Computational methods for large distributed parameter estimation problems with possible discontinuities

Uri M. Ascher* Eldad Haber†

Abstract

This paper considers problems of distributed parameter estimation from data measurements on solutions of diffusive partial differential equations (PDEs). A nonlinear 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 have developed methods for the resulting constrained optimization problem which directly address the discretized, strongly coupled PDE system that defines a critical point of the Lagrangian. Such methods have been applied for electromagnetic data inversion in 3D, both in frequency and in time domains.

In practice the reconstruction process depends crucially on additional, a priori information on the model, which is introduced through the regularization term. We explore the use of the so-called Huber's norm for the recovery of piecewise smooth model functions. Since our forward operators are diffusive, care must be taken not to conclude too much from discontinuity reconstructions. We supply examples.

1 Introduction

The success of many practical results for ill-posed problems depends crucially on the circumstances in which they are derived. Let us therefore start by describing in some detail the class of problems under consideration here.

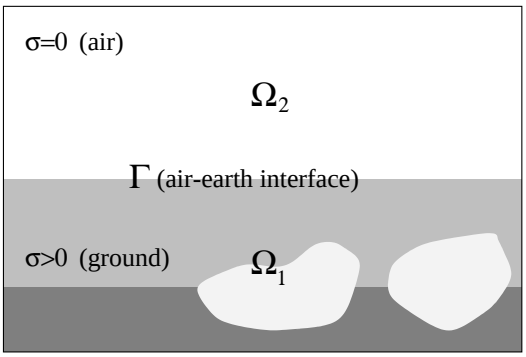

Figure 1: A typical domain cross-section for a geophysical application

Several different applications give rise to a partial differential equation (PDE) of the form

$$\nabla \cdot (\sigma \nabla \mathrm{u}) = \mathrm{q}. \tag{1}$$

Here, the field $\mathrm{u}(\mathbf{x})$ satisfies the PDE plus boundary conditions for given sources $\mathrm{q} = \mathrm{q}(\mathbf{x})$ and a conductivity function $\sigma = \sigma(\mathbf{x})$. Applications include DC resistivity, magnetotelluric inversion, diffraction tomography, impedance tomography, oil reservoir simulation and aquifer calibration [15, 13, 6, 5]. Other applications of interest give rise to the more involved Maxwell's equations, written for the time harmonic case as

$$\nabla \times (\mu^{-1} \nabla \times \mathbf{E}) - \imath\omega\,\sigma\,\mathbf{E} = \imath\omega\mathbf{s}. \tag{2}$$

See, e.g., [11, 7] and references therein. A typical domain Ω in 3D on which such a PDE is defined is depicted in Figure 1.

In a typical application we may have several PDEs such as (1) or (2) corresponding, e.g., to different frequencies and different sources and

*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.

†Department of Mathematics and Computer Science, Emory University, Atlanta, Ga, 30322, USA. (haber@mathcs.emory.edu).

sinks. Whatever of these we have, we group them as well as the given boundary conditions into a unified notation

$$\mathcal{A}(\mathrm{m})\mathrm{u} = \mathrm{q}$$

where the *model* $\mathrm{m}(\mathbf{x})$ is typically a function of σ, e.g.,

$$\sigma(\mathbf{x}) = e^{\mathrm{m}(\mathbf{x})},\ \mathbf{x} \in \Omega.$$

The obtained boundary value PDE is discretized on a tensor product grid, not necessarily uniform, using a finite volume technique [7, 1]. We assume the material properties to be constant in each cell and call the resulting grid functions u, q and m. If needed, they are ordered into vectors. This yields our discretized problem

$$A(m)u = q. \tag{3}$$

The *forward problem* is to find u satisfying (3) given m (and q, which is always assumed given). The *inverse problem* is to recover the model m, given measurements b on the field u such that (3) holds.

However, it is well-known that while the forward problem is well-posed the inverse problem is not. Indeed, in practice for the available, noisy data typically there is no unique solution, i.e., there are many models m which yield a field u which is close to b to within the noise level, and moreover, such models m may vary widely.

Thus, we must add a priori information and isolate noise effects. Then the problem becomes to choose, from all possible model solutions, the one closest to the a priori information. This regularization leads to the optimization problem

$$\begin{aligned} \min_{m,u}\ \ & \phi = \frac{1}{2}\|Qu - b\|^2 + \beta R(m) \\ \text{subject to}\ \ & A(m)u - q = 0. \end{aligned} \tag{4}$$

Here Q is a matrix consisting of unit rows which projects to data locations, $R(m)$ is a regularization term, and $\beta > 0$ is the regularization parameter whose choice has been the subject of many papers. Also, throughout this article the notation $\|\cdot\|$ refers to the l_2 norm.

For the regularization term, consider a same-grid discretization of

$$\mathrm{R}(\mathrm{m}) = \int_\Omega \rho(|\nabla \mathrm{m}|) + \hat{\alpha}(\mathrm{m} - \mathrm{m}_{\mathbf{ref}})^{\mathbf{2}} \tag{5}$$

where $\hat{\alpha} \geq 0$ is a small parameter and $\mathrm{m}_{\mathrm{ref}}$ is a given reference function. The selection of $\mathrm{m}_{\mathrm{ref}}$ can be crucial in geophysical exploration, but in this article we set $\hat{\alpha} = 0$ and concentrate on the choice of the function ρ. The latter relates directly to the a priori information we have about the smoothness of the model. A typical choice is the weighted l_2 norm, but we consider choices of total variation and Huber's function in §4.

Whichever choice is made for ρ, there is a very large, nonlinear optimization problem to solve in (4). This problem is significantly harder to solve when total variation or Huber's function are used, and it also gets harder the smaller β is. It is not unusual in our experience to solve problems of the form (4) for half a million unknowns, so care must be taken to do this effectively: One cannot simply assume the existence of a general software package to do the job.

Fortunately, the matrices appearing in the necessary conditions for (4) are all very sparse and correspond to discretizations of elliptic PDEs, so we exploit in §2 the special structure thus afforded. The details are taken from [8, 2]. An example [9] inverting 3D electromagnetic data is recalled in §3. This example uses a weighted l_2 regularization.

Often, available a priori information about the model suggests a piecewise smooth surface function. A method for reconstructing such surfaces is developed in §4. This is currently a hot area of research. However, for compact forward problems such as ours one should re-evaluate the potential of using discontinuity reconstruction techniques, because in practical situations there is enough uncertainty in the data that arbitrary shifts of discontinuity interfaces within the allowable set of model solutions cannot be precluded. We demonstrate this by examples.

2 Solving the optimization problem

The most obvious first step in order to solve (4), taken by many, is to eliminate the field u using the forward problem. Thus we obtain a much smaller, although still large (and dense) unconstrained minimization problem

$$\min_m\ \phi(m) = \frac{1}{2}\|QA(m)^{-1}q - b\|^2 + \beta R(m). \tag{6}$$

The *unconstrained approach* is to devise numerical methods for solving this unconstrained optimization problem.

Newton and Gauss-Newton methods are well-known. The Gauss-Newton iteration involves positive definite linear systems of the form

$$H_{red}\ \delta m \quad \equiv \quad (J^T J + \beta R'')\ \delta m = -p \quad (7)$$

where $J = -QA^{-1}G$ is the *sensitivity matrix* and $G = \frac{\partial A(m)u}{\partial m}$. The vector p and the matrices J and R'' are evaluated based on a current iterate m and the next iterate is obtained as $m \leftarrow m + \tilde{\alpha}\ \delta m$, for a suitable step size $0 < \tilde{\alpha} \leq 1$.

Whereas G is sparse, the sensitivity matrix J is large and full for applications of the type considered here, so it is never evaluated or stored. To solve (7) we use a Preconditioned Conjugate Gradient method, i.e. the conjugate gradient method for

$$M^{-1}(J^T J + \beta R'')\delta m = -M^{-1}p,$$

where M is a preconditioner, e.g. $M = R''$. This conjugate gradient method requires only the evauation of matrix-vector products involving H_{red}.

Unfortunately, however, *evaluating* $H_{red}v$ for a given vector v is expensive! It involves evaluating Jv and $J^T v$. While multiplying a vector by G or Q or their adjoints is fast, there is A^{-1} in J as well. The forward and adjoint problems must be solved for this purpose to a relatively high accuracy. In the large, sparse context the superiority of the unconstrained approach may thus be challenged. Let us proceed to widen the scope.

Introducing the Lagrangian for the constrained formulation (4),

$$\mathcal{L}(m,u,\lambda) = \frac{1}{2}\|Qu-b\|^2+\beta R(m)+\lambda^T(A(m)u-q),$$

where λ is the vector (or more generally, a grid function like u) of Lagrange multipliers, the necessary conditions for a minimum lead to the large system of nonlinear equations

$$\begin{aligned} \mathcal{L}_\lambda &= Au - q = 0, && (8a)\\ \mathcal{L}_u &= Q^T(Qu - b) + A^T\lambda = 0, && (8b)\\ \mathcal{L}_m &= \beta R'(m) + G^T\lambda = 0. && (8c) \end{aligned}$$

All matrices appearing in (8) are sparse.

For the numerical solution of (8) consider using a variant of Newton's method (e.g. Lagrange-Newton, SQP, Gauss-Newton; see for instance [12]). A Gauss-Newton instance could read

$$\begin{pmatrix} A & 0 & G \\ Q^TQ & A^T & 0 \\ 0 & G^T & \beta R'' \end{pmatrix} \begin{pmatrix} \delta u \\ \delta\lambda \\ \delta m \end{pmatrix} = - \begin{pmatrix} \mathcal{L}_\lambda \\ \mathcal{L}_u \\ \mathcal{L}_m \end{pmatrix}. \quad (9)$$

As for the unconstrained formulation (7), the system (9) describes the equations to be solved at each iteration in order to find the correction vector for a current iterate (u, λ, m). We can distinguish the following *tasks* at each iteration: (i) calculate the Gradients and Hessian; (ii) solve the large, sparse Hessian system; and (iii) update the iteration. Of these tasks the first is straightforward in the present context. The third is textbook material (e.g. [12]) involving methods such as damped Newton, trust region, directly updating λ, etc. We thus concentrate on the remaining, second task.

There are two approaches here: the reduced Hessian approach and the simultaneous, all-at-once approach. In the first of these approaches we eliminate δu from the first block row of (9) (this involves solving the forward problem), then $\delta\lambda$ from the second block row of (9) (this involves solving the adjoint problem), obtaining for δm a reduced system of the form (7). Next, consider the *all-at-once* approach, developed in [8, 2, 3, 4].

The rationale is simple: When the iterate for the update direction is far from the solution to the linear system, it is wasteful to eliminate some variables accurately in terms of others! Thus, one wants to balance accuracies inside the linear solver. This is where the unconstrained approach and the reduced Hessian approach described above fall over.

We were thus led to consider methods where δu, $\delta\lambda$ and δm are eliminated *simultaneously*. This approach is more natural also when considering the origin of the systems (8) and (9). These are, after all, discretizations of systems of PDEs, and the approach of eliminating some PDEs in terms of others instead of solving the given system as one is less usual. Unfortunately, however, the matrix of (9) is no longer positive definite. Moreover, β is typically small, so the problem corresponds to a *strongly coupled* PDE system.

In [8] we proposed a preconditioned QMR method for the symmetrized system (9). The preconditioner consists of applying iterations towards the solution of the reduced problem (7).

The catch is that this no longer has to be performed to a high accuracy. In the next section we demonstrate the performance of this method.

In [2] we proposed a multigrid method for the strongly coupled system (9). The resulting method is very nice and fast, but we have applied it only for the problem (1). The staggered discretization applied for the Maxwell equations in §3 complicates the multigrid technique significantly, so we chose the preconditioned QMR approach for the latter application.

3 Example 1: 3D electromagnetic data inversion

The inversion results reported in this section were obtained using the weighted least squares norm on ∇m for the regularization functional R, where

$$\mathrm{m}(\mathbf{x}) = \ln \sigma(\mathbf{x}).$$

This transformation automatically takes care of the positivity constraint on σ, and it also reduces the contrast in the conductivity, which is reasonable for geophysical mining exploration applications and commensurate with the use of a weighted l_2 norm on $|\nabla \mathrm{m}|$ in R.

Maxwell's equations for the time-harmonic case are written as

$$\begin{aligned} \nabla \times \mathbf{E} + \alpha\mu\mathbf{H} &= \mathbf{s}_H \quad \text{in } \Omega, \\ \nabla \times \mathbf{H} - \hat{\sigma}\mathbf{E} &= \mathbf{s}_E \quad \text{in } \Omega, \\ \vec{n} \times \mathbf{H} &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

where $\hat{\sigma} = \sigma + \alpha\epsilon$ and $\alpha = -\imath\omega$. The permeability $\mu(\mathbf{x})$ and permittivity $\epsilon(\mathbf{x})$ are assumed known throughout. Typical parameter regimes in these applications satisfy $0 \le \epsilon \ll 1$ and exclude high frequencies ω.

Appropriate discretization and solution techniques for this forward model, including boundary conditions, have been described in [7, 9]. They involve a finite volume discretization applied on a staggered grid. Note that the forward problem is linear, while the inverse problem is not.

In a multiple source/frequency experiment, a linear system arises from such a discretization for each pair of frequency and source. The forward problem (3) then consists of the collection of all these subsystems. Obvious parallelism can be exploited since they depend on each other only through the model.

$\beta = 100$	misfit = 0.06		
n-it	KKT-it	infeas	rel-grad
1	4	$3e-2$	2e-1
2	4	$2e-4$	3e-2
3	3	$2e-6$	5e-4
$\beta = 1e0$	misfit = 0.03		
n-it	KKT-it	infeas	rel-grad
1	8	$1e-6$	3e-3
2	6	$8e-7$	9e-4

Table 1: Inverting electromagnetic data in the frequency domain.

In [9] a data inversion experiment is described. It involves a transmitter and receiver geometry from an actual CSAMT field survey, although the conductivity model is synthesized. The transmitter is a 1km grounded wire a few kilometers away from survey area – this is dealt with using a special procedure. The data is measured on 5 field components at frequencies 16, 64 and 512 Hz at 28 stations spaced $50m$ apart on each of 11 lines with linespacing of $100m$. This gives a total of 308 data locations and 4620 data values. The "true model" has two conductive, boxy bodies and one resistive box. This is used to generate "true data", which are then contaminated by Gaussian noise, 2% in amplitude and 2 degrees in phase. The $3350m \times 3000m \times 2000m$ volume is discretized into $64 \times 50 \times 30 = 96,000$ cells.

A simple continuation process in β is applied, where we start with a large β and decrease it until the data is deemed to be fitted sufficiently well. Results are accummulated in Table 1. The "misfit" reported here and later on is $\|Qu - b\|/\|b\|$, i.e., the relative l_2 norm of the predicted minus the observed data at the end of the continuation step. Also, "n-it" is the nonlinear iteration counter, "KKT-it" is the number of preconditioned QMR iterations required for each nonlinear iteration, "infeas"$= \|Au - q\|/\|q\|$, and "rel-grad" is the relative gradient error in ϕ of (4). Note that an *inexact* Newton-type method is used. The overall number of KKT iterations is rather reasonable here, and not one forward problem is solved accurately before the end of

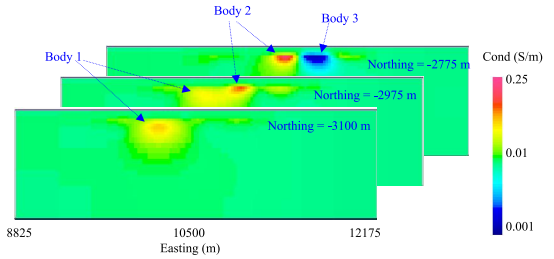

Figure 2: Three slices through the recovered 3D conductivity model obtained by inverting synthetic CSAMT data.

the entire continuation step. The reconstructed model is shown in Figure 2.

We have also inverted 3D electromagnetic data in time; see [9] for details and a full example.

4 Discontinuous solutions and "Huber's norm"

Recall that our argument leading to the minimization problem (4) has been that we introduce a priori information about smoothness of the model this way. But in many cases, including the examples of the previous section, our a priori knowledge is that the model probably contains jump discontinuities! So, in the regularization term

$$R(m) = \left[\int_\Omega \rho(|\nabla \mathrm{m}|) + \hat{\alpha}(\mathrm{m} - \mathrm{m_{ref}})^2\right]_h$$

(where the subscript h implies that the integral has been discretized) we want to limit the effect of penalty through a jump discontinuity in m, as this should not be penalized for non-smoothness.

Consider choosing a seminorm in $R(m)$ that will tolerate discontinuities [14, 17]. We have the multiresolution view of exploring functions on a given grid as corresponding to a discretization of some limit process.

Note that for $|\nabla \mathrm{m}| \to \infty$, $\int |\nabla \mathrm{m}|$ is integrable but $\int |\nabla \mathrm{m}|^2$ is not. This suggests inadequacy of the least squares norm in the presence of discontinuities. On the other hand, for $|\nabla \mathrm{m}| \to \mathbf{0}$, $\int |\nabla \mathrm{m}|$ yields problems when differentiating it to obtain necessary conditions, while $\int |\nabla \mathrm{m}|^2$ does not.

These observations suggest to combine the two, which yields the so-called Huber's norm [10, 14]

$$\begin{aligned} \rho(\tau) &= \begin{cases} \tau, & \tau \geq \gamma, \\ \tau^2/(2\gamma) + \gamma/2, & \tau < \gamma \end{cases} \qquad (10) \\ R'(\mathrm{m}) &\leftarrow \nabla \cdot \left(\min\{\frac{1}{\gamma}, \frac{1}{|\nabla \mathrm{m}|}\}\, \nabla \mathrm{m}\right). \end{aligned}$$

We have preferred to consider (10) over regularizing total variation, even though it's rougher and harder to analyze, because the parameter γ appears more naturally. Indeed, γ is the maximal size of a local change in $|\nabla \mathbf{m}|$ which is still interpreted as a smooth change in m on the scale of the current grid. This quantity may depend on the application.

We have found the following *automatic choice* to be particularly useful in practice:

$$\gamma = \frac{h}{|\Omega|_h}\left[\int_\Omega |\nabla \mathbf{m}|\right]_h. \qquad (11)$$

Thus, γ depends on the solution and is adjusted through the iteration in an obvious fashion. Others choose this parameter using an expression from robust statistics involving medians of $|\nabla \mathbf{m}|$: see [14] and references therein. Also, in the image processing literature one often chooses to penalize even less through discontinuities than when using Huber's norm (e.g. using a Gaussian function or a Tukey biweight). However, this leads to non-convex functionals and local minima, which seems excessive in our more complex context where the forward problem is not simply the identity.

To solve (7) or (9) we apply the method of *lagged diffusivity*, whereby in the current iterate we set

$$R''\delta m \approx \nabla \cdot \left(\min\{\frac{1}{\gamma}, \frac{1}{|\nabla \mathbf{m}|}\}\, \nabla \delta \mathbf{m}\right), \qquad (12)$$

and solve the resulting weighted least squares problem. This iteratively reweighted least squares (IRLS) algorithm has been analyzed for a fixed γ in the image denoising context to show global convergence; see [17] and references therein.

Difficulties do arise, of course, when the forward model is (1) or (2), where even the l_2 regularization functional can be testy. The corresponding algebraic equations when using (12) are harder to solve than for the l_2 case.

Recovering discontinuous surfaces from diffusive forward operators?

We have obtained good results using the techniques outlined above for applications in denoising and in SPECT tomography. However, in the context of the present paper one wonders, is there really enough accurate data in applications to allow an honest identification of discontinuities using our diffusive forward operators?! For instance, in Example 1 the recovered field approximates the unpolluted data well, whereas the quality of the reconstructed model is much poorer [9]. It would appear that many, rather different models for m could fit the data within the noise tolerance. Does it then make sense at all to seek a model with pinpointed discontinuities?

The answer depends on the quality of the a priori information available. Just fitting the data to within a hypothetical noise level, however, provides insufficient information to determine discontinuities responsibly, as the following examples demonstrate.

Example 2

This example is in $2D$, and $\Omega = [-1,1]^2$. To allow maximum chance for recovery of discontinuities we consider the differential operator (1) in the form

$$\nabla \cdot (\mathrm{m}^{-1}\nabla \mathrm{u}) = \mathrm{q},$$

i.e. without the exponential transformation from σ to m, with natural boundary conditions, and assume (unrealistically) that data on u are available everywhere on a given grid.

The right hand side is chosen with source and sink,

$$\begin{aligned} q &= \exp(-10((x+0.6)^2+(y+0.6)^2)) \\ &- \exp(-10((x-0.6)^2+(y-0.6)^2)), \end{aligned}$$

and the "true model", depicted in Figure 3, contains discontinuities. We use this true model to generate a field on a 129×129 cell-centered grid and contaminate this with 1% white noise to yield the "observed data", b.

A standard finite volume discretization is applied to the forward problem (with harmonic averaging for m^{-1}). For the optimization problem we use a conjugate gradient solver with a multigrid preconditioner [16] which employs operator-induced, node-based prolongation.

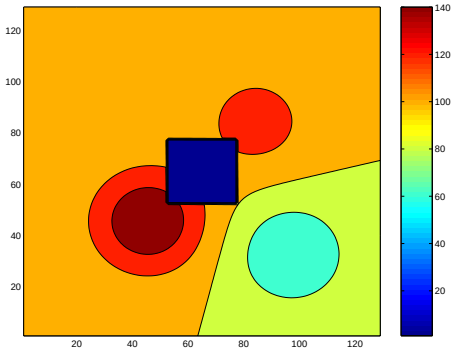

Figure 3: Contour plot of the "true model" for Example 2.

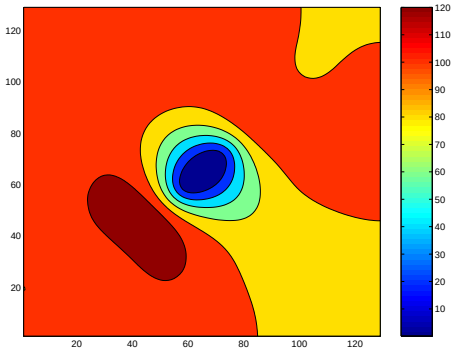

Figure 4: Recovered model using l_2 regularization of $|\nabla \mathrm{m}|$ with $\beta = 3 \times 10^{-6}$ for Example 2.

Figure 4 displays the recovered model using l_2 regularization with $\beta = 3 \times 10^{-6}$. The resulting misfit is 1.50×10^{-2}. Reducing β to 10^{-6} produces a model where noise-related artifacts are apparent, so β should not be reduced further.

In comparison, our result using Huber's norm with $\beta = 10^{-5}$, which yields a final $\gamma = 4.6$ and an almost ideal misfit 1.01×10^{-2}, is displayed in Figure 5.

Like for Example 1, the reconstructed fields, u, in all of these inversions are far smoother than the observed data and closer to the noiseless data (i.e. the exact discrete field).

The good news when comparing these figures is that the more careful Huber norm does yield bet-

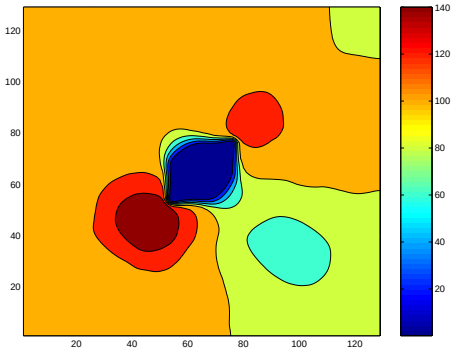

Figure 5: Recovered model using Huber's norm with $\beta = 10^{-5}$ for Example 2.

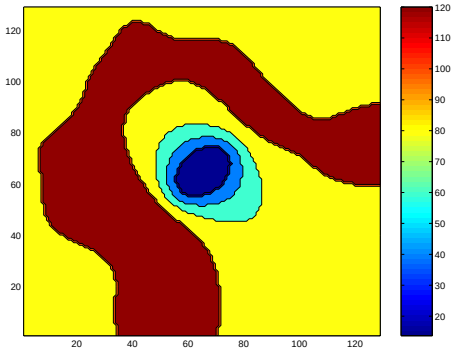

Figure 6: The model of Figure 4 replaced by a piecewise constant approximation with 5 constant values.

ter results, both in terms of misfit and in terms of closeness to the "true model".

On the other hand, the difference between a misfit of 1% and 1.5% is much too fine to define a cutting edge between a "good" and a "bad" model in realistic situations, where we do not know a "true model" either. In fact, in most practical cases the noise level is unknown and one uses other statistical techniques such as GCV [17] to evaluate it. Furthermore, the data are scarcely given everywhere - see the realistic Example 1.

Of course, it can be said that, using the a priori information that the model contains discontinuities, we generate one such model which fits the data well enough and that is that. However, there are other models with discontinuities which fit the data well! For the present example we took the recovered model of Figure 4 and applied a simple thresholding procedure, whereby the range $[m_{\min}, m_{\max}]$ was divided into 5 equal subintervals, and then all values of m within each such subinterval were replaced by the midpoint value. The result is displayed in Figure 6. Clearly, the resulting model is rather far from the "true model" of Figure 3, and yet the misfit is a respectable 2.2×10^{-2}. The point of the above thresholding experiment is that in more realistic examples we may not know whether our recovered discontinuities are, even approximately, in the right place.

Example 3

We use the same setting as in Example 2 with the following exceptions: (i) The grid size is a coarser 33×33; (ii) an exact solver is used for the linear equations (7); and most importantly, (iii) the "true model" only has two values and is given by the function

$$\mathrm{m}(x,y) = \begin{cases} 10 & -.2 < x, y < .2, \\ 10 & x^2 + y^2 > .8^2, \\ 1 & \text{otherwise} \end{cases} .$$

Thus, we have a ring of height 1 with a circular outer boundary and a square inner boundary immersed in a background of height 10. However, unlike with a typical level set method, we do not assume such additional, specific knowledge of this particular model. The result using Huber's norm with $\beta = 10^{-6}$ and $\gamma = 2.2$ is displayed in Figure 7. It is rather far from the "true" ring model (in particular, the cross diagonal symmetry is strongly violated), even though the misfit is an almost ideal 1.02×10^{-2}. Repeating the same experiment with a smaller $\gamma = .34$ yields a misfit of 8.7×10^{-3}, but the quality of the reconstructed model is not improved.

From these two examples it is clear that simply trusting the reconstruction because the misfit is small cannot be advocated. (Considering the maximum norm of the predicted minus the observed fields proves insufficient as well.) It can then be argued that displaying a smooth blob, such as when using least squares regularization (Figures 2 and 4) is less committing than displaying a discontinuous solution (especially with only a few constant values), and as such is more commensurate with the actual information at hand.

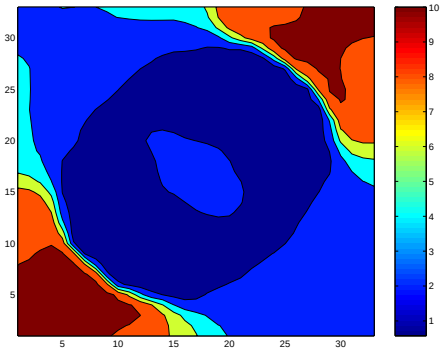

Figure 7: The recovered model for Example 3 using Huber's norm with $\beta = 10^{-6}$.

References

[1] U. Ascher and E. Haber. Grid refinement and scaling for distributed parameter estimation problems. *Inverse Problems*, 17:571–590, 2001.

[2] U. Ascher and E. Haber. A multigrid method for distributed parameter estimation problems. *J. ETNA*, 18:1–18, 2003.

[3] G. Biros and O. Ghattas. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization I: the Krylov-schur solver. *SIAM J. Scient. Comput.*, 2004. To appear.

[4] G. Biros and O. Ghattas. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization II: the Lagrange-Newton solver, and its application to optimal control of steady viscous flows. *SIAM J. Scient. Comput.*, 2004. To appear.

[5] L. Borcea, J. G. Berryman, and G. C. Papanicolaou. High-contrast impedance tomography. *Inverse Problems*, 12, 1996.

[6] A. J. Devaney. The limited-view problem in diffraction tomography. *Inverse Problems*, 5:510–523, 1989.

[7] 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.

[8] E. Haber and U. Ascher. Preconditioned all-at-one methods for large, sparse parameter estimation problems. *Inverse Problems*, 17:1847–1864, 2001.

[9] E. Haber, U. Ascher, and D. Oldenburg. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 2004. To appear.

[10] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Stats.*, 35:73–101, 1964.

[11] G. Newman and D. Alumbaugh. Three-dimensional massively parallel electromagnetic inversion–I. theory. *Geophysical journal international*, 128:345–354, 1997.

[12] J. Nocedal and S. Wright. *Numerical Optimization*. New York: Springer, 1999.

[13] R. L. Parker. *Geophysical Inverse Theory*. Princeton University Press, Princeton NJ, 1994.

[14] G. Sapiro. *Geometric Partial Differential Equations and Image Analysis*. Cambridge, 2001.

[15] 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.

[16] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, 2001.

[17] C. Vogel. *Computational methods for inverse problem*. SIAM, Philadelphia, 2002.