

On effective methods for implicit piecewise smooth surface recovery

U. M. Ascher* E. Haber† H. Huang‡

September 22, 2005

Abstract

This paper considers the problem of reconstructing a piecewise smooth model function from given, measured data. The data are compared to a field which is given as a possibly nonlinear function of the model. A regularization functional is added which incorporates the a priori knowledge that the model function is piecewise smooth and may contain jump discontinuities. Regularization operators related to total variation (TV) are therefore employed.

Two popular methods are modified TV and Huber's function. Both contain a parameter which must be selected. The Huber variant provides a more natural approach for selecting its parameter, and we use this to propose a scheme for both methods. Our selected parameter depends both on the resolution and on the model average roughness; thus, it is determined adaptively. Its variation from one iteration to the next yields additional information about the progress of the regularization process.

The modified TV operator has a smoother generating function; nonetheless we obtain a Huber variant with comparable, and occasionally better, performance.

For large problems (e.g. high resolution) the resulting reconstruction algorithms can be tediously slow. We propose two mechanisms to improve efficiency. The first is a multilevel continuation approach aimed mainly at obtaining a cheap yet good estimate for the regularization parameter and the solution. The second is a special multigrid preconditioner for the conjugate gradient algorithm used to solve the linearized systems encountered in the procedures for recovering the model function.

*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).

‡Institute of Applied Mathematics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada (hzhzhiyan@math.ubc.ca). Supported in part under NSERC Research Grant 84306.

Keywords: Inverse problem, Total variation, Huber regularization, Lagged diffusivity, Gauss-Newton, Multilevel continuation, multigrid preconditioner

1 Introduction

We consider the problem of reconstructing a piecewise smooth model function in several dimensions from given, measured data. In general, the data are viewed as a nonlinear function of the model:

$$b = F(m) + \epsilon,$$

where F is the forward modeling operator, $m = m(\mathbf{x})$ is the model to be recovered, b is the data, and ϵ is measurement noise. We assume that the model is defined on a 2D domain with unit volume, Ω , although extensions to 3D are possible and in principle straightforward.

The operators F under consideration vary considerably from an identity in simple denoising problems [33, 36, 34] through a linear, constant coefficient form in deblurring and tomography problems [36, 22, 21, 14] to the solution of a diffusive partial differential equation (PDE) system involving Poisson or Maxwell's differential equations [9, 27, 4, 20]. Due to either ill-posedness of the problem or the presence of noise in the measured data (or both) the problem requires regularization. In a Tikhonov-type approach this leads to the optimization problem

$$\min_m T(m) \equiv \frac{1}{2} \|F(m) - b\|^2 + \beta R(m), \quad (1)$$

where $R(m)$ is the regularization operator which is the focus of attention in the present work, and $\beta \geq 0$ is the regularization parameter. We use the least squares norm for the data fitting term, although other norms are possible and may be more suitable depending on the noise statistics.

For simplicity of exposition we employ the following notational convention. The domain Ω is discretized by a uniform grid with N^2 square cells of length $h = 1/N$ each. With a slight abuse of notation the function $m(\mathbf{x})$ becomes an $(N+1) \times (N+1)$ grid function m which is occasionally further reshaped into a vector, $m \in \mathfrak{R}^K$, $K = (N+1)^2$. Likewise, the data b , given on a subset of the grid's nodes or cells, is occasionally rearranged as a vector, $b \in \mathfrak{R}^M$. Then $F : \mathfrak{R}^K \rightarrow \mathfrak{R}^M$ is a vector function, assumed continuously differentiable. And yet, we shall also denote the discretized gradient of m , say, simply as ∇m . Finally, we define the *sensitivity matrix*

$$J(m) = \frac{\partial F}{\partial m}.$$

Then the necessary conditions for optimum in (1) are written as a system of generally nonlinear algebraic equations,

$$J^T (F - b) + \beta R_m = 0. \quad (2)$$

For the regularization term, one typically considers a same-grid discretization of

$$R(m) = \int_{\Omega} \rho(|\nabla(m - m_{ref})|) + \hat{\alpha}(m - m_{ref})^2 \quad (3)$$

where $\hat{\alpha} \geq 0$ is a small parameter and m_{ref} is a given reference function. The selection of m_{ref} can be crucial, for instance in geophysical exploration, but here we set $\nabla m_{ref} = 0$ 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. Note that

$$R_m(m) = -\nabla \cdot (\sigma(|\nabla m|)\nabla m), \quad \text{where } \sigma(\tau) = \frac{\rho'(\tau)}{\tau}.$$

For recovering smooth surfaces it is common to choose least squares, $\rho(\tau) = \frac{1}{2}\tau^2$, yielding $\sigma \equiv 1$. However, this choice is inadequate in the presence of discontinuities. The famous total variation (TV) choice (e.g. [10, 31, 36, 16]) uses the l_1 -norm instead, $\rho(\tau) = \tau$, implying $\sigma(\tau) = \frac{1}{\tau}$. Unfortunately, $\sigma(\tau) = 1/\tau$ is unbounded when τ shrinks, i.e., when $|\nabla m| \rightarrow 0$. Three remedies have been proposed in the literature to alleviate this. The first, and most recent, is to view $|\nabla m|$ as dual variables in a primal-dual approach. See [5] and references therein. This leads to some beautiful, more complex mathematical analysis. However, the other two possibilities to be explored below are so simple, *general*, popular and effective in practice that the case for their replacement is necessarily not crucial. Moreover, the adaptive variants proposed in this article often yield better results than the pure TV would.

The *modified TV* approach replaces $|\nabla m|$ in ρ and therefore in σ by

$$|\nabla m|_{\varepsilon} = \sqrt{m_x^2 + m_y^2 + \varepsilon^2}. \quad (4)$$

See, e.g., [8, 7, 36]. The parameter $\varepsilon \geq 0$ must be selected to be small enough so as to sharply preserve discontinuities and large enough so as not to cause undue numerical hardship, especially when iterative methods such as preconditioned conjugate gradients are used. We have found surprisingly little in the literature regarding the practical selection of this parameter.

Experiment 1

To see that there is an issue here, consider the popular example of denoising the picture given in the upper right panel of Figure 1 using a suitably modified TV regularization [31, 8]. For our grid function $m = \{m_{i,j}\}_{i,j=0}^N$ ($N = 256$, $h = 1/N$), denoting

$$D_{+,x}m_{i,j} = \frac{m_{i+1,j} - m_{i,j}}{h}$$

and similarly for $D_{+,y}m_{i,j}$, we define $|\nabla m_{i,j}|_h$, etc., by

$$|\nabla m|_{\varepsilon} = |\nabla m|_{h,\varepsilon} = \sqrt{(D_{+,x}m)^2 + (D_{+,y}m)^2 + \varepsilon^2}.$$

This yields a symmetric term in the sum replacing the integral for $R(m)$ in (3) which appears in (1). We next vary ε and record work units and CPU times: the actual method used is described later, in Sections 3 and 4; here we merely note that an increase in computing time relates directly to an increase in the total number of preconditioner calls in our procedure for solving the resulting linear systems.

| | | | | |
|---------------|-----------|-----------|----|------|
| ε | 10^{-6} | 10^{-2} | 4 | 3000 |
| Work units | 37 | 30 | 12 | 10 |
| CPU sec | 127 | 115 | 50 | 42 |

Table 1: Work units and CPU time vs the modification parameter ε .

Table 1 clearly indicates that more computational effort is required for smaller values of ε . Next, the results displayed in Figure 1 indicate that the choice $\varepsilon = 3000$ is not very good, especially for segmentation purposes; the other three values yield very similar, and satisfactory, results.

The combination of quality and cost therefore indicates not only that (i) as is well-known, values of ε which are too large or too small should be avoided, but also that (ii) the best value among those displayed, $\varepsilon = 4$, does not automatically leap to mind! It should be noted that the performance indicators in Table 1 do not change significantly in the range $4 \leq \varepsilon \leq 100$. Thus, the quest is not to find an optimal value of ε , but rather, to select it from “the right range”. This parameter range, however, need not necessarily include values $\varepsilon \ll 1$ which are commonly used by practitioners. Unless one is content with a highly expensive trial and error approach, a procedure is required that would yield an effective, automatic selection of ε .



An alternative to modified TV uses the Huber switching function [23, 16, 33]

$$\rho(\tau) = \begin{cases} \tau, & \tau \geq \gamma, \\ \tau^2/(2\gamma) + \gamma/2, & \tau < \gamma \end{cases} \quad (5)$$

$$R_m(m) \leftarrow \nabla \cdot \left(\min\left\{\frac{1}{\gamma}, \frac{1}{|\nabla m|}\right\} \nabla m \right).$$

Unlike for the smooth modified TV function, here ρ is not C^2 , although its second derivative is bounded. Nonetheless we show in Section 2 and Appendix A that the lagged diffusivity algorithm converges globally under the same assumptions as for the modified TV, extending the proof of [8]. The advantage of the choice (5) is that the selection of the parameter γ is much more transparent, and we capitalize on this in the present article.

Indeed, γ in (5) is viewed as answering the question, “what can be interpreted as a jump?” Thus, it is the minimal size of a local change in $|\nabla m|$ which is still interpreted by a TV functional and hence jumps will not be necessarily smeared. The

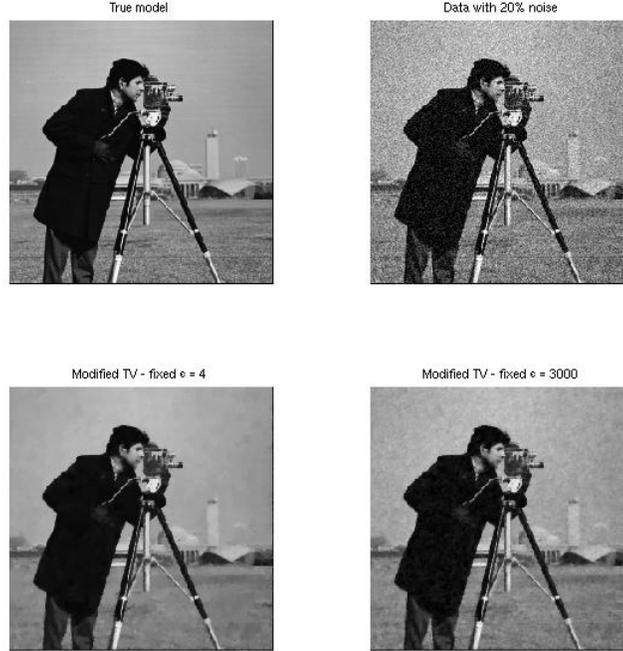


Figure 1: Denoising: The “true image” (upper left) is corrupted by adding 20% white noise (upper right) and the latter is fed as data b to a denoising program using the usual Tikhonov regularization with modified TV. The smaller value of switching parameter ε (lower left) produces a more pleasing result than the larger value (lower right). Taking $\varepsilon = .01$ or $\varepsilon = 10^{-6}$ yields results similar to those displayed in the lower left panel but at a higher cost.

TV interpretation is desirable, particularly if edges are to be respected, but it must be avoided in an $O(h)$ band, where the resolution limitation does not allow a true distinction between what may be zero and what is not.

With this in mind, we propose the following *automatic, adaptive choice*:

$$\gamma = h \int_{\Omega} |\nabla m|. \quad (6)$$

The integral is of course replaced by a sum upon discretization. If the volume of Ω differs from unity then we divide by it to obtain the average of $|\nabla m|$. In (6), $\gamma = \gamma(m)$ depends on the solution, and indeed we adjust its value through the iteration in an obvious fashion. What this choice basically says is that we avoid the possibility of a jump *only when* $|\nabla m|$ *is below* h *in the relative sense*; otherwise, a jump *may* happen.

Note that the average value of $|\nabla m|$ may well depend on the application, its scal-

ing, and the amount of noise still present in a given iteration. Moreover, the rate of change in γ from one iteration to the next indicates to what extent the discontinuities have already been distinguished from the smooth parts of the surface m to be recovered.

Our choice of γ also leads to a selection of the parameter ε in (4), as elaborated upon in Section 2.

Another, earlier choice uses so-called “robust statistics theory” where γ is interpreted as answering the question, how large the model’s gradient can be before it is considered to be an outlier. One arrives at the formula (see, e.g., [33], p. 231, and references therein)

$$\gamma = c \cdot \text{median}[|\nabla m - \text{median}(|\nabla m|)|], \quad (7)$$

with $c \approx 1.5$. However, it is not entirely clear under what conditions the statistical arguments leading to (7) are valid, nor is it clear how this formula scales with h . In fact, the rationale leading to (7), unlike ours in deriving (6), does not favour TV whenever possible, in recognition of the tendency of TV regularization to flatten natural scenes.

While the selection of γ and ε is settled, that of the regularization parameter β , which is notoriously challenging in practical problems even for least squares regularization [36, 15], does not become easier here. Several systems must be solved in a search for a good β , each involving difficult nonlinear optimization and the solution of often ill-conditioned linear systems. For large problems (corresponding to high resolution) the resulting reconstruction algorithms can become tediously slow.

In Section 3 we propose two mechanisms to improve efficiency. The first is a multilevel continuation approach [1]. While applying continuation also in a procedure to grow the model solution gradually this technique is aimed mainly at obtaining a cheap yet good estimate for the regularization parameter β . The key observation is that β should not vary much between different grids with our choice of γ .

The second item in Section 3 is a special multigrid preconditioner for the conjugate gradient algorithm used to solve the linearized systems encountered in the procedures employed for recovering the model function. Because of jumps in the coefficient function σ a usual multigrid cycle can be ineffective [37].

We close this paper with numerical experiments demonstrating the proposed techniques on several inverse problems. These techniques work very well for problems such as denoising, deblurring and emission tomography. But for a problem such as Electrical Impedance Tomography (EIT) [4] or electromagnetic geophysical prospecting [20] the reconstruction of a piecewise continuous model is fundamentally difficult. A Conclusions section seals the paper.

2 Modified TV and Huber regularization

Let us consider first the Huber regularization (5) and the choice of the parameter γ given by (6). It is not difficult to see that, despite the appearance of the term $1/\gamma$ in (5), the regularization term $R(m)$ remains bounded as $h \rightarrow 0$. Indeed, on the continuous level fix γ and let

$$\Phi = \{\mathbf{x} : |\nabla m(\mathbf{x})| < \gamma\}. \quad (8a)$$

Then (1) is written as

$$\min_m \frac{1}{2} \|F(m) - b\|^2 + \beta \left[\frac{1}{2\gamma} \int_{\Phi} |\nabla m|^2 + \frac{\gamma}{2} \int_{\Phi} 1 + \int_{\Omega \setminus \Phi} |\nabla m| \right]. \quad (8b)$$

But by (8a),

$$\frac{1}{2\gamma} \int_{\Phi} |\nabla m|^2 + \frac{\gamma}{2} \int_{\Phi} 1 < \gamma \int_{\Phi} 1,$$

and also $\int_{\Phi} |\nabla m| < \gamma \int_{\Phi} 1$. Hence, letting $\gamma \rightarrow 0$ we get $R(m) \rightarrow \int_{\Omega} |\nabla m|$.

Now, the regularization term on a grid with meshwidth h approximates the continuous $R(m)$ up to $O(h)$ by straightforward consistency. We summarize this as:

Lemma 1

With γ given by (6) the regularization term satisfies

$$R(m) = \int_{\Omega} |\nabla m| + O(h). \quad (9)$$

◆

Next, consider the modified TV regularization (4). When $\varepsilon \gg |\nabla m|$ we obviously get approximately the least squares norm on $|\nabla m|$. Thus, the considerations for selecting ε should follow similar lines to those for γ . We choose

$$\varepsilon = 0.5\gamma, \quad (10)$$

which yields agreement with (5) at $\tau = 0$, where these formulae are most useful if considered as modifications of TV. It is important to note that while both forms of stabilizing the TV seminorm are roughly equivalent the choice of the stabilization parameter is natural in the Huber case but is less obvious for the modified TV. If edges are not to be emphasized then the choice of a larger γ again has a more obvious meaning in the Huber case.

Note that as $h \rightarrow 0$, $|\nabla m|_{h,\varepsilon} \rightarrow |\nabla m|$ in a well-defined manner.

2.1 Lagged diffusivity

Upon differentiating (1) with respect to m we obtain the following system of equations,

$$\beta[L(m)m]_{i,j} + [J^T(m)(F(m) - b)]_{i,j} = 0, \quad \text{where} \quad (11a)$$

$$(L(m)m)_{i,j} = \frac{1}{h^2}[\sigma_{i+1/2,j}(m_{i,j} - m_{i+1,j}) + \sigma_{i,j+1/2}(m_{i,j} - m_{i,j+1}) + \sigma_{i-1/2,j}(m_{i,j} - m_{i-1,j}) + \sigma_{i,j-1/2}(m_{i,j} - m_{i,j-1})], \quad (11b)$$

$$\sigma_{i+1/2,j} = \min\left\{\frac{1}{\gamma}, \frac{h}{\sqrt{(m_{i,j} - m_{i+1,j})^2 + (m_{i,j} - m_{i,j+1})^2}}\right\}. \quad (11c)$$

Here, h is the step size of a uniform grid and $i, j = 1, \dots, N - 1$. The five point discretization of the regularization term above can be thought of as a consistent finite volume discretization of the differential operator $-\beta \nabla \cdot \left(\min\left\{\frac{1}{\gamma}, \frac{1}{|\nabla m|}\right\} \nabla m \right)$ with natural boundary conditions.

In order to solve the nonlinear system (11) numerically, a fixed point iteration called *lagged diffusivity* has been proposed and studied extensively (see, e.g., [36] and references therein). Given an initial iterate m^0 , one repeatedly and approximately solves

$$\beta L(m^{s-1})m^s + J^T(m^s)(F(m^s) - b) = 0, \quad s = 1, 2, \dots \quad (12)$$

Thus, if F is linear then for each s we solve a linear, anisotropic diffusion equation. This method works very well for low accuracy requirements (i.e., it converges rapidly at first but then slows down as its asymptotic convergence rate is only linear), especially if $J^T J$ is nonsingular.

Global convergence, i.e. from any starting point m^0 , for the modified TV norm was shown in several references in the case that $J^T J$ is constant and bounded away from singularity; see [11] and references in Chapter 8 of [36], where it is also claimed that some proofs easily extend to Huber's function with γ fixed.

Here, for the sake of completeness, we extend the proof by Chan and Mulet [8] to the case where $F(m) = Jm$, J constant with a full column rank, using Huber's function (5) with γ adaptively defined by (6).

Theorem 2

Assume that $F(m) = Jm$, where J is a constant matrix with a full column rank in (1),(2). Let $\{m^s\}$ be the sequence of iterates obtained by solving the discrete version of (12). Then the following holds:

1. $T(m^s) \leq T(m^{s-1})$ for all integers $s \geq 1$.
2. $\lim_{s \rightarrow \infty} |m^s - m^{s-1}| = 0$.
3. The sequence $\{m^s\}$ converges to the unique global minimum.

The proof is given in Appendix A.

Another possibility proposed by Chan, Golub and Mulet [7] and advocated in [36] is a primal-dual Newton method: dual variables corresponding to the flux (in (11) these are the discretized $\mathbf{v} = \min\{\frac{1}{\gamma}, \frac{1}{|\nabla m|}\} \nabla m$) are considered as independent variables along with m and a Newton method is defined, observing maximum size constraints for \mathbf{v} . However, the typical number of iterations required using lagged diffusivity to reach the rough error tolerances which makes sense in the present context is so small that the improved convergence rate of the primal-dual method may kick in too late to be of practical interest.

Lagged diffusivity and damped Gauss-Newton

Returning to lagged diffusivity, in case that $F(m)$ is nonlinear it makes sense to also linearize the second term in (12) in a comparable manner. In particular, it makes sense to use a single Gauss-Newton iteration, linearizing $F(m^s)$ around m^{s-1} , which leads to the following linear system to be solved at each iteration:

$$(J^T J + \beta L) \delta m = J^T (b - F(m^{s-1})) - \beta L m^{s-1}, \quad (13)$$

with $J = J(m^{s-1})$, and $L = L(m^{s-1})$ defined in (11). In general global convergence is no longer guaranteed by simply setting $m^s = m^{s-1} + \delta m$, and a line search (i.e., setting $m^s = m^{s-1} + \eta \delta m$, $0 < \eta \leq 1$; see for instance [29]) or a trust region method may be necessary to guarantee sufficient reduction of the objective function $T(m)$ of (1).

2.2 On choosing the regularization parameter

In general, the optimization problem (1) must be solved many times by the process described above for different regularization parameter values β . This contributes in a major way to the total cost of the solution process. A solution is accepted when some stopping criterion is satisfied. This is not our focus in the present article; however, leading to the next section we mention the simplest such criterion, namely, the discrepancy principle. Thus, one requires

$$\phi_d(m(\beta)) = \|F(m) - b\|^2 \leq \text{tol} \quad (14)$$

for some specified tolerance level tol which depends on the noise level and is assumed given. A procedure of continuation in β is then developed, where previous values of ϕ_d are used to estimate the next value of β , e.g. by a secant iteration.

In practice the tolerance is derived from statistical arguments, such as χ^2 distribution, and it must be noted that one rarely has more than a rough estimate of what tol should be. This is true even for the case of a smooth model recovery.

3 Numerical techniques for large problems

Multilevel continuation

Lemma 1 assures us that $R(m)$ does not vary much as the resolution is increased on the same domain. The same holds true, of course, for the data fitting term in (1). We therefore conclude that *with γ selected by (6) the ideal parameter β varies little on all sufficiently fine grids!*

As in [1, 18] we develop a multilevel continuation methodology, aimed chiefly at exporting the value of β constructed on a coarse grid to finer grids.

Starting on the coarse grid and subsampling the data as needed, the algorithm calculates an approximate solution and an approximate regularization parameter. Next, we interpolate the solution to a finer grid using a piecewise linear interpolation (cf. [1]) and solve the optimization problem using the interpolated solution as a starting guess, with the regularization parameter obtained on the coarse grid possibly slightly adjusted. The process continues until we reach the finest grid.

The reason that we may need to adjust β as the grid is refined is that the discretization error may play a more major role on coarser grids, particularly when the forward problem also involves the inversion of a discretized differential equation. (That “the coarsest grid must be fine enough” is a well-known requirement in the multigrid literature; see [1] in our particular context.) Thus, a search for β commences as in Section 2.2, so that a criterion such as (14) with a “soft tolerance” is respected before work starts on the next finer grid.

A multigrid preconditioner

Although multilevel continuation decreases the number of iterations needed to compute a solution on the finer grids, each iteration is still expensive because (13) is a large system of linear equations which corresponds to a differential or integro-differential equation with non-smooth coefficients. Therefore, an essential component in any fast algorithm for the solution of the problem is a fast algorithm for the solution of the linear system.

To simplify notation we rewrite the system as

$$H \delta m \equiv (J^T J + \beta L) \delta m = \text{rhs}. \quad (15)$$

There are two cases to consider here. First, if $J^T J$ is sparse with a sparsity pattern similar to that of L (e.g. denoising, super-resolution) then a preconditioned conjugate gradient (PCG) method with a multigrid cycle as a preconditioner can be directly applied to the system. If on the other hand $J^T J$ is dense then one can approximate $J^T J$ either by a sparse matrix, for example $\text{diag}\{J^T J\}$, or by a low rank matrix (see [18]) and obtain an approximate $\hat{J}^T \hat{J} + \beta L$ which is used as a preconditioner. We now briefly discuss how to obtain such an approximation. Obtaining an efficient

approximation to J is a non-trivial matter, see for example [17, 36, 3], which is beyond the scope of the present paper.

We therefore apply a multigrid preconditioner based on the approximate inversion of L plus a positive definite sparse matrix. Note that since L is a discretization of a differential operator with non-smooth coefficients standard multigrid methods are ineffective. One can resort to algebraic multigrid (AMG) (see, e.g., [35]); however, this may not be as efficient as geometric multigrid which utilizes the underlining grid that exists for our problem. We therefore turn to geometric multigrid methods which are developed for problems with non-smooth coefficients [24, 25].

Let m_F be a grid function on the finest grid of $(N + 1)^2$ points, or nodes. We set the next coarser grid function, m_C , to be the grid function on $(N/2 + 1)^2$ points chosen in the obvious manner. The difference from standard multigrid is in the use of *operator-induced interpolation*. This approach combines algebraic multigrid with geometric multigrid methods to obtain the interpolation operator. As common to algebraic multigrid methods we divide the fine grid points to coarse points (C) and fine points (F). The approximate Gauss-Newton step is reordered as

$$\begin{pmatrix} H_{FF} & H_{FC} \\ H_{FC}^T & H_{CC} \end{pmatrix} \begin{pmatrix} \delta m_F \\ \delta m_C \end{pmatrix} = \text{rhs.}$$

We next define the *exact interpolation*

$$P = \begin{pmatrix} H_{FF}^{-1} H_{FC} \\ -I \end{pmatrix}$$

and, eliminating the fine grid points, obtain the *exact coarse grid equations*

$$P^T \begin{pmatrix} H_{FF} & H_{FC} \\ H_{FC}^T & H_{CC} \end{pmatrix} P \delta m_C = \text{rhs}_C.$$

The obvious problem with the exact interpolation and coarse grid operators is that the matrix H_{FF} does not have a sparse inverse. Different approximations for its inverse define different methods. For a survey and comparison of methods see [25]. Here we have chosen a similar approach to the one suggested by Dendy [24] to obtain sparse interpolation and a coarse grid operators. The restriction operator is chosen to be the transpose of the interpolation and symmetric Gauss-Seidel is used as a smoother. Combining the above components into a V(3,1) cycle we have found that in most cases a single cycle is sufficient to reduce the residual norm below 10^{-2} .

4 Numerical experiments

Experiment 1 (cont.)

For simple denoising, $F(m) = m$ and hence $J = I$, the identity. We apply PCG iterations using our multigrid V-cycle preconditioner. The coarsest grid in all cases has $h_{\text{coarsest}} = 2^{-4}$.

For the example considered in Section 1, $h = 2^{-8}$. Let us complete its details. The quality of the results strongly depends on β and some search was required before settling on $\beta = .07$ for a 20% noise. Using the Huber function with γ determined adaptively according to the formulas (6) and (7) we obtain after 10 lagged diffusivity iterations results which are very similar, indeed indistinguishable to the naked eye, to those displayed in the lower left and right panels, respectively, of Figure 1. The choice of 10 lagged diffusivity iterations has been made because that is when γ in (6) changes by less than 10% from one iteration to the next, but further iterations have not revealed significantly different results.

The respective final values of γ are 8.34 and 6.56×10^3 , and this gives rise to the values of ε used in the experiment reported in Section 1. There has been no notable difference in the performance of the Huber vs the modified TV variants in this case, and the results reported earlier therefore demonstrate both the general utility of the formula (6) and its superiority over (7). The (non-optimized) CPU times reported in Table 1 correspond directly to the total number of PCG iterations required, and the latter count, for 10 nonlinear iterations in each case, is reported as 'work units' in Table 1. The V-cycle preconditioner is therefore seen to perform rather effectively as well in our algorithm. The additional cost of carrying out the evaluation of the adaptive γ , by either formula (6) or (7), is negligible. In summary, then, our proposed algorithm performs very well indeed for this denoising problem.

Here is another denoising experiment. The "clean" data are synthesized from MATLAB's supplied function 'peaks' modified as follows: *if $|m_{i,j}| > 0.01$ then $m_{i,j} = m_{i,j} + d \text{ sign}(m_{i,j})$* . We set $d = 10$. Adding 10% ($nlevel = 0.1$) Gaussian noise and setting $\beta = .06$ we run for different values of h and record in Table 2 the values of γ for the final m in each run. The value of γ is determined by (6). The linear dependence of γ on h clearly emerges. For the relative tolerance value of 10^{-2} on

| | | | | | |
|----------|----------|----------|----------|----------|----------|
| h | 2^{-4} | 2^{-5} | 2^{-6} | 2^{-7} | 2^{-8} |
| γ | 4.05 | 2.03 | 1.04 | 0.518 | 0.255 |

Table 2: γ by (6) vs mesh width h .

both the lagged diffusivity iteration and the PCG iteration, three lagged diffusivity iterations are required in all cases for both Huber and modified TV. A total of three PCG iterations (i.e., one for each lagged diffusivity iteration) are needed for modified TV while an additional PCG iteration is required for the Huber function. A typical result is depicted in Figure 2.

Note also that the same value of β has yielded quality results on all 5 grids, satisfying $\|F(m) - b\| < 2\|m\|nlevel$. On the other hand, when the coarsest grid was set coarser than $h = 2^{-4}$ it proved useless for the multilevel process.

A multilevel continuation process as described in Section 3 has been set up as

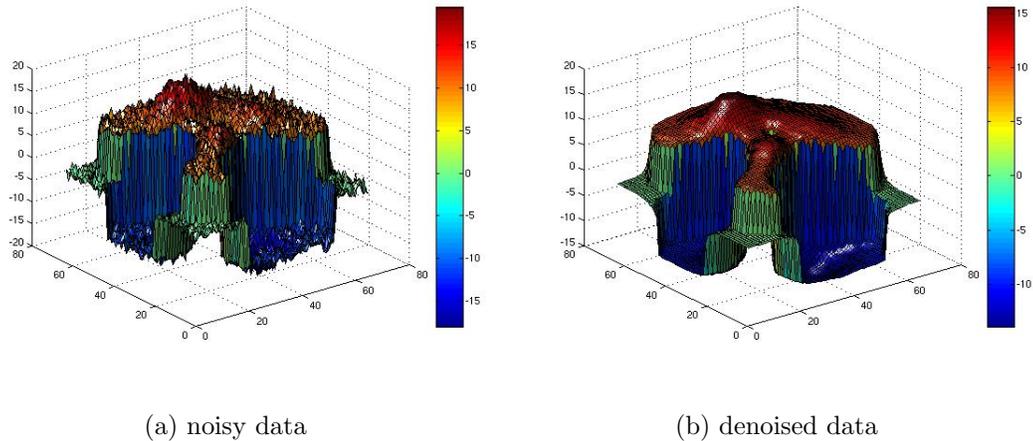


Figure 2: Denoising a surface with discontinuity.

well. Bilinear interpolation from each grid to the next finer one proves more efficient than a similar piecewise constant interpolation. It cuts the iteration count on the finest grid to just one iteration and the overall elapsed time by a factor of more than two.

Repeating the above experiments with the lower noise level of 1% yields a faster convergence on the finest grid starting from the data (and requiring one or at most two iterations for convergence), so the multilevel continuation process is not useful. However, exporting the value of $\beta = .006$ determined on a coarse grid with the same $h_{coarsest}$ as before to finer grids does prove to work well here, too.

We have also made comparative runs choosing γ by the formula (7). The resulting value is significantly larger than γ of (6), especially on fine grids. Thus, the least squares norm is used more often in the regularization term R , typically resulting in a possibly better approximation in smooth regions but in less sharp edges as well. An example of this sort is obtained here upon setting $d = 2$ in the above model. To make the point about the discontinuity smearing clearer to visualize, though, we plot in Figure 3 the results of a reconstruction of a piecewise constant surface. \blacklozenge

Experiment 2

Next we consider the limited angle Single Photon Emission Tomography problem, SPECT. A brief description of the problem follows; see [21] for more detail. A chemical tracer is injected into a patient. This chemical emits photons in all directions. The chemical is targeted to bond to a specific functional area in the body (for example it attaches itself to active muscles). As a result, a few minutes after the injection, some active parts of the body "glow" while others do not. In order to find these parts, the patient is put under a collimated camera. The camera can detect the glow only from

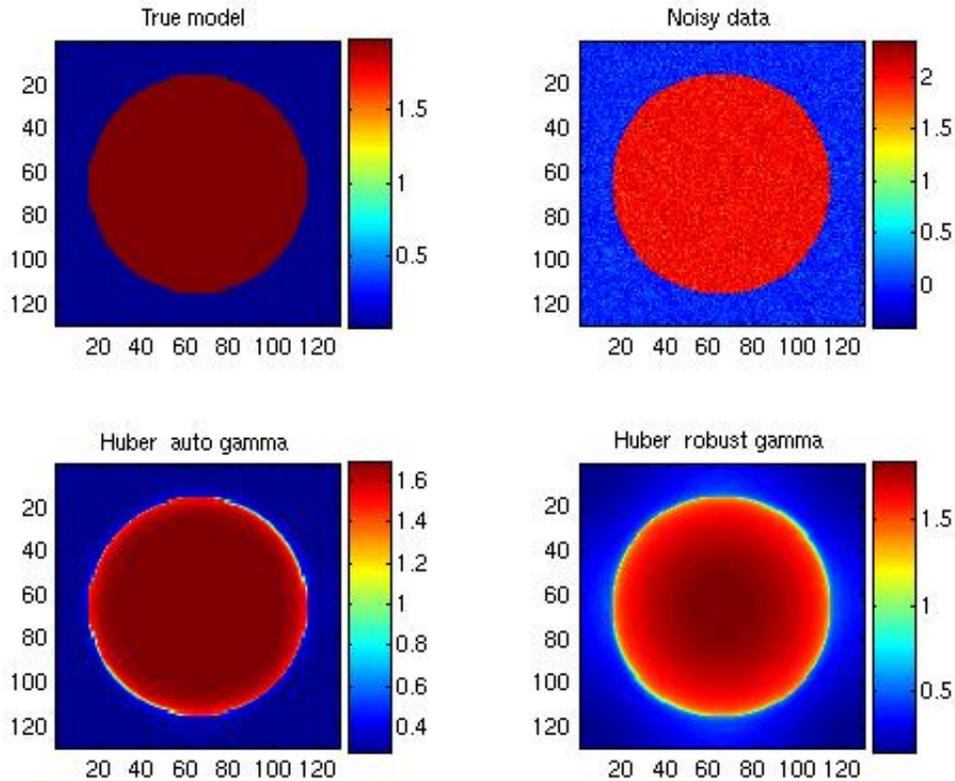


Figure 3: Denoising a piecewise constant surface: The true model takes the value 2 inside a circular pie subdomain (brown) and 0 in the background (blue). The run parameters are $nlevel = .1$, $h = 2^{-7}$, $\beta = .06$. The 'robust' $\gamma = 2.83$ determined by (7) is much larger than the 'auto' $\gamma = .0276$ determined by (6). Consequently, the discontinuity is more smeared in the lower right plot than in the lower left one.

a single direction. Mathematically, the number of photons that are detected at each collimator are

$$b_i = \int_{\text{ray}_i} m(x, y) dx dy. \quad (16)$$

The set of all the rays at the same direction forms a projection of the object onto this direction. The goal is to find m based on the projection data. In order to discretize the line integral (16) the model is divided into cells. We then trace each of the rays and obtain a matrix J . The entries of the matrix J_{ij} are the length of the i^{th} ray in the j^{th} cell. Since each ray usually passes through only a few cells, J is typically sparse.

Here, the forward problem is linear,

$$F(m) = Jm.$$

The sensitivity matrix J is constant and has full row rank, but the problem is typically underdetermined and thus $J^T J$ is singular. For instance, one has to solve the problem (2) given by

$$J^T Jm + \beta R_m = J^T b,$$

where J is 2048×4225 , for a 65×65 grid function m . So, we must have $\beta > 0$. Indeed, β may neither be too large nor too small.

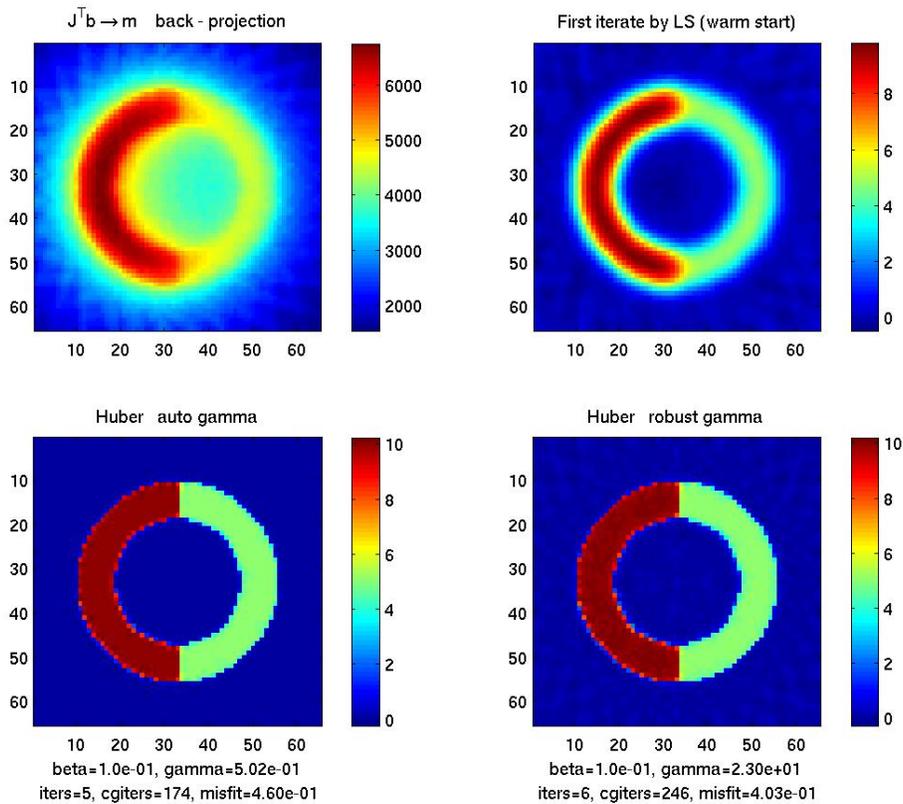


Figure 4: SPECT: noiseless data synthesized from two halves of a circular ring. Here $\beta = .1$. The back-projection image appears unfocussed and the solution obtained by least squares regularization is useful as a starting iterate. The image obtained using Huber's function with (6) is excellent (lower left). The 'robust' γ (7) gives somewhat poorer results.

We have used a piecewise constant 'true model' m_{true} which is well approximated in the lower left panel of Figure 4 to synthesize data b . At first we add no noise, so $b = Jm_{true}$. Results are pictured in Figure 4. Using the simple back-projection

method, $m = J^T b$, which is common in practice, leads to a blurred, out of scale image. We then set $\beta = 0.1$ and, keeping the same iteration tolerance values as in the previous experiment, apply our regularization techniques. (A significantly smaller β has led to an increased number of iterations due to the increased ill-conditioning of the linear system.) Using a least squares regularization ($\gamma = \infty$) yields blurred results which nonetheless prove useful as a starting iterate for a TV regularization. The Huber and modified TV regularizations produce excellent results with γ determined by (6). The TV regularization yields identical looking results and costs one additional work unit. The formula (7) for γ produces somewhat poorer results here; the difference can be more easily observed when reading an electronic copy of this article.

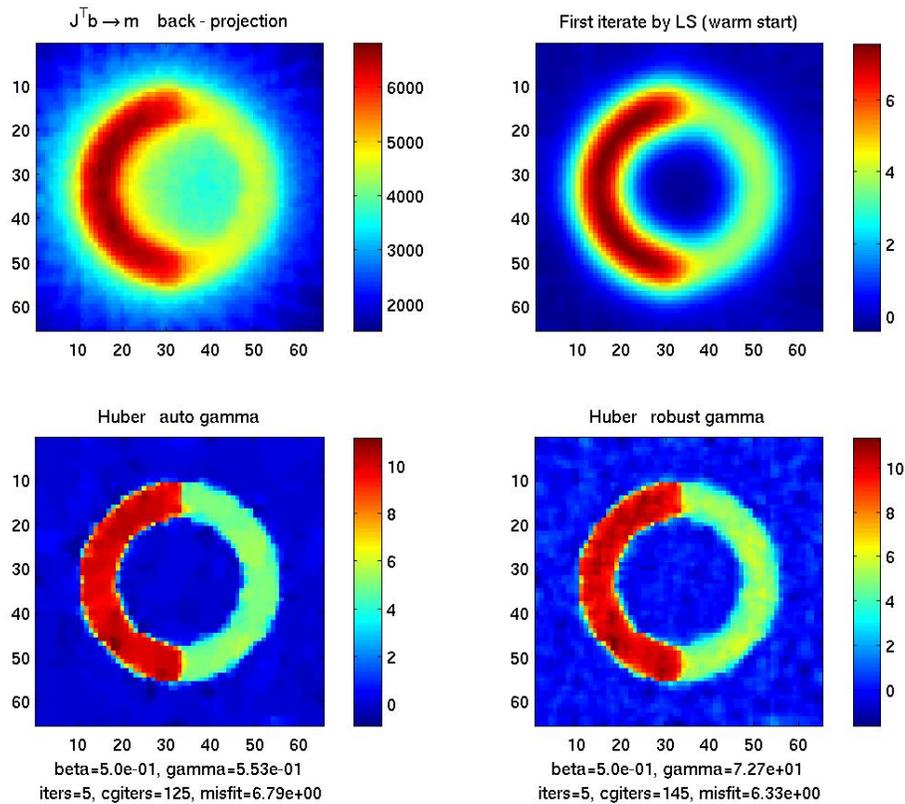


Figure 5: SPECT: data synthesized from two halves of a ring with 10% noise added on top. Here $\beta = .5$. The back-projection image appears unfocused and the solution obtained by least squares regularization is useful as a starting iterate. The image obtained using Huber’s function with (6) is excellent (lower left). The ‘robust’ γ gives inferior results.

Next we add 10% normally distributed random noise to the data, set $\beta = .5$ and repeat the calculations. The results are gathered in Figure 5: the observations are similar to those in the noiseless case, except that now the performance of the “robust

γ'' is more clearly worse.

In Figure 6 we record the history of a continuation process with the same β . The total cost is only a fraction less than the one-grid process described above. The relative efficiency of multilevel continuation improves for larger values of β , though. Note how the dependence of γ of (6), not only on h but also on the smoothness of the surface, is expressed here. \blacklozenge

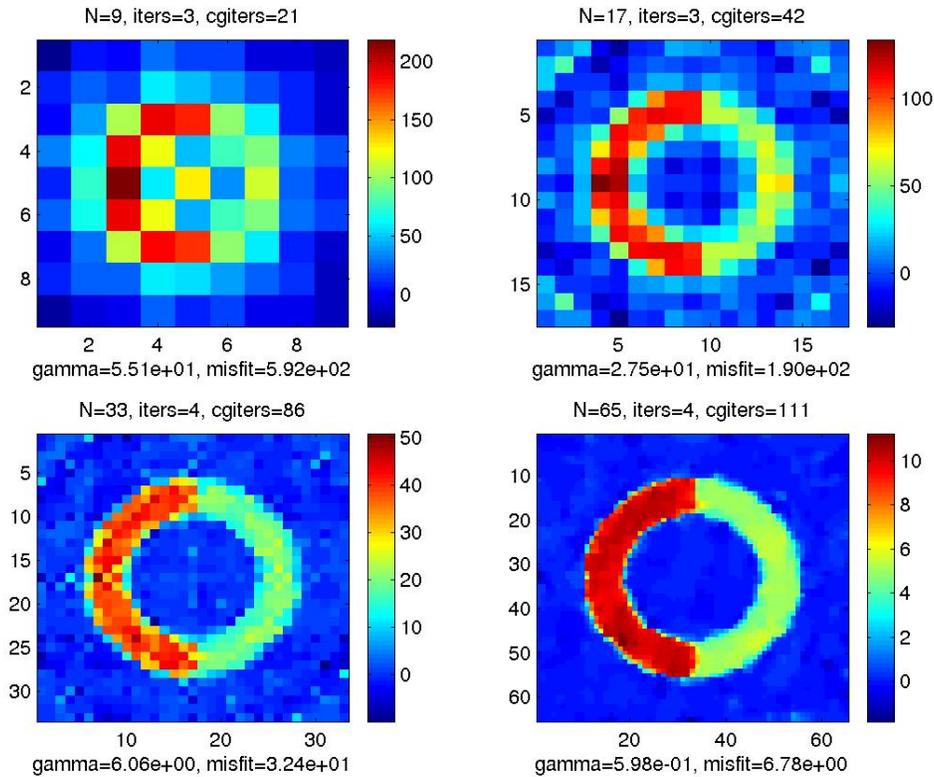


Figure 6: SPECT: data synthesized from two halves of a ring with 10% noise added on top. Here continuation with $\beta = .5$ is used. Note that γ get smaller faster than h because the surface becomes smoother.

We have also experimented with standard deblurring problems (e.g., [36, 22]) and found rather satisfactory results - and very close performance indicators - using either the modified TV or Huber's function with the switching parameter of (6). We do wish to repeat a general note of caution, though, that here too the TV regularization shines especially when recovering piecewise constant surfaces, although it does well also for a combination of rolling hills and steep canyons.

Experiment 3

The third experiment is conducted on problems (1) with $F(m) = Qu$, where u is the solution of the forward problem

$$A(m)u = q \tag{17}$$

and Q indicates where in the domain the field u is being measured. Thus, $F(m) = QA(m)^{-1}q$. The matrix A discretizes an elliptic operator such as $\nabla \cdot (\sigma(\mathbf{x})\nabla)$, where $\sigma(\mathbf{x}) > 0$ is bounded away from 0, or Maxwell’s equations in the frequency domain for a low frequency, and q are given sources. Both A and Q are typically large and sparse; see, e.g. [9, 4, 28, 19, 20, 2, 4, 6].

The discretized inverse problem in this case is much more ill-conditioned than in the previous problem instances. See [20] for realistic examples where the recovered field u approximates the unpolluted data b well while the quality of the reconstructed model is significantly poorer. Thus, one wonders in the context of piecewise continuous surfaces whether there is enough accurate data in applications to allow for an honest identification of discontinuities using such diffusive forward operators.

As a representative example we have considered the resistivity problem

$$\nabla \cdot (m^{-1}\nabla u) = q \tag{18}$$

with natural boundary conditions on a unit domain in $2D$. The right hand side is chosen with source and sink,

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

and the “true model”, depicted in Figure 7(a), contains discontinuities as well as smooth, nonconstant parts. 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} .

Figure 7(b) displays the recovered model using Huber’s switching with $\beta = 10^{-5}$, which yields a final $\gamma = 4.6$ and an almost ideal misfit 1.01×10^{-2} . This is an excellent result.

However, the difference between a misfit of 1% and 2% is much too fine to define a cutting edge between a “good” and a “bad” model in more realistic situations, where we do not know the “true model” either. In fact, in most practical cases the noise level is unknown and one uses other statistical techniques such as GCV (e.g. [36]) to evaluate it. As it turns out one can vary the discontinuity locations significantly and still retain a misfit within 2%; see [2]. Unfortunately, therefore, in realistic experiments we may not know whether our recovered discontinuities are in the right place, not even approximately.

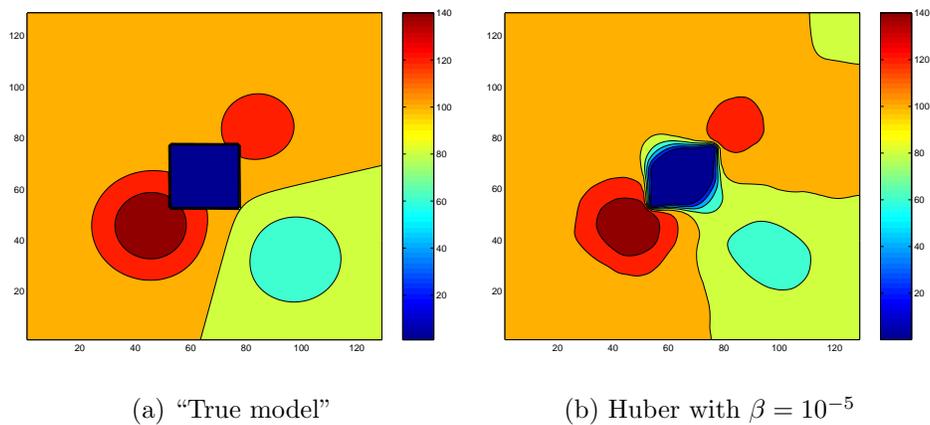


Figure 7: The “true model” and the recovered model for Example 1 of Experiment 3.

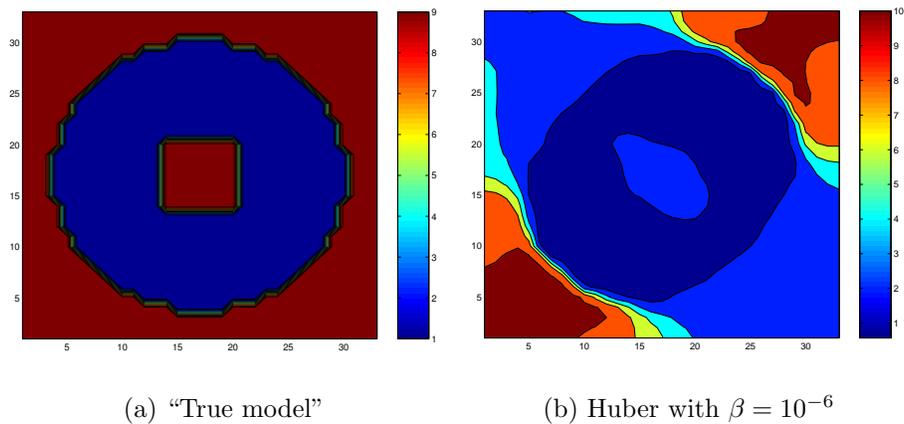


Figure 8: The “true model” and the recovered model for Example 2 of Experiment 3.

We then have considered a “true model” which only has two values and is given by the function

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

See Figure 8(a). The same setting as before is employed except that an exact solver is used for the linear equations encountered during the solution process and the grid size is 33×33 .

The result using the Huber regularization with $\beta = 10^{-6}$ yielding $\gamma = 2.2$ is displayed in Figure 8(b). 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 “sufficiently small” cannot be advocated. Considering the maximum norm of the predicted minus the observed fields proves insufficient as well. It can then even be argued that displaying a smooth blob, such as obtained when using the least squares regularization with β small, 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.



5 Conclusions and further comments

A new formula for selecting the Huber switching parameter γ , having selected β before, is proposed. This formula *adaptively “decides”* what must be smooth and what can be a jump. The question of the automatic selection of γ (or ε for modified TV) is usually ignored among TV users, although it is important and has been considered by others who proposed a criterion based on robust statistics. However, we have demonstrated that our adaptive choice (6) is better, particularly for picking up edges.

Once γ has been determined, the constant ε in modified TV, if one insists on using this variant, is selected as $\varepsilon = 0.5\gamma$. We have shown this to be a good choice. The varying nature of either parameter provides further important information on the progress of the iteration process, and this together with the possibility of choosing larger values γ yield a method with capabilities beyond pure TV regularization.

A global convergence result for lagged diffusivity, similar to that known for modified TV, has been proved for the Huber variant. Moreover, with the above choice of ε numerical performance of the two variants has proved very close in many cases tried. This gives further reinforcement for our proposed formula.

Multilevel continuation has been proposed by keeping β fixed when refining the grid and by using bilinear interpolation of m for the next first iterate. The advantage

in not having to calculate β from scratch on the fine grid can be huge.

Encouraging results are obtained for both the denoising and the SPECT problems. However, for diffusive forward problems the use of Huber or modified TV has tended to be unreliable. Note that in such problems it is possible to vary the discontinuity interface location significantly without violating the discrepancy principle for a given noise level.

Recent work has been carried out using level sets to recover the model under the assumption (which we do not utilize with the Huber and modified TV regularizations) that it is piecewise constant with known constant values (thus, it is a shape optimization problem [32, 26, 18, 12]). The stronger assumptions allow such methods to recover surfaces in situations where the Huber and TV regularizations fail, especially when the data are prescribed (more realistically) only in a ring next to the boundary, possibly in multiple sets. If the true model happens not to conform to these strong model assumptions, however, then there is no natural mode of recovery from determining a wrong model in this situation, especially when the data are sparse.

Both formulae (6) and (7) may fail, as will usual approaches for determining β , when processing an image with highly differing noise characteristics in different regions. A more local approach is then required.

We remark that in the image processing literature one often chooses to penalize even less through discontinuities than when using TV, e.g. using a Gaussian function or a Tukey biweight; see, e.g., [33, 30, 13]. However, this leads to non-convex functionals and occasionally also to several local minima, which seems excessive in our more complex context where the forward problem operator is not simply the identity. We do not consider these more adventurous choices further here.

Appendix A

In this appendix we prove Theorem 2. To recall, we consider minimization of the functional $T(m)$ given by (1) where $F(m) = Jm$, J is a constant matrix with a full column rank. Set matrices $A_{ij}^T = \nabla_{i,j} \in \mathbb{R}^{N \times 2}$, $1 \leq i, j \leq N$, so that the necessary conditions (2) can be written as

$$T_m(m) = \beta \sum_{i,j=1}^N \left(\frac{A_{ij}^T A_{ij}^T m}{|A_{ij}^T m|_\gamma} \right) + J^T (Jm - b) = 0$$

where $|A_{ij}^T m|_\gamma = \max\{\gamma, |A_{ij}^T m|\}$ > 0 . Then we have the following for T :

Lemma 3

The functional T is coercive and strictly convex.

Proof: Since $T(m) \geq \frac{1}{2}\|Jm - b\|^2$, T is coercive and bounded below by 0. Since T is twice *Fréchet* differentiable, we have for its Hessian

$$T_{mm} = \beta \sum_{i,j=1}^N A_{ij} \left(\frac{1}{|A_{ij}^T m|_\gamma} (I_2 - \frac{A_{ij}^T m \otimes A_{ij}^T m}{|A_{ij}^T m|_\gamma^2}) + \rho''(|A_{ij}^T m|) \frac{A_{ij}^T m \otimes A_{ij}^T m}{|A_{ij}^T m|_\gamma^2} \right) A_{ij}^T + J^T J$$

where ρ is defined by (5) and \otimes denotes the Kronecker product. By Cauchy-Schwarz, this yields

$$(A_{ij}^T m, A_{ij}^T v)^2 \leq (A_{ij}^T m, A_{ij}^T m)(A_{ij}^T v, A_{ij}^T v) \leq |A_{ij}^T m|_\gamma^2 |A_{ij}^T v|^2 \implies |A_{ij}^T v|^2 - \frac{(A_{ij}^T m, A_{ij}^T v)^2}{|A_{ij}^T m|_\gamma^2} \geq 0.$$

Moreover, we have $\rho''(\tau) \geq 0$ if $\tau \geq 0$. So, the Hessian of T satisfies, $\forall v \neq 0$,

$$T_{mm}(v, v) = \beta \sum_{i,j=1}^N \left(\frac{1}{|A_{ij}^T m|_\gamma} (|A_{ij}^T v|^2 - \frac{(A_{ij}^T m, A_{ij}^T v)^2}{|A_{ij}^T m|_\gamma^2}) + \rho''(|A_{ij}^T m|) \frac{(A_{ij}^T m, A_{ij}^T v)^2}{|A_{ij}^T m|_\gamma^2} \right) + v^T J^T J v > 0.$$

Hence T is a strictly convex functional. \blacklozenge

Thus problem (1) has a unique solution. We then define

$$G(v, m) = T(m) + (v - m, T_m(m)) + \frac{1}{2}((v - m), C(m)(v - m))$$

and follow [8] almost verbatim in recognizing that the lagged diffusivity algorithm is an instance of the generalized Weiszfeld method, so that the following lemma holds:

Lemma 4

1. $C(m) = \beta \sum_{i,j=1}^N (\frac{A_{ij} A_{ij}^T}{|A_{ij}^T m|_\gamma}) + J^T J$ is continuous.
2. $\lambda_{\min}(C(m)) \geq \mu = \lambda_{\min}(J^T J) > 0, \forall m$.
3. $T(v) \leq G(v, m), \forall v$.

\blacklozenge

Using Lemmas 3 and 4 we can now establish the global convergence for the Huber switching function with γ adaptively defined by (6), as stated in Theorem 2. We address each item in turn:

1. From the above properties, we have $T(m^s) \leq G(m^s, m^{s-1})$. Moreover, the fixed point iteration equation is equivalent to the necessary condition for the generalized Weiszfeld method, which yields

$$m^s = \underset{v}{\operatorname{argmin}} G(v, m^{s-1}) \implies G(m^s, m^{s-1}) \leq G(m^{s-1}, m^{s-1}) = T(m^{s-1}).$$

2. Recall that

$$\begin{aligned} T(m^s) &\leq G(m^s, m^{s-1}) \\ &= T(m^{s-1}) + (m^s - m^{s-1}, T_m(m^{s-1})) + \frac{1}{2}((m^s - m^{s-1}), C(m^{s-1})(m^s - m^{s-1})) \\ &= T(m^{s-1}) - \frac{1}{2}((m^s - m^{s-1}), C(m^{s-1})(m^s - m^{s-1})). \end{aligned}$$

Then taking into account that $\lambda_{\min}(C(m^{s-1})) \geq \mu > 0$, we have

$$\frac{1}{2}\mu|m^s - m^{s-1}|^2 \leq \frac{1}{2}((m^s - m^{s-1}), C(m^{s-1})(m^s - m^{s-1})) \leq T(m^{s-1}) - T(m^s).$$

It transpires that

$$0 \leq |m^s - m^{s-1}| \leq \sqrt{2\mu^{-1}(T(m^{s-1}) - T(m^s))}.$$

Note that T is bounded below and $\{T(m^s)\}$ is monotonically decreasing. Thus, $\{T(m^s)\}$ converges, implying that $\lim_{s \rightarrow \infty} |T(m^s) - T(m^{s-1})| = 0$, hence also

$$\lim_{s \rightarrow \infty} |m^s - m^{s-1}| = 0.$$

3. Since $\{T(m^s)\}$ is monotonically decreasing, it is bounded above of course. The fact that T is coercive then guarantees that $\{m^s\}$ is bounded. By compactness, it suffices to show that the limit of every convergent subsequence of $\{m^s\}$ is the global minimum of (1). Let $\{m^{s_i}\}$ be a subsequence of $\{m^s\}$ that converges to m^* . Since $G_v(v, m) = T_m(m) + C(m)(v - m)$, T is continuously differentiable and C is continuous, we have that G_v is continuous. From the above $\lim_{s_i \rightarrow \infty} |m^{s_i} - m^{s_i-1}| = 0$, and we also have $\lim_{s_i \rightarrow \infty} m^{s_i} = \lim_{s_i \rightarrow \infty} m^{s_i-1} = m^*$. Then, taking limits for $G_{m^{s_i}}(m^{s_i}, m^{s_i-1})$, we deduce the following equation:

$$\begin{aligned} 0 &= \lim_{s_i \rightarrow \infty} G_{m^{s_i}}(m^{s_i}, m^{s_i-1}) = G_{m^{s_i}}(\lim_{s_i \rightarrow \infty} m^{s_i}, \lim_{s_i \rightarrow \infty} m^{s_i-1}) \\ &= G_{m^{s_i}}(m^*, m^*) = T_m(m^*) + C(m^*)(m^* - m^*) = T_m(m^*). \end{aligned}$$

Hence the sequence $\{m^s\}$ converges to the unique global minimum m^* of our problem (1).

◆

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. Computational methods for large distributed parameter estimation problems with possible discontinuities. *Proc. Symp. Inverse Problems, Design & Optimization*, pages 201–208, 2004. M. Colaco, H. Orlande and G. Dulikravich (eds.).

- [3] G. Biros and O. Ghattas. Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization Parts I,II. *SIAM J. Scient. Comput.*, 2004. To appear.
- [4] L. Borcea, J. G. Berryman, and G. C. Papanicolaou. High-contrast impedance tomography. *Inverse Problems*, 12:835–858, 1996.
- [5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math Imaging and Vision*, 20:89–97, 2004.
- [6] T. Chan and X. Tai. Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM J. Scient. Comput.*, 25:881–904, 2003.
- [7] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Scient. Comput.*, 20:1964–1977, 1999.
- [8] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM J. Numer. Anal.*, 36:354–367, 1999.
- [9] M. Cheney, D. Isaacson, and J.C. Newell. Electrical impedance tomography. *SIAM Review*, 41:85–101, 1999.
- [10] J. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38:826–844, 1973.
- [11] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34:1779–1791, 1997.
- [12] O. Dorn, E.L. Miller, and C.M. Rappaport. A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Problems*, 16, 2000. 1119-1156.
- [13] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGraph*, pages 257–266. ACM, 2002.
- [14] I. Elbakri and J. Fessler. Statistical image reconstruction for polyenergetic x-ray computed tomography. *IEEE Trans. Med. Imag.*, 21(2):89–99, 2002.
- [15] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, 1996.
- [16] C. Farquharson and D. Oldenburg. Non-linear inversion using general measures of data misfit and model structure. *Geophysics J.*, 134:213–227, 1998.

- [17] E. Haber. *Numerical Strategies for the Solution of Inverse Problems*. PhD thesis, University of British Columbia, 1997.
- [18] E. Haber. A multilevel, level-set method for optimizing eigenvalues in shape design problems. *J. Comp. Phys.*, 198:518–534, 2004.
- [19] E. Haber and U. Ascher. Preconditioned all-at-one methods for large, sparse parameter estimation problems. *Inverse Problems*, 17:1847–1864, 2001.
- [20] 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*, 69:1216–1228, 2004.
- [21] E. Haber, D. Oldenburg, and A. Celler. Recovery of dynamic parameters in SPECT. *IEEE Trans. Nuc. Sci.*, 67:129–135, 1997.
- [22] P. C. Hansen. *Rank Deficient and Ill-Posed Problems*. SIAM, Philadelphia, 1998.
- [23] P. J. Huber. Robust estimation of a location parameter. *Ann. Math. Stats.*, 35:73–101, 1964.
- [24] J.E. Dendy Jr. Black box multigrid. *J. Comput. Phys.*, 48:366–386, 1982.
- [25] S. Knapek. Matrix-dependent multigrid homogenization for diffusion problems. *SIAM J. Sci. Comp.*, 20(2):515–533, 1999.
- [26] A. Leitao and O. Scherzer. On the relation between constraint regularization, level sets, and shape optimization. *Inverse Problems*, 19:L1–L11, 2003.
- [27] S. McCormick and J. Wade. Multigrid solution of a linearized, regularized least-squares problem in electrical impedance tomography. *Inverse problems*, 9:697–713, 1993.
- [28] G.A. Newman and D.L. Alumbaugh. Three-dimensional magnetotelluric inversion using nonlinear conjugate gradients. *Geophysical journal international*, 140:410–418, 2000.
- [29] J. Nocedal and S. Wright. *Numerical Optimization*. New York: Springer, 1999.
- [30] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [31] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [32] F. Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM Controle Optim. Calc. Var.*, 1:17–33, 1996.

- [33] G. Sapiro. *Geometric Partial Differential Equations and Image Analysis*. Cambridge, 2001.
- [34] O. Scherzer. Scale-space methods and regularization for denoising and inverse problems. *Advances in Image and Electron Physics*, 128:445–530, 2003.
- [35] U. Trottenberg, C. Oosterlee, and A. Schuller. *Multigrid*. Academic Press, 2001.
- [36] C. Vogel. *Computational methods for inverse problem*. SIAM, Philadelphia, 2002.
- [37] C.R. Vogel. Negative results for multilevel preconditioners in image deblurring. In *Proc. Scale Space*, 1999.