mathbf{J}$ in (2.8a) can be eliminated. This yields a system of $(N+1)^3$ linear algebraic equations for u (for a given m) based on a 7-point discretization stencil. As before, a constant null-space is eliminated in a standard way.

These discretizations of the forward model and of the regularization term now form the discrete optimization problem (1.3). We proceed from here to form the necessary conditions for an optimum (1.5) and the linearized iteration (1.6). It is not difficult to see that now the h-ellipticity measure remains bounded away from 0 for all $\beta > 0$ (unlike (2.7)).

Note that the formation of all matrices appearing in (1.5) and (1.6) remains straightforward and cheap to evaluate, and it produces very sparse, structured matrices.

However, the design of a point relaxation scheme for this staggered-grid discretization becomes difficult precisely when $\beta \ll h^4$. Neighboring values of m, u and λ become strongly coupled then, and a box (or Vanka) relaxation [40] would involve all unknowns in both one cell of the primary grid and an overlapping cell in the dual grid. This is prohibitively expensive, so we stay with the cell-centered alternative in the sequel.

3. Elements of the multigrid method. The cell-based, finite volume discretization described in §2.2 may suggest at first use of a cell-based coarsening within a multigrid algorithm. However, this is not mandatory. Indeed, once the finest-grid equations are generated we may treat the resulting system as a non-staggered discretization of some continuous problem with averaged coefficients, and proceed with a standard vertex-based coarsening [40, 14]. The "vertices" here are the nodes of the dual grid. Thus, a trilinear prolongation and its adjoint, full-weight restriction operator are applied for each grid function δu, δm and $\delta\lambda$. Note that we are treating m here as a nodal representation of a continuous, piecewise linear function, whereas in the derivation of the discretization m is treated as piecewise constant: This inconsistency of interpretation, although seldom noticed, is standard and produces an acceptable approximation method [4].

For the coarse grid operators we employ a standard Galerkin operator.

As mentioned before the PDE system which (1.5) discretizes is strongly coupled when β is small. In such cases (which are the ones of interest if the recovered image of m is to be sharp) we employ a collective point-relaxation, where all three unknowns corresponding to one grid point are relaxed simultaneously. A popular choice is red-black Gauss-Seidel. Another possibility is weighted (under-relaxed) collective Jacobi [11, 40]. We use the latter (with a weight, or damping factor, of 0.8), although Gauss-Seidel variants have better smoothing properties, because of ease of vectorization. In particular, using MATLAB we can treat block diagonal matrices very efficiently and generate a rather fast code.

Rates of smoothing using standard local Fourier analysis (see [11, 40]) are reported in Table 3.1 for $\boldsymbol{\alpha} = (1,1,1)^T$. The rates are seen to be acceptable so long as $\beta \geq h^4$. Thus, assuming h-ellipticity, these point relaxations provide effective smoothers and handle the PDE coupling on a local level.

TABLE 3.1
Smoothing rates for collective Gauss-Seidel and for collective weighted Jacobi (weight = 0.8) relaxations.

β	h	Gauss-Seidel	weighted Jacobi
.05	.05	.56	.74
.0025	.05	.57	.75
.635e-5	.05	.76	.80
.3125e-6	.05	1.24	1.30
.1	.1	.57	.72
.01	.1	.59	.72
.001	.1	.64	.74
.0001	.1	.76	.88
.00001	.1	1.07	1.65

However, for very small β/h^4 the discrete system no longer possesses a good h-ellipticity measure (cf. (2.7)). To overcome this difficulty we add "artificial regularization" by increasing β inside H (but not at the right hand side) of (1.6). This depends on the grid size at each level. Specifically, we set according to (2.7)

$$\beta_h = \max(\beta, \beta_{\min}) \tag{3.1}$$

where

$$\beta_{\min} = \begin{cases} h^4\eta & Q \to I \\ h^2\eta & Q \to \nabla \end{cases}$$

$$\eta = \frac{1}{8}\max_{\mathbf{x}}\left[\frac{\mathrm{u}_x^2}{\mathrm{a}_{11}} + \frac{\mathrm{u}_y^2}{\mathrm{a}_{22}} + \frac{\mathrm{u}_z^2}{\mathrm{a}_{33}}\right], \quad \mathrm{a} = \mathrm{diag}\{\mathrm{a}_{11}, \mathrm{a}_{22}, \mathrm{a}_{33}\}.$$

Note that if we use a fine enough *finest* grid then $\beta_h > \beta$ only *inside* the multigrid cycle and not on the finest level. In such a case, the *outer* nonlinear iteration is not affected. If, on the other hand, we need to increase β on the finest level as well then this may slow down the convergence of the nonlinear iteration. The effect is similar to increasing the parameter in the Levenberg-Marquardt method which is commonly used for highly ill-conditioned problems [33].

These elements are combined into a Newton-multigrid method where a Newton variant, with a global convergence control, is applied at each iteration (see [23, 33]), and a linear V(2,2)-cycle of multigrid inner iteration based on the above method commences thereupon.

4. Numerical results. Similar to [22], we synthesize data for our numerical experiment by assuming that the "true model" $\mathrm{m}(\mathbf{x})$ is given by

$$\begin{aligned}\mathrm{m}(x,y,z) = [&3(1-x)^2\exp(-(x^2)-(y+1)^2-3(z+1)^2)\\ &-10(x/5-x^3-y^5-z^5)\exp(-x^2-y^2-3z^2)\\ &-1/3\exp(-(x+1)^2-y^2-3z^2)-2]/4.\end{aligned}$$

Four sets of data are then calculated for this model. First, we assume that data are available everywhere on the finest grid[3] and calculate either the field u or its gradient ∇u. Then, we assume that the data (u or ∇u) are limited and given on an 8^3-grid uniformly spaced in the cube $[-0.6, 0.6]^3$.

To obtain the data, we use the cell-based discretization for the forward problem (1.2) and calculate the fields on a 129^3-grid. This simulates the continuous process and we can evaluate the discretization error based on the solution of this very fine grid problem. The finest grid used for the actual inverse problem calculations is chosen to be uniform and consist of 49^3 cells, which leads to $352,947$ unknowns. Comparing the very fine scale (129^3 cells) solution of the forward problem to the solution on 49^3 cells we find that the discretization error for the latter is roughly 0.2%. We then pollute the data with Gaussian noise with a larger maximum norm than the discretization error's and recover the corresponding model. The initial iterate is described in [22].

4.1. Solving the linear system for different β-values and different grids. In the first set of experiments we solve the linear system which arises for the first nonlinear iteration. We set $\mathrm{a} = I$, i.e. W corresponds to the discretization of the gradient operator, and thus obtain an isotropically smooth solution. The noise level in the data is set to be $5\%, 2\%$ and 0.5%. For each of these noise levels we pick an appropriate value for β as described in [4] and record the rate of convergence and the number of multigrid levels used. The higher the number of grids the coarser is the coarsest grid.

As expected from the theory, stabilization of the cell-centered discretization is required for small values of the regularization parameter, where the h-ellipticity measure is small. This effect arises sooner (i.e. at finer grids) for ∇u-data than for u-data.

4.2. Solving the linear systems within an inexact Newton-type method. The calculations in §4.1 were performed with a stringent tolerance for the linear system. But within the nonlinear iteration framework for (1.3), the linear KKT system (1.6) at each iteration may be solved only approximately [33, 28, 29, 13, 40]. Thus, our proposed approach becomes even more advantageous when employing an inexact Newton-type method, as very few (usually only one) V-cycles are needed per inexact Newton iteration.

Using an inexact strategy in the constrained optimization context raises the prospect of global convergence. In our code we employ a line search strategy with the l_1 merit function as formulated in [33]. We also use the option of secondary correction [33] to the step after each inexact outer iteration, which can be considered as an inexact version of the procedure N3 proposed in [23]. We carry out the process by iterating towards the solution of the forward problem (the constraint) *after* each inexact step is computed, applying one V(2,2) cycle to the forward problem (1.5a). If we use this option then feasibility is reached earlier than optimality. This may be important because in

[3]This assumption is common in optimal control.

TABLE 4.1

Rates of convergence for different noise levels (and corresponding values of β) and different discretizations. CC - cell-centered, SCC - stabilized cell-centered. The notation $(\cdot)_I$ implies limited data, with the field interpolated to 8^3 measurement locations. No convergence is denote by ''.*

Measurement	β	noise	Levels	CC	SCC
∇u	2.5e − 4	5%	5	*	0.21
			4	0.48	0.21
	1.7e − 5	2%	5	*	0.21
			4	0.53	0.21
	4.6e − 6	0.5%	5	*	0.21
			4	*	0.21
u	3.5e − 2	5%	5	*	0.23
			4	0.46	0.23
	2.0e − 3	2%	5	*	0.23
			4	0.49	0.23
	2.7e − 4	0.5%	5	*	0.23
			4	0.57	0.23
$(\nabla \mathrm{u})_I$	1.3e − 4	5%	5	*	0.32
			4	0.51	0.32
	8.6e − 6	2%	5	*	0.32
			4	*	0.32
	2.1e − 6	0.5%	5	*	0.32
			4	*	0.32
u_I	1.5e − 2	5%	5	*	0.33
			4	0.53	0.33
	1.1e − 3	2%	5	*	0.33
			4	0.58	0.33
	1.2e − 4	0.5%	5	*	0.33
			4	*	0.33

most applications the optimization process merely helps to obtain a reasonable solution whereas the constraint approximates the physics of the problem and should as such be taken more seriously.

We therefore conduct the following experiments where we solve (1.6) to a rough accuracy of 0.5 using the stabilized cell-centered discretization. We also experiment with different matrix functions $\mathrm{a}(\mathbf{x})$ in the regularization operator.

- i. As in the first experiment we set $\mathrm{a} = I$ (the model is smooth in all directions).
- ii. Common to many geophysical scenarios we set the weight in the depth variable lower than in the other directions: $\mathrm{a} = \mathrm{diag}\{3, 3, 1\}$. Thus we use a priori information which should lead to a smoother model in the $x - y$ directions than in the z direction.

It is important to note that any robust computational technique for this type of problems should be able to incorporate such simple a priori information as it may vary with the application and the scenario [30]. As before, the noise level is set and the regularization parameter is then chosen correspondingly. Obviously, the correct regularization parameter value is different for different choices of W and Q.

Table 4.2 shows that combining our solver with an inexact Newton-type iteration is very powerful for all the above scenarios. Indeed, the whole nonlinear problem is solved in a cost which is not much higher than solving a few forward problems. In Figure 4.1 we plot the norms of the gradient of $\mathcal{L}$ and the residual of the constraint. Note that the secondary correction does help to obtain a reasonable feasible solution well before optimality is reached.

TABLE 4.2
Numerical Experiment 3: Number of nonlinear iterations (= number of V(2,2)-cycles) needed to solve the nonlinear problem when using an inexact Newton-multigrid method.

Data	Regularization	Noise	Inexact Newton iterations
$(\nabla u)_I$	i	5%	16
		2%	24
	ii	5%	17
		2%	25
u_I	i	5%	15
		2%	23
	ii	5%	17
		2%	24

As with most multigrid variants, our method can fail if $\mathrm{a}(\mathbf{x})$ has jump discontinuities (as in total variation regularization) unless an operator-induced prolongation (and restriction) is used. We do not report on the latter sort of experiments in this paper.

It is also important to see that adding the a priori information makes a difference in the recovered models. Figure 4.2 displays an $x-z$ slice through the center of the model, reconstructed using the different regularization functionals. The models are obviously quite different. Finally, in Figure 4.3 we plot a slice through the true model and the reconstructed model using the isotropic regularization $\mathrm{a}=I$.

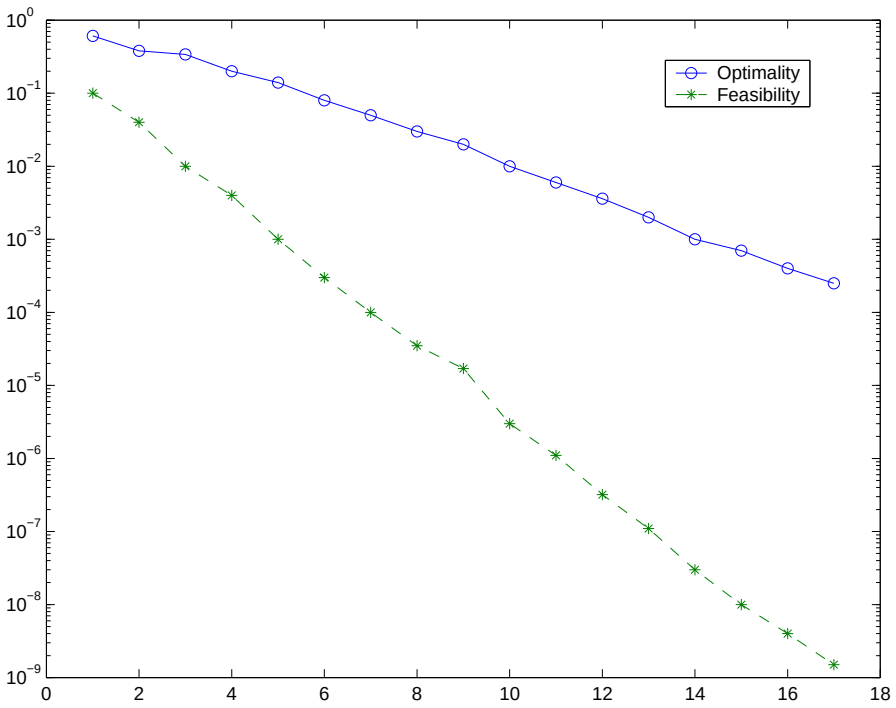

FIG. 4.1. *Convergence curve for the experiment (ii) with* $(\nabla u)_I$ *data.*

5. Summary and discussion. An efficient multigrid method has been developed and demonstrated for large, sparse inverse problems arising from distributed parameter estimation in the context of elliptic PDEs in 3D. Our guiding principle has been to balance the various iterative processes

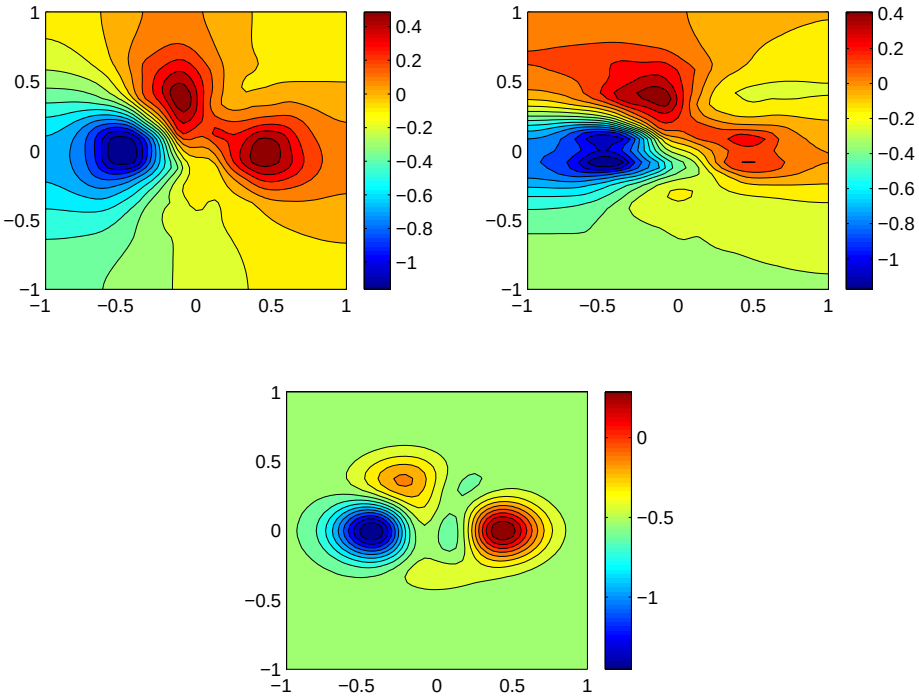

FIG. 4.2. *Slices in $x-z$ plane through the true model (bottom) and the reconstructed models using the isotropic gradient (left) and the non-isotropic gradient (right) in the regularization functional for the a priori information.*

involved, avoiding a costly premature elimination of some unknowns in terms of others as well as utilizing *inexact* Newton-type methods.

Given the discretized constrained optimization problem (1.3), we therefore solve simultaneously the discretized PDE system which results from the necessary conditions for a critical point of the Lagrangian. The relevant underlying PDE system is strongly coupled when the regularization parameter β attains a very small value. Moreover, the h-ellipticity measure for a cell-centered discretization deteriorates as βh^{-4} (or βh^{-2} when $Q \to \nabla$) becomes small. Thus, we use a collective weighted Jacobi relaxation and increase β inside H of (1.6) if necessary.

For the experiments reported in §4 we always choose the value of β based on the Morozov discrepancy principle (see, e.g., [4] and references therein). If β is arbitrarily decreased with the noise level held fixed then the computational method may start fitting the noise in the data. This in turn may cause the obtained model m to have unphysical high frequency oscillations. The convergence of the multigrid method could subsequently slow down as well, then, basically because the solution looks significanly different in different grid resolutions. Thus, one should avoid using unreasonably small values of β – for more than one reason.

Furthermore, we advocate use of an inexact Newton-type method for the constrained optimization problem (1.3) with a $V(2,2)$ cycle (or two) as the inexact solver for the linearized system at each iteration. Combining all these elements results in a very effective method, as demonstrated in §4.

There seem to be surprisingly few multigrid methods developed in the literature for inverse problems; we mention [26, 32, 31, 2, 9], none of which is close to our proposed approach.

Although multigrid may not be the approach of choice for all inverse problems (e.g., it may be hard to generate a robust multigrid method for the forward problem) it clearly holds promise; moreover, there are some interesting observations for general inverse problems that arise from our

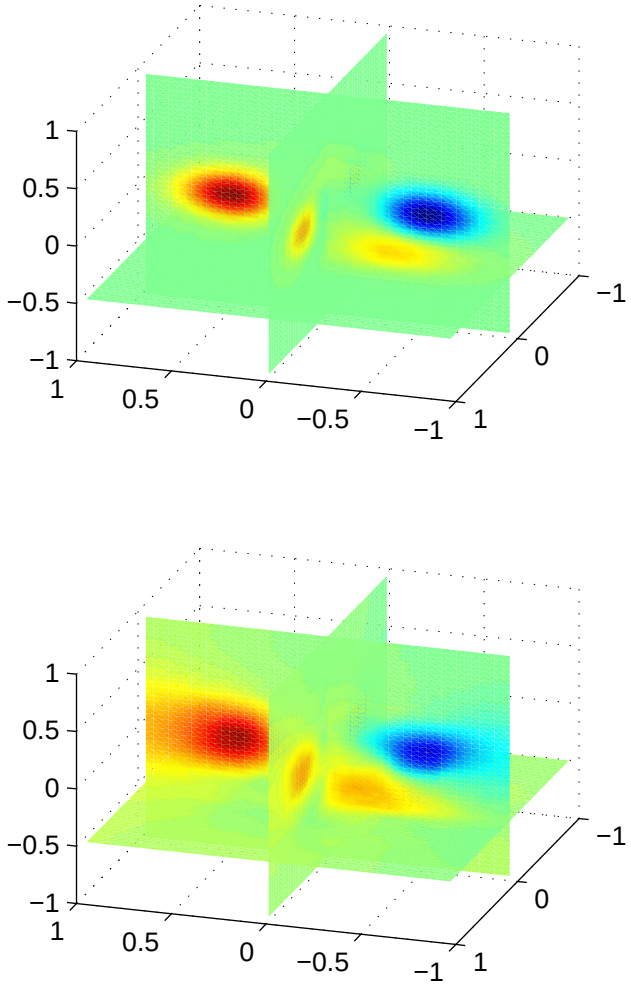

FIG. 4.3. *Slices in the true model (top) and the reconstructed model using the isotropic gradient (i).*

multigrid analysis. First, compact dicretizations of the forward problem and of the regularization term do not necessarily yield a compact discretization for the combined PDE system arising from the constrained approach for very small values of β. This can affect other inversion techniques, too. Note that such a discretization enhances the null space of the sensitivity matrix $\frac{\partial Qu}{\partial m}$. This arises in other methods, too, for example in the popular unconstrained Gauss-Newton formulation, for reasons that relate to the discretization rather than the ill-posedness of the problem. In these cases, regularization may be needed to make the problem well-posed even for very accurate data. Second, we have explored another discretization technique for our inverse problem. While this discretization is stable and may still yield a fast multigrid method for the forward problem, using multigrid for the inverse problem can be quite expensive because the smoothing cannot be done collectively. Finally, similar to our work in [22] we note that by using the constrained approach and applying an inexact Newton version for the optimization problem we can dramatically reduce the cost of solving the inverse problem.

REFERENCES

[1] U. Ascher. Stabilization of invariants of discretized differential systems. *Numerical Algorithms*, 14:1–23, 1997.
[2] U. Ascher and P. Carter. A multigrid method for shape from shading. *SIAM J. Numer. Anal.*, 30:102–115, 1993.
[3] U. Ascher, H. Chin, L. Petzold, and S. Reich. Stabilization of constrained mechanical systems with daes and invariant manifolds. *J. Mech. Struct. Machines*, 23:135–158, 1995.
[4] U. Ascher and E. Haber. Grid refinement and scaling for distributed parameter estimation problems. *Inverse Problems*, 17:571–590, 2001.
[5] J. Betts and W. Huffman. Path constrained trajectory optimization using sparse sequential quadratic programming. *AIAA J. Guidance, Cont. Dyn.*, 16:59–68, 1993.
[6] J. T. Betts. *Practical Methods for Optimal Control using Nonlinear Programming.* SIAM, Philadelphia, 2001.
[7] G. Biros and O. Ghattas. Parallel Newton-Krylov methods for PDE-constrained optimization. In *Proceedings of CS99, Portland Oregon*, 1999.
[8] G. Biros and O. Ghattas. Parallel preconditioners for KKT systems arising in optimal control of viscous incompressible flows. In *Proceedings of Parallel CFD*. North Holland, 1999. May 23-26, Williamsburg, VA.
[9] L. Borcea. Nonlinear multigrid for imaging electrical conductivity and permittivity at low frequency. *Inverse Problems*, 17:329–360, 2001.
[10] L. Borcea, J. G. Berryman, and G. C. Papanicolaou. High-contrast impedance tomography. *Inverse Problems*, 12, 1996.
[11] A. Brandt. Multigrid techniques: 1984 Guide with applications to fluid Dynamics. The Weizmann Institute of Science, Rehovot, Israel, 1984.
[12] A. Brandt and N. Dinar. Multigrid solution to elliptic flow problems. In *Numerical Methods for PDE's*, pages 53–147. Academic Press, New York, 1979.
[13] P. N. Brown and Y. Saad. Hybrid krylov methods for nonlinear systems of equations. *SIAM J. Sci. Statist.*, 11:450–481, 1990.
[14] J.E. Dendy, Jr. Black box multigrid. *J. Comput. Phys.*, 48:366–386, 1982.
[15] A. J. Devaney. The limited-view problem in diffraction tomography. *Inverse Problems*, 5:510–523, 1989.
[16] R. Ewing (Ed.). *The mathematics of reservoir simulation.* SIAM, Philadelphia, 1983.
[17] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer, 1996.
[18] C. Farquharson and D. Oldenburg. Non-linear inversion using general measures of data misfit and model structure. *Geophysics J.*, 134:213–227, 1998.
[19] G. Golub and Q. Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Scient. Comp.*, 21:1305–1320, 2000.
[20] S. Gomez, A. Perez, and R. Alvarez. Multiscale optimization for aquifer parameter identification with noisy data. In *Computational Methods in Water Resources XII, Vol. 2*, 1998.
[21] E. Haber and U. Ascher. Fast finite volume simulation of 3D electromagnetic problems with highly discontinuous coefficients. *SIAM J. Scient. Comput.*, 22:1943–1961, 2001.
[22] E. Haber and U. Ascher. Preconditioned all-at-one methods for large, sparse parameter estimation problems. *Inverse Problems*, 2001. To appear.
[23] E. Haber, U. Ascher, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse problems*, 16:1263–1280, 2000.
[24] P. C. Hansen. *Rank Deficient and Ill-Posed Problems.* SIAM, Philadelphia, 1998.
[25] M. Heinkenschloss and L.N. Vicente. Analysis of inexact trust region SQP algorithms. Technical report, TR 99-18, Rice University, September 1999.
[26] V. Henson, M. Limber, S. McCormick, and B. Robinson. Multilevel image reconstruction with natural pixels. *SIAM J. Scient. Comput.*, 17:193–216, 1996.
[27] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Stats.*, 35:73–101, 1964.
[28] C.T. Kelley. *Iterartive Methods for Linear and Nonlinear Equations.* SIAM, Philadelphia, 1995.
[29] C.T. Kelley. *Iterartive Methods for Optimization.* SIAM, Philadelphia, 1999.
[30] Y. Li and D. Oldenburg. Incorporating geologic dip information into geophysical inversions. *Geophysics*, 65:148–157, 2000.
[31] S. McCormick and J. Wade. Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography. *Inverse problems*, 9:697–713, 1993.
[32] S.G. Nash. A multigrid approach to discretized optimization problems. Technical report, 1999. Manuscript.
[33] J. Nocedal and S. Wright. *Numerical Optimization.* New York: Springer, 1999.
[34] R. L. Parker. *Geophysical Inverse Theory.* Princeton University Press, Princeton NJ, 1994.
[35] V. Schulz. *Reduced SQP methods for large scale optimal control problems in DAE's with application to path planning problems for satellite mounted robots.* PhD thesis, University of Heidelberg, 1996.
[36] A.R. Shenoy, M. Heinkenschloss, and E.M. Cliff. Airfoil design by an all-at-once method. *Int. J. Comp. Fluid Mechanics*, 11:3–25, 1998.
[37] N.C. Smith and K. Vozoff. Two dimensional DC resistivity inversion for dipole dipole data. *IEEE Trans. on geoscience and remote sensing*, GE 22:21–28, 1984. Special issue on electromagnetic methods in applied geophysics.
[38] M. Steinbach, H. G. Bock, and R. W. Longman. Time-optimal extension or retraction in polar coordinate robots: a numerical analysis of the switching structure. In *Proceedings IAAA Guidance, Navigation and Control*,

Boston, 1989.
[39] A.N. Tikhonov and V.Ya. Arsenin. *Methods for Solving Ill-posed Problems*. John Wiley and Sons, Inc., 1977.
[40] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, 2001.
[41] C. Vogel. Sparse matrix computation arising in distributed parameter identification. *SIAM J. Matrix Anal. Appl.*, 20:1027–1037, 1999.
[42] C. Vogel. *Computational methods for inverse problem*. SIAM, Philadelphia, 2001.
[43] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.