Electronic Transactions on Numerical Analysis.
Volume 6, pp. 255-270, December 1997.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

A COMPARISON OF MULTILEVEL METHODS FOR TOTAL VARIATION REGULARIZATION*

P.S. VASSILEVSKI† AND J.G. WADE‡

Abstract. We consider numerical methods for solving problems involving total variation (TV) regularization for semidefinite quadratic minimization problems $\min_u \|\mathcal{K}u - z\|_2^2$ arising from illposed inverse problems. Here $\mathcal{K}$ is a compact linear operator, and z is data containing inexact or partial information about the "true" u. TV regularization entails adding to the objective function a penalty term which is a scalar multiple of the total variation of u; this term formally appears as (a scalar times) the L^1 norm of the gradient of u. The advantage of this regularization is that it improves the conditioning of the optimization problem while *not penalizing discontinuities* in the reconstructed image. This approach has enjoyed significant success in image denoising and deblurring, laser interferometry, electrical tomography, and estimation of permeabilities in porus media flow models.

The Euler equation for the regularized objective functional is a quasilinear elliptic equation of the form $\left[\mathcal{K}^*\mathcal{K} + A(u)\right]u = -\mathcal{K}^*z$. Here, $A(u)$ is a standard self-adjoint second order elliptic operator in which the coefficient κ depends on u, by $[\kappa(u)](x) = 1/|\nabla u(x)|$. Following the literature, we approach the Euler equation by means of fixed point iterations, resulting in a sequence of linear subproblems.

In this paper we present results from numerical experiments in which we use the preconditioned conjugate gradient method on the linear subproblems, with various multilevel iterative methods used as preconditioners.

Key words. total variation, regularization, multilevel methods, inverse problems.

AMS subject classifications. 65N55, 35R30, 65F10.

1. Introduction.

We examine the numerical properties of various multilevel preconditioners for a class of quasilinear elliptic operators arising in total variation minimization. These operators typically have discontinuous and highly varying coefficients which may, for increasingly fine discretizations, have arbitrarily small margin of coercivity.

The outline of the paper is as follows. We first introduce the class of linear illposed inverse problems, which we formulate as minimization problems, and we give two examples. We provide a rather detailed motivation for the use of regularization methods in general and for total variation regularization in particular. Next, we exhibit the Euler equation (first-order necessary condition) for the minimization problem with total variation regularization. The Euler equation is a quasilinear integro-partial differential equation and, following the experience reported in the literature, we approach it computationally by means of fixed-point iteration. In this way we arrive at a sequence of linear operator problems of the form

$$\left(\mathcal{K}^*\mathcal{K} + \alpha A\right)w = b,$$

where $\mathcal{K}$ is a bounded (usually compact) operator which typically has properties similar to those of an integral operator, and A is a second order elliptic operator with rapidly varying and discontinuous coefficients. It is preconditioners for the elliptic part of this system which form the focus of this paper. In particular, we consider (after discretization) four preconditioners: the standard variational multigrid method, the hierarchical basis (HB) multilevel method, and the "approximate wavelet-modified" hierarchical basis (AWM-HB) method, and the AWM-HB method with a weighted L^2 norm. Our experience indicates that the standard multigrid method is the most effective of these for this problem. Next, we provide an example of the

* Received June 3, 1997. Accepted for publication October 29, 1997. Communicated by V. Henson.
† Center of Informatics and Computer Technology, Bulgarian Academy of Sciences, Sofia, Bulgaria. This work was conducted while the first author was a visiting scholar at Bowling Green State University, Bowling Green, Ohio.
‡ Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio. (gwade@math.bgsu.edu)

ETNA
Kent State University
etna@mcs.kent.edu

numerical solution of the inverse problem with this method. Finally, we summarize the paper and our conclusions and indicate possible directions for future work.

2. The Inverse Problem and Regularization.

2.1. Description of the Inverse Problem. Let Ω be the unit square in two dimensions. We asume we are given some linear operator $\mathcal{K}$ which is defined on a subset $U \subseteq L^1(\Omega)$ and has range in some Hilbert space Z. We assume that $\mathcal{K} : U \mapsto Z$ is compact. Also we assume we are given some "data" $z \in Z$ for which

$$z = \mathcal{K}\bar{u} + \delta \tag{2.1}$$

for some $\bar{u} \in U$ and some "error" δ.

The inverse problem is to determine $\bar{u}$, at least approximately. Since z may not lie in the range of $\mathcal{K}$, the inverse problem is formulated as a least-squares minimization, in which the goal is to minimize

$$\Phi(u) \stackrel{\text{def}}{=} \frac{1}{2}\|\mathcal{K}u - z\|_Z^2 \tag{2.2}$$

over some subset of $L^1(\Omega)$.

2.2. Discretization. All of the numerical approximations in this paper are based on discretizations involving finite element approximations with piecewise bilinear finite elements on a uniform mesh.

Specifically, let J be a fixed integer. For each integer k satisfying $1 \le k \le J$ we set $n_k = 2^k$ and $h_k = 2^{-k}$, and we partition the unit square into a collection $\mathcal{T}_k$ of n_k^2 uniform squares. The corners of these squares form the mesh whose node set we denote by $\mathcal{N}_k$. On this mesh we define the usual continuous piecewise bilinear elements whose span forms the finite element space which we denote by V_k.

Since we are considering multilevel methods, we shall have need of intergrid transfer operators. Since $V_{k-1} \subset V_k$, any element in V_{k-1} with nodal coefficient vector u_{k-1} can be represented exactly as an element of V_k by a unique nodal coefficient vector u_k. We take as I_{k-1}^k the matrix such that $u_k = I_{k-1}^k u_{k-1}$ for such u_{k-1} and u_k. Thus I_{k-1}^k is the matrix representation (with respect to the nodal bases of V_{k-1} and V_k) of the identity map from V_{k-1} to V_k. This is our "prolongation" operator. For the "restriction" operator, we take that operator whose matrix representation is given by the transpose, e.g.,

$$I_k^{k-1} \stackrel{\text{def}}{=} (I_{k-1}^k)^T. \tag{2.3}$$

2.3. Examples of $\mathcal{K}$.

2.3.1. Image deblurring. Here the operator $\mathcal{K}$ is a first kind Fredholm integral operator with translation invariance, e.g., a convolution. It is of the form

$$(\mathcal{K}u)(\vec{x}) = \int_\Omega k(\vec{x} - \vec{x}')u(\vec{x}')\, d\vec{x}',$$

where k is a Gaussian kernel of the form

$$k(\vec{\xi}) = \frac{1}{2\pi\sigma^2} \exp\Big(\frac{-|\xi|^2}{2\sigma^2}\Big)$$

for some $\sigma > 0$. The use of total variation regularization in conjunction with deconvolution or "deblurring" with this model for image processing has been investigated by a number of authors [1, 5, 9, 10, 11, 13, 21].

Matrix representations of $\mathcal{K}^*\mathcal{K}$ are generally dense. However, as discussed in [9], the action of $\mathcal{K}^*\mathcal{K}$ on a nodal representation of an FEM function may be carried out in $\mathcal{O}(n\log n)$ operations by the use of the FFT. In our numerical investigations we used Vogel's implementation [19] of this idea.

As a test pattern for the image reconstruction problem we use the piecewise constant function shown in Figure 2.1. It is given by $U_{TRUE} = \chi_{\Omega_1} + \chi_{\Omega_2} + \chi_{\Omega_3}$, where the $\Omega_i \subset \Omega$ are given by

$$\begin{aligned}\Omega_1(x,y) &= \Big\{(x-1/2)^2 + (y-1/2)^2) < 1/6^2\Big\},\\ \Omega_2(x,y) &= \Big\{1/5 < x < 4/5 \text{ and } 19/40) < y < 21/40\Big\},\\ \Omega_3(x,y) &= \Big\{9/10 < x+y < 11/10 \text{ and } 1/8 < x < 7/8 \text{ and } 1/8 < x < 7/8\Big\},\end{aligned}$$

and $\chi(\omega)$ denotes the characteristic function of subsets ω of Ω.

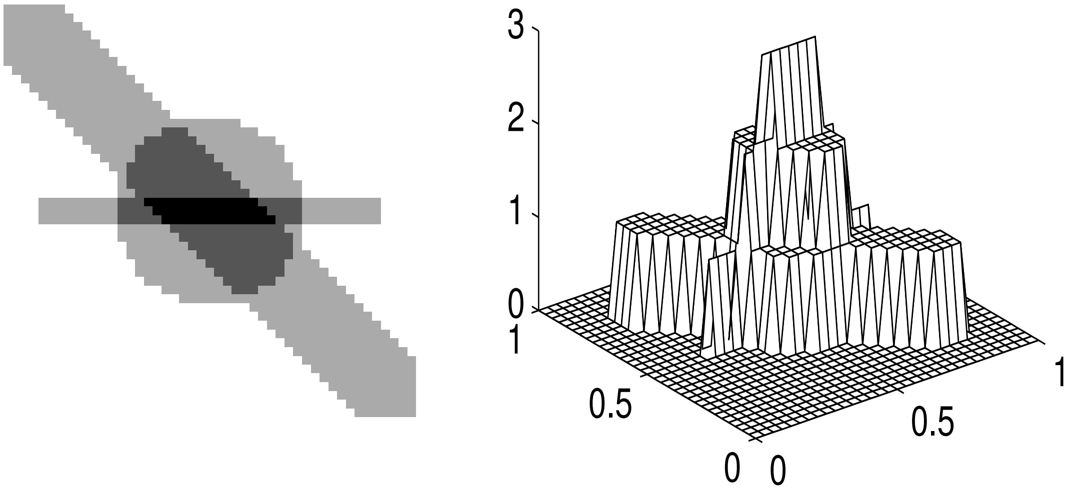

FIG. 2.1. *"Test pattern" U_{TRUE} used for image reconstruction problems. See §2.3.1.*

2.3.2. Electrical Impedance Tomography. As a second example we consider the linearized problem from electrical impedance tomography (EIT). In this case $Z = \Big[L^2(\partial\Omega)\Big]^m$ for some $m \geq 1$. Elements $z = \mathcal{K}u \in Z$ in the range of $\mathcal{K}$ are of the form $\{z_j\}_{j=1}^m$, where each z_j lies in $H^{1/2}(\partial\Omega) \subset L^2(\partial\Omega)$ and is given by

$$z_j(s) = \phi_j(s;u), \quad s \in \partial\Omega, \tag{2.4}$$

where $\phi_j(\cdot;u)$ is a potential function satisfying

$$\begin{aligned}\nabla\cdot\Big(\alpha(x)\nabla\phi_j(x;u)\Big) &= -\nabla\cdot\Big(u(x)\nabla\bar{\phi}_j(x;u)\Big), \quad \text{in } \Omega,\\ \Big(\alpha(x)\nabla\phi_j(x;u)\Big)\cdot\vec{n}(x) &= 0 \quad \text{on } \partial\Omega.\end{aligned} \tag{2.5}$$

Here, $\alpha \in L^\infty(\Omega)$ with $\inf\{\alpha(x) : x \in \Omega\} > 0$ and $\bar{\phi}_j \in H^1(\Omega)$ are fixed, given functions. For j and for each given $u \in L^\infty(\Omega)$ we obtain a z_j from this set of equations. This defines the map $\mathcal{K}$ for this example.

The interpretation of the operator $\mathcal{K}$ in this example is that is that it is the Fréchet derivative of the "conductivity to Dirichlet" map in the EIT problem. See [4, 22] for details. As demonstrated in [22], $\mathcal{K}$ is continuous as a linear map from $U \subset L^1(\Omega)$ into $H^{1/2}(\partial\Omega)$ if U is of the form

$$U = \{u \in L^1(\Omega) \ : \ \|u\|_{L^\infty(\Omega)} < C \ \text{ and } TV(u) \leq \gamma\},$$

where C and γ are constants and TV is the total variation functional discussed below in Section 2.5. The TV-regularization method discussed below implicitly ensures that $\mathcal{K}$ is restricted to such subsets.

2.4. Illposedness. Minimizers $\tilde{u}$ of the least-squares functional (2.2) must satisfy the normal equation

$$\mathcal{K}^*\mathcal{K}\tilde{u} = \mathcal{K}^* z. \tag{2.6}$$

However, the compactness of $\mathcal{K}$ implies that, unless $\mathcal{K}$ has finite dimensional range, the eigenvalues of the operator $(\mathcal{K}^*\mathcal{K})$ cluster at the origin so that $(\mathcal{K}^*\mathcal{K})^{-1}$ is unbounded. Hence $\tilde{u}$ generally will not exist as an element of $L^1(\Omega)$, so that the inverse problem is illposed.

Further insight into the nature of the illposedness is furnished by (attempted) numerical approximaton of the inverse problem. Specifically, if (2.6) is discretized as in Section 2.2 and $\tilde{u}_n$ are computed solutions of the these discrete problems, then $\tilde{u}_n$ will exhibit unwanted oscillations which increase in frequency and magnitude as $n \to \infty$.

We provide an example of this with the deconvolution example of Section 2.3.1, with a "synthetic data" set z and full matrix representations of $\mathcal{K}$ and $\mathcal{K}^*$ at various grid levels. To generate the synthetic data, we first formed a numerical representation of U_{TRUE} as described in §2.3.1 and given in Figure 2.1, on the level $k = 5$ grid (e.g., $2^5 \times 2^5$). From this we computed $\mathcal{K}U_{TRUE}$, and added 1% noise to it (that is, each each of the 33^2 grid points we added noise which was normally distributed with zero mean and standard deviation $1/100$) to obtain the data z. We then attempted to solve the inverse problem (2.6) with this data, using the psuedo-inverse of the (full) matrix representations of $\mathcal{K}$ on levels 3, 4 and 5. (To represent the synthetic data z on levels 3 and 4 we projected it using the restriction operator given in (2.3).)

The results are shown in Figure 2.2. The behavior illustrated there is typical of distributed parameter inverse problems. In order to (attempt to) capture the salient features, one must use a sufficiently fine grid; however, this leads to a highly oscillatory solution which is due to unboundedness of $(\mathcal{K}^*\mathcal{K})^{-1}$ as $n \to \infty$. These "spurious oscillations" means that the illposedness is a serious practical matter.

2.5. H^1 and total variation regularization. Strategies for approximately recovering $\bar{u}$ must account for the illposedness in some way. Generally, either the search for the solutions must be restricted to some subset of $L^1(\Omega)$ which is sufficiently constrained so as to avoid spurious oscillations in computed solutions, or, essentially equivalently, we must "regularize" the problem by modifying the objective functional (2.2) so as to supress the unwanted oscillations.

Below we adopt a regularization based on penalizing the *total variation* of candidate solutions $\tilde{u}$. Total variation regularization is an alternative to the better-known H^1 regularization, which is based on penalizing the square of the H^1 norm of candidate solutions. Both of these regularizations have the advantage that they implicitly limit the minimization of (2.2)

ETNA
Kent State University
etna@mcs.kent.edu

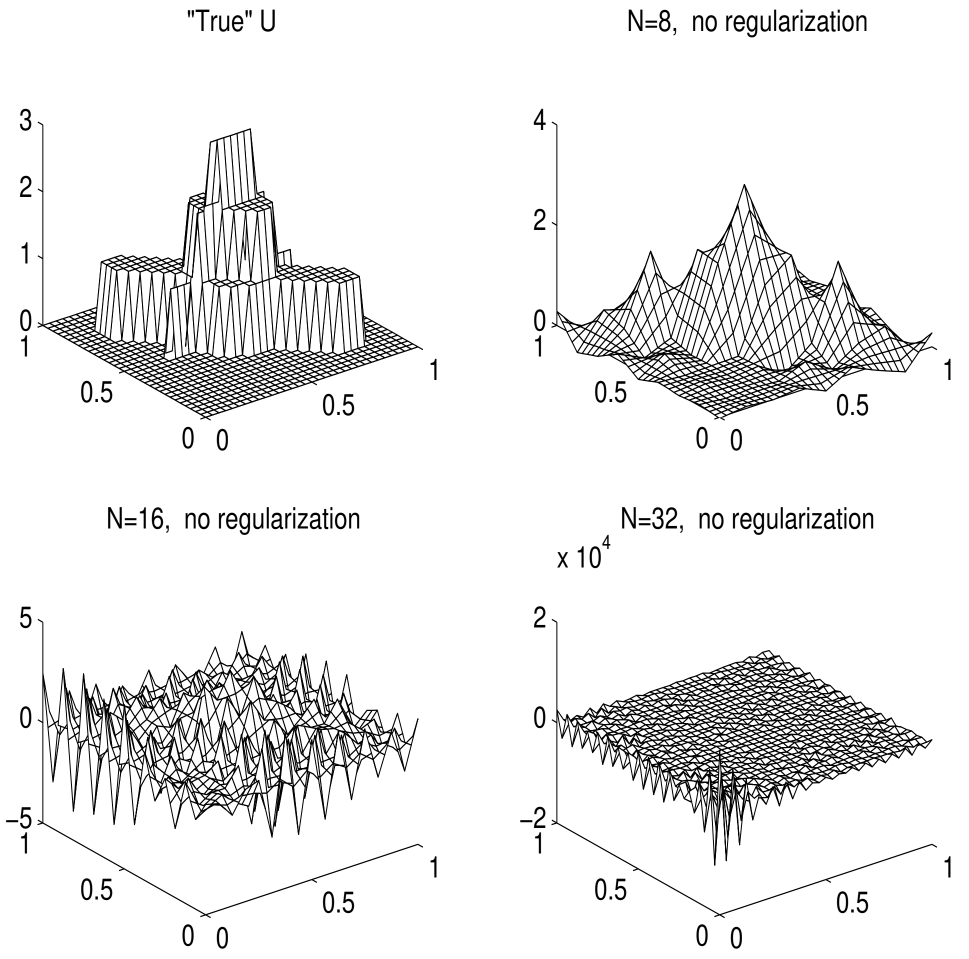

FIG. 2.2. *Results from the computations reported in Section 2.4, illustrating the illposedness of unregularized inverse problem. Note the difference in scale in the fourth subplot.*

to compact subsets of $L^1(\Omega)$ so that minimizers are guaranteed to exist. The H^1 regularization is mathematically and computationally more tractable because it yields a quadratic minimization problem; however, as illustrated in the examples below, it tends to oversmooth the solution. The key advantage of total variation regularization is that is it *permits discontinuities* in the computed solutions. However, as discussed below, it results in a nonquadratic optimization problem, so that the mathematical and numerical analysis are both more involved.

The total variation $TV(u)$ of $u \in L^1(\Omega)$ is defined [7] as

$$TV(u) \stackrel{\text{def}}{=} \sup\Big\{ \int_\Omega u(x) \nabla \cdot g(x)\, dx \;:\; g \in C_0^1(\Omega; \mathbf{R}^n), \|g\|_{L^\infty} = 1 \Big\}. \tag{2.7}$$

The regularized objective functional is then

$$\frac{1}{2}\|\mathcal{K}u - z\|^2_{L^2(\Omega)} + \alpha TV(u) \tag{2.8}$$

for some small parameter α. The advantage of this regularization is that it permits functions u with jump discontinuities, yet sets of the form $\{u \in L^1(\Omega) : TV(u) \leq \gamma\}$ are compact in $L^1(\Omega)$ [7]. Hence approximate reconstructions of u can be stably computed from (2.8) even when the "true" u is discontinuous.

For u in the Sobolev space $W^{1,1}(\Omega)$, the expression (2.7) becomes

$$TV(u) = \int_\Omega |\nabla u(x)| \, dx.$$

We shall use this throughout. Also, we use a modification, which is based on experience reported in the literature, e.g., [1, 3, 6, 20], and serves the purpose of making the TV functional TV differentiable, namely, for a fixed $\beta > 0$,

$$TV_\beta u \stackrel{\text{def}}{=} \int_\Omega \sqrt{|\nabla u(x)|^2 + \beta^2} \, dx. \tag{2.9}$$

Our regularized objective function is then, for given α and β,

$$\Phi^{\alpha,\beta}(u) \stackrel{\text{def}}{=} \frac{1}{2}\|\mathcal{K}u - z\|^2 + \alpha TV_\beta u. \tag{2.10}$$

Minimizers of (2.10) must satisfy the first order necessary condition (the Euler equation) for this functional. It is a quasilinear elliptic equation of the form

$$\mathcal{K}^*\mathcal{K}u + \alpha A(u)u = \mathcal{K}^* z \tag{2.11}$$

in Ω, subject to homogeneous Neumann boundary conditions. Here, for a given v, $A(v)$ is the self-adjoint second order elliptic operator whose action is given by

$$A(v)u = -\nabla \cdot \big(\kappa(v)\nabla u\big), \tag{2.12}$$

where the coefficient κ depends on v by

$$[\kappa(v)](x) = \frac{1}{\sqrt{|\nabla v(x)|^2 + \beta^2}}. \tag{2.13}$$

2.5.1. Example contrasting H^1 and TV regularizations. As noted above, the chief advantage of the TV regularization is that it allows discontinuities in the reconstruction $\tilde{u}$, whereas H^1 regularization oversmooths them. We illustrate this point with a numerical experiment.

On the level 5 mesh we computed U_{TRUE} and correspondingly noisy data z as in Section 2.4, as well as a full matrix representation of $\mathcal{K}$ and a sparse matrix discretizations of $A(U_{TRUE})$ based on (2.12). For various values of α in (2.11), we then computed the solution u by a direct method. In Figure 2.4, we present graphical results of this for three different values of α, one of which shows over-regularization, one under-regularization, and one which lies in between. We also performed the same calculations but with H^1 regularization, e.g., with $A(U_{TRUE})$ replaced by $A(1)$ which is (minus) the Laplacian scaled by $1/\beta$. These results are also shown in Figure 2.4.

3. Numerical solution of the minimization problem. As discussed above in §2.5, the Euler equation for the inverse problem with total variation regularization is given by (2.11), where A is the quasilinear elliptic operator given by (2.12). A straightforward computational

ETNA
Kent State University
etna@mcs.kent.edu

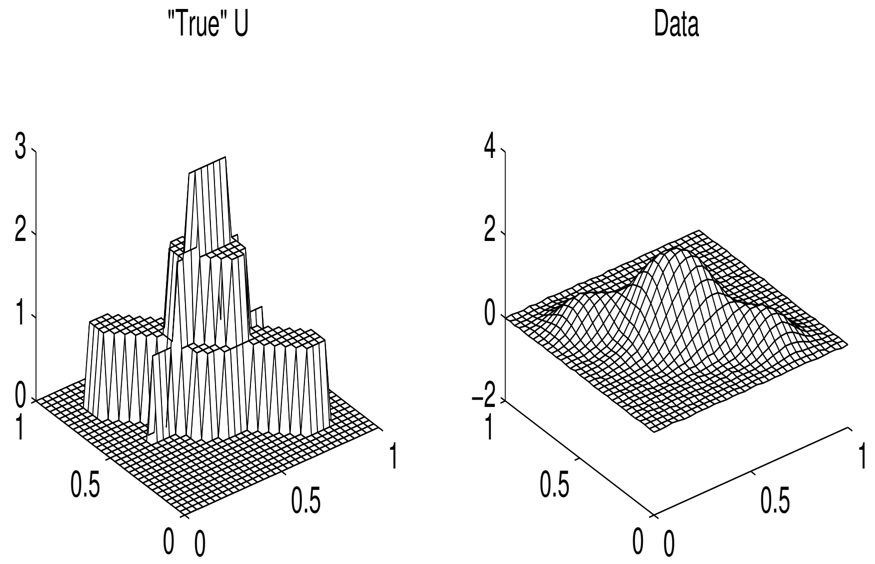

FIG. 2.3. *The "true" u and the "data" for the numerical examples discussed in Section 2.5.1, where the H^1 and TV regularizations are compared, and in Section 4.2, where numerical approximations of the solution of the inverse problem with TV regularization are presented.*

approach to solving (2.11) is fixed point iteration. That is, give some initial guess $u^{(0)}$, we compute $\{u^{(m)}\}$ by $u^{(m+1)} = u^{(m)} + \delta u^{(m)}$, where $\delta u^{(m)}$ satisfies

$$[\mathcal{K}^*\mathcal{K} + \alpha A(u^{(m)})]\delta u^{(m)} = \mathcal{K}^* z - [\mathcal{K}^*\mathcal{K} + \alpha A(u^{(m)})]u^{(m)}. \tag{3.1}$$

As reported in the literature (e.g., [6, 20]), fixed point iteration is fairly robust and effective for this problem. Our own experience, reported below, confirms this.

The focus of this work is on the solution of (some discretized versions of) the linear problems (3.1) in the fixed-point iterations. Hence we shall consider the fixed-point counter m to be fixed, drop the dependence on m from the notation, and write equation (3.1) as

$$[\mathcal{K}^*\mathcal{K} + \alpha A]w = f, \tag{3.2}$$

with the understanding that A is an ellptic operator of the form (2.12), with v possessing possibly large gradient.

In the computations presented in Sections 2.4, $\mathcal{K}$ was represented as a full matrix. This is of course impractical for all but the most modest levels of discretization and in general there is a need for iterative methods and preconditioners. In the following sections we examine the performance of the preconditioned conjugate gradient method for (3.2).

3.1. Multilevel Preconditioners. We turn now to the construction of the multilevel preconditioners for solving discretized versions of the linearized PDE (3.2), where A is given by (2.12) for a given parameter β.

If the action of A^{-1} were available and inexpensive, then a straightforward choice of preconditioner for (3.2) would be A^{-1}. The obvious advantage of this would be that for each fixed $\alpha > 0$ the condition number of $A^{-1}(\mathcal{K}^*\mathcal{K} + \alpha A)$ would be bounded independently of mesh size, so that we could expect favorable performance from the PCG scheme; however, the condition number would deteriorate as $\alpha \downarrow 0$. This approach has been investigated in

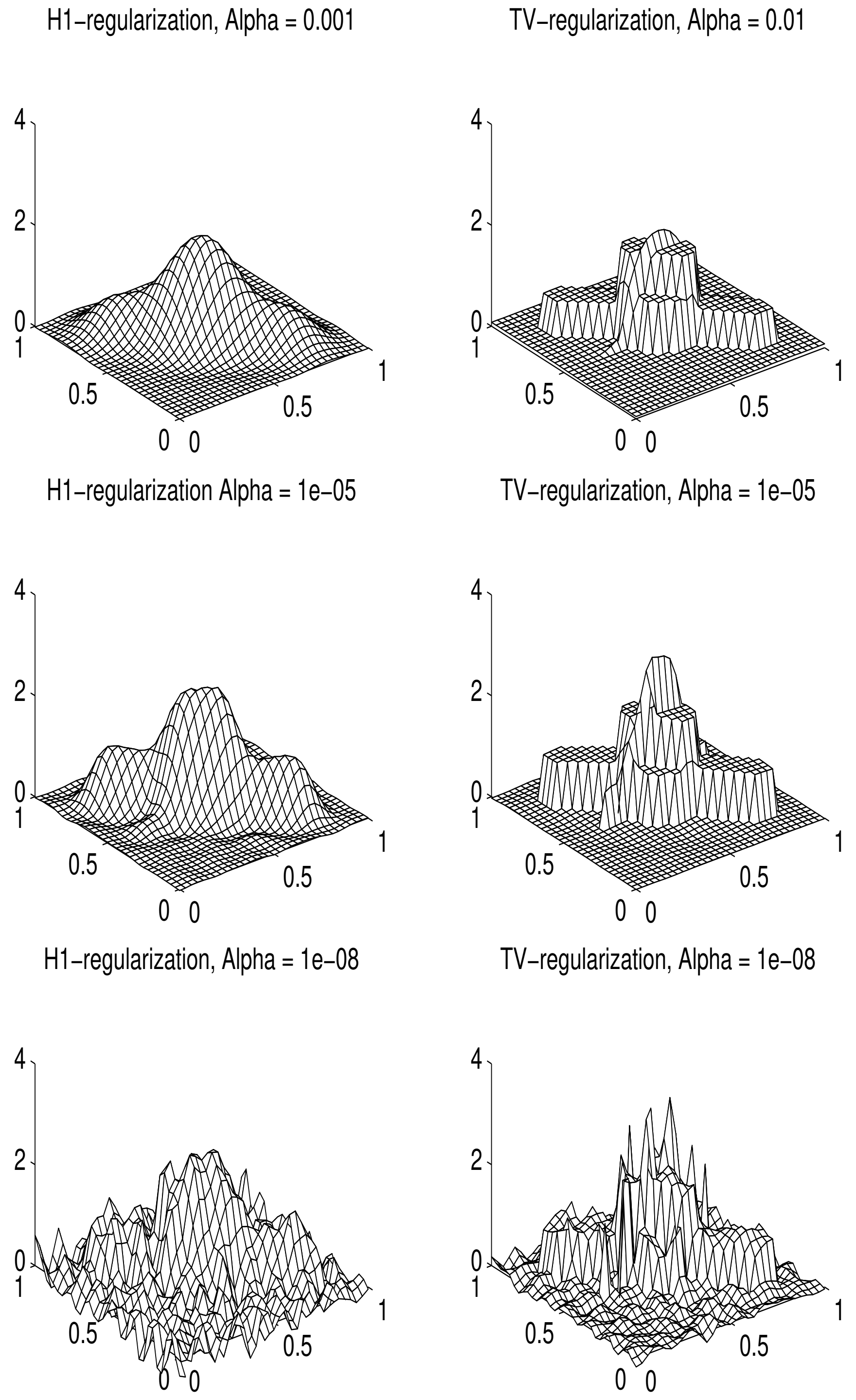

FIG. 2.4. *Results of the numerical investigation, discussed in Section 2.5.1, of the effects of different α, and the clear advantage of TV regularization over H^1 regularization in resolving discontinuous u.*

some detail numerically in [18]. Based on the experience reported in [18] and the overall simplicity and generality of this method, it merits some attention in our judgement.

The preconditioners which we study below are based on this idea. In particular, we examine the numerical properties of the the PCG scheme for (3.2) where the preconditioner is one or two multilevel V-cycles for approximating the action of A^{-1}.

A major difficulty comes from the gradient of the given function v in (3.2) which may have very large values. In fact, in our application, this is a typical situation; namely, v will have highly oscillatory behavior, though being bounded away from zero and bounded above — see Figure 4.1. It is clear that to capture the oscillatory behavior of the coefficient and the respective solution w we have to discretize the problem on a relatively fine grid and discretizations on coarse grids may generally not have good approximation properties. Therefore, to develop efficient iterative schemes based on effective preconditioners such as the multilevel ones, we have to create the coarse problems in an *algebraic* manner, rather than using discretizations of (3.2) on respective coarse grids. That is, we first generate a discretization of the problem (3.2), using finite elements for example, getting the respective stiffness matrix $A = A_h$ coming from the $\nabla \cdot (k\nabla w)$ part of the problem on a sufficiently fine mesh and then use algebraic coarsening to define coarse–grid stiffness matrices. That is, let J as in Section 2.2 be $J \simeq \log h^{-1}$ and define $A^{(J)} = A_h$. Assuming that the mesh $\mathcal{T}_h$ has been obtained by $J \geq 1$ steps of uniform refinement of an initial coarse mesh $\mathcal{T}_0$, then $A^{(k-1)} = \left(I^k_{k-1}\right)^T A^{(k)} I^k_{k-1}$, for $k = J, J-1, \ldots, 2, 1$, where I^k_{k-1} is the intergrid transfer matrix discussed in Section 2.2. The sparsity pattern of $A^{(k)}$ remains the same as that of A_h; namely, in terms of stencil, we have nine–point stencil representation of $A^{(k)}$ at each level k.

Below we consider three types of multilevel preconditoners $B^{(k)}$, which are approximations of $A^{(k)}$ and easily invertible. The first of these is based on the standard multigrid method, discussed in Section 3.1.1. The other two are based on the so–called two–level partitioning of the matrix $A^{(k)}$ that corresponds to a two–space decomposition of the current finite element space V_k, which we now discuss.

We assume a decomposition

$$V_k = V^1_k + V_{k-1}, \tag{3.3}$$

which is not necessarily a direct one, and define the following block partitioning of $A^{(k)}$,

$$\widehat{A}^{(k)} = \left[\begin{array}{cc} \widehat{A}^{(k)}_{11} & \widehat{A}^{(k)}_{12} \\ \widehat{A}^{(k)}_{21} & A^{(k-1)} \end{array}\right] \begin{array}{l} \}\ \ V^1_k \\ \}\ \ V_{k-1} \end{array} . \tag{3.4}$$

We use the notation "$\widehat{A}^{(k)}$" to distinguish the representation of the elliptic operator in the computational bases of V^1_k and V_{k-1} from the notation "$A^{(k)}$" representing the operator in standard nodal basis of V_k. We have, $\widehat{A}^{(k)}_{11} = Y_1^{(k)^T} A^{(k)} Y_1^{(k)}$, $\widehat{A}^{(k)}_{12} = Y_1^{(k)^T} A^{(k)} Y_2^{(k)}$, $\widehat{A}^{(k)}_{21} = Y_2^{(k)^T} A^{(k)} Y_1^{(k)}$ and $A^{(k-1)} = Y_2^{(k)^T} A^{(k)} Y_2^{(k)}$. The block $Y_2^{(k)} = I^k_{k-1}$ is the natural coarse–to–fine (interpolation) transfer matrix, whereas the block $Y_1^{(k)}$ comes from the subspace V^1_k and represents the natural imbedding of V^1_k into V_k. For the time being we will not specify the space V^1_k and its corresponding transfer matrix $Y_1^{(k)}$. We only mention that in the extreme case one can have $V^1_k = V_k$ and hence $Y_1^{(k)} = I$. We assume, though, that the actions of $Y_1^{(k)}$ and $Y_1^{(k)^T}$ are readily available and inexpensive.

Based on the block–partitioning (3.4) we are now in a position to define our multilevel preconditioner $B^{(k)}$ for $A^{(k)}$, based on the choice of $V_1^{(k)}$, by a routine recursive argument.

DEFINITION 1 (MULTILEVEL PRECONDITIONERS).

ETNA
Kent State University
etna@mcs.kent.edu

- *$B^{(0)} = A^{(0)}$;*
- *For $1 \le k \le J$, define $B^{(k)}$ by ${B^{(k)}}^{-1} = \left[Y_1^{(k)},\ Y_2^{(k)}\right] {\widehat{B}^{(k)}}^{-1} \left[Y_1^{(k)},\ Y_2^{(k)}\right]^T$, where*

$$\widehat{B}^{(k)} = \begin{bmatrix} B_{11}^{(k)} & 0 \\ \widehat{A}_{21}^{(k)} & B^{(k-1)} \end{bmatrix} \begin{bmatrix} I & {B_{11}^{(k)}}^{-1} \widehat{A}_{12}^{(k)} \\ 0 & I \end{bmatrix}.$$

The $\hat{A}_{ij}^{(k)}$ here are from (3.4), and the $B_{11}^{(k)}$ is depends the choice of $V_1^{(k)}$; we discuss it further below.

The block $B_{11}^{(k)}$ in this definition is an approximation of the block $\widehat{A}_{11}^{(k)}$ of $\widehat{A}^{(k)}$ in a space complementary to V_{k-1} (in V_k). It gives rise to the so–called smoothing iteration in the multigrid method. Hence the choice of $B_{11}^{(k)}$ depends on the properties of the space V_k^1 and the means by which $\widehat{A}_{11}^{(k)}$ is approximated.

To implement the action of ${B^{(k)}}^{-1}$ one needs the actions of ${B_{11}^{(k)}}^{-1}$ and of the transformation matrices $Y_r^{(k)}$ as well as the actions of their transpositions ${Y_r^{(k)}}^T$, $r = 1, 2$ at every level k. The factored form of $\widehat{B}^{(k)}$ can be utilized to get the inverse actions of $\widehat{B}^{(k)}$ in the usual forward and backward elimination sweeps. Algorithmically, for a given $b \in V_k$ represented in the standard nodal basis, the computation $w = {B^{(k)}}^{-1} b$ in the nodal basis may be expressed as follows.

1. Transform b to the two-level basis:
   $b^1 = {Y_1^{(k)}}^T b$ and $b_{k-1} = {Y_2^{(k)}}^T b$.
2. Perform the forward elimination, creating intermediate vectors ϕ and ψ:
   $\phi = {B_{11}^{(k)}}^{-1} b_k^1$,
   $\psi = (B^{(k-1)})^{-1}\big(b_{k-1} - \widehat{A}_{21}^{(k)} b_k^1\big)$.
3. Perform the backward elimination:
   $w_{k-1} = \psi$ and $w_1 = \phi - {B_{11}^{(k)}}^{-1} \widehat{A}_{12}^{(k)} w_{k-1}$.
4. Transform (w_1, w_{k-1}) to the nodal basis:
   $w = Y_1^{(k)} w_1 + Y_2^{(k)} w_{k-1}$.

We turn now to three main choices we have made in our numerical tests.

3.1.1. Multigrid method. We denote this preconditioner by $B_{MG}^{(k)}$. It is a standard variational multigrid "V(1,1)" cycle; it is not of the form of $B^{(k)}$ given in Definition 1. However, it can be discussed in terms of the decomposition 3.3: here $V_k^1 = V_k$, so that $Y_1^{(k)} = I$ and $\widehat{A}_{11}^{(k)} = A^{(k)}$. For $B_{11}^{(k)}$ we have chosen the symmetric Gauss–Seidel approximation to $A^{(k)}$. Namely, if $A^{(k)} = D^{(k)} - L^{(k)} - U^{(k)}$ is split into diagonal, strictly lower triangular and strictly upper triangular parts, then $B_{11}^{(k)} = (D^{(k)} - L^{(k)}){D^{(k)}}^{-1}(D^{(k)} - U^{(k)})$.

Algorithmically, for a given $b \in V_k$, the computation $w = (B_{MG}^{(k)})^{-1} b$ may be expressed in a manner similar to the algorithm given above for ${B^{(k)}}^{-1}$, as follows.

1. Compute the projection b_{k-1} of b upon the coarse grid:
   $b_{k-1} = {Y_2^{(k)}}^T b$.
2. Perform one symmetric Gauss–Seidel "smoothing iteration":
   $\phi = {B_{11}^{(k)}}^{-1} b$.
3. Compute the residual $b - A\phi$, project it to the coarse grid, and solve the coarse grid equation:
   $\psi = (B_{MG}^{(k-1)})^{-1}\big(b_{k-1} - \widehat{A}_{21}^{(k)} b_k^1\big)$.

4. Perform the "coarse–grid update":
   $w = \phi + Y_2^{(k)}\psi.$
5. Perform one symmetric Gauss–Seidel "smoothing iteration":
   $w = w + B_{11}^{(k)^{-1}}(b - Aw).$

For more details on multigrid we refer to Bramble [2] or Oswald [12].

3.1.2. Hierarchical basis method. The classical HB method of Yserentant [23] corresponds to the case in which V_k^1 is the standard two-level hierarchical complement of V_{k-1} in V_k. It is given by $(I_k - I_{k-1})V_k$, where I_k stands for the nodal interpolation; namely, for any continuous function v, $I_k v \in V_k$ is defined as

$$I_k v = \sum_{x_i \in \mathcal{N}_k} v(x_i)\varphi_i^{(k)},$$

where $\{\varphi_i^{(k)},\ x_i \in \mathcal{N}_k\}$ stands for the nodal basis of V_k and $\mathcal{N}_k$ is the nodal set (the vertices of the rectangles from $\mathcal{T}_k$) at level k. That is, $(I_k v)(x_i) = v(x_i)$ for all $x_i \in \mathcal{N}_k$. In this case the block $Y_1^{(k)}$ is given by

$$Y_1^{(k)} = \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{array}{l} \} \quad \mathcal{N}_k \setminus \mathcal{N}_{k-1} \\ \} \quad \mathcal{N}_{k-1} \end{array} . \tag{3.5}$$

3.1.3. Approximate wavelet-modified hierarchical basis method. Here we consider the approximate wavelet-modified hierarchical basis or AWM-HB preconditioner. The block $Y_1^{(k)}$ has a more complicated structure, coming from a corresponding space $V_k^1 = (I - Q_{k-1}^a)(I_k - I_{k-1})V_k$, where I_k is as in Section 3.1.2, and Q_{k-1}^a stands for an approximate L^2–projection operator with some readily-available matrix representation Π_k. The exact L^2–projection operator Q_k is defined in the usual way; namely,

$$(Q_k v, \varphi) = (v, \varphi), \qquad \text{for all } \varphi \in V_k.$$

We remark that if we let $Q_{k-1}^a = 0$ then we recover the classical HB method. Due to the (modification) term $-Q_{k-1}(I_k - I_{k-1})V_k$ the above method is called approximate wavelet modified HB. The name "wavelet" stands for the extreme case of $Q_{k-1}^a = Q_{k-1}$, since it that case one gets the wavelet (L^2–orthogonal) decomposition $V_k = W_k \oplus V_{k-1}$, where $W_k = (Q_k - Q_{k-1})V_k$. The latter is impractical to use since no simple locally supported bases of W_k are available.

Note that to compute the actions of the exact projection Q_k, one has to solve a mass–matrix problem at level k i.e., with the mass matrix

$$G^{(k)} = \left\{(\varphi_j^{(k)},\ \varphi_i^{(k)})\right\}_{x_i, x_j \in \mathcal{N}_k} . \tag{3.6}$$

Although the mass matrices are well–conditioned it may become too costly to evaluate the exact projections Q_k. To define an optimal order preconditioner $B^{(k)}$ it turns out that it is sufficient to have a good approximations Q_k^a to Q_k. For an analysis and implementations of the AWM–HB–preconditioners we refer to Vassilevski and Wang [16], [17], see also Vassilevski and Wang [15] and the survey Vassilevski [14]. The choice we have made in the present numerical tests for Q_k^a is based on very simple approximation of the inverse of the coarse grid mass matrix $G^{(k-1)^{-1}}$, namely,

$$\widetilde{G}^{(k)^{-1}} = D^{(k-1)^{-1}}. \tag{3.7}$$

ETNA
Kent State University
etna@mcs.kent.edu

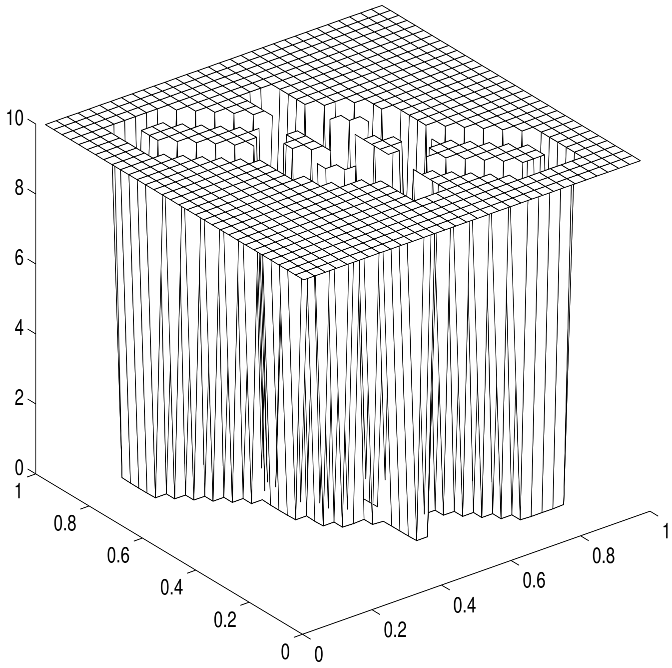

FIG. 4.1. *The coefficient κ, as given by (2.13) in the operator $A = A(U_{TRUE})$, with U_{TRUE} is described in Section 2.3.1, and $\beta = 1/10$. The mesh in this figure is 32×32 ($k = 5$).*

where $D^{(k-1)}$ is the main diagonal of $G^{(k-1)}$. With this, for our matrix prepresentation Π_k of Q_k^a we used

$$\Pi_k = I_{k-1}^k \widetilde{G}^{(k-1)^{-1}} \left(I_{k-1}^k\right)^T D^{(k)}.$$

The matrix representation of the transformation matrix block $Y_1^{(k)}$ then reads as

$$Y_1^{(k)} = [I - \Pi_k] \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{array}{l} \} \quad \mathcal{N}_k \setminus \mathcal{N}_{k-1} \\ \} \quad \mathcal{N}_{k-1} \end{array} . \tag{3.8}$$

Finally, the block $B_{11}^{(k)}$ in Definition 1 corresponded to the symmetric Gauss–Seidel approximation to $\widehat{A}_{11}^{(k)} = Y_1^{(k)^T} A^{(k)} Y_1^{(k)}$. The latter we formed explicitly as a sparse matrix using the fact that $Y_1^{(k)}$ is a sparse matrix.

4. Numerical Experiments.

4.1. Performance of the preconditioned CG methods. To gain some insight into the effectiveness of the methods of Section 3.1 for the probem (3.1), we performed a set of comptutations using the preconditioned conjugate gradient method to find approximate solutions of the form $Au = f$, where $A = A(v)$ as given in (2.12) with $\beta = 1/10$ and with $v = U_{TRUE}$ as described in Section 2.3.1. For a level-5 mesh (32×32), the resulting a coeffient κ, as given by (2.13), is shown in Figure 4.1. We took $f = Aw$ where $w(x, y) = \cos(3x)\cos(5y)$.

We computed approximations of the solution of $Au = f$ using preconditioned conjugate gradients with various multilevel V-cycle preconditioners on levels $k = 4, 5, 6$ and 7. The preconditioners which we used in these computations were: the multigrid method, the HB

ETNA
Kent State University
etna@mcs.kent.edu

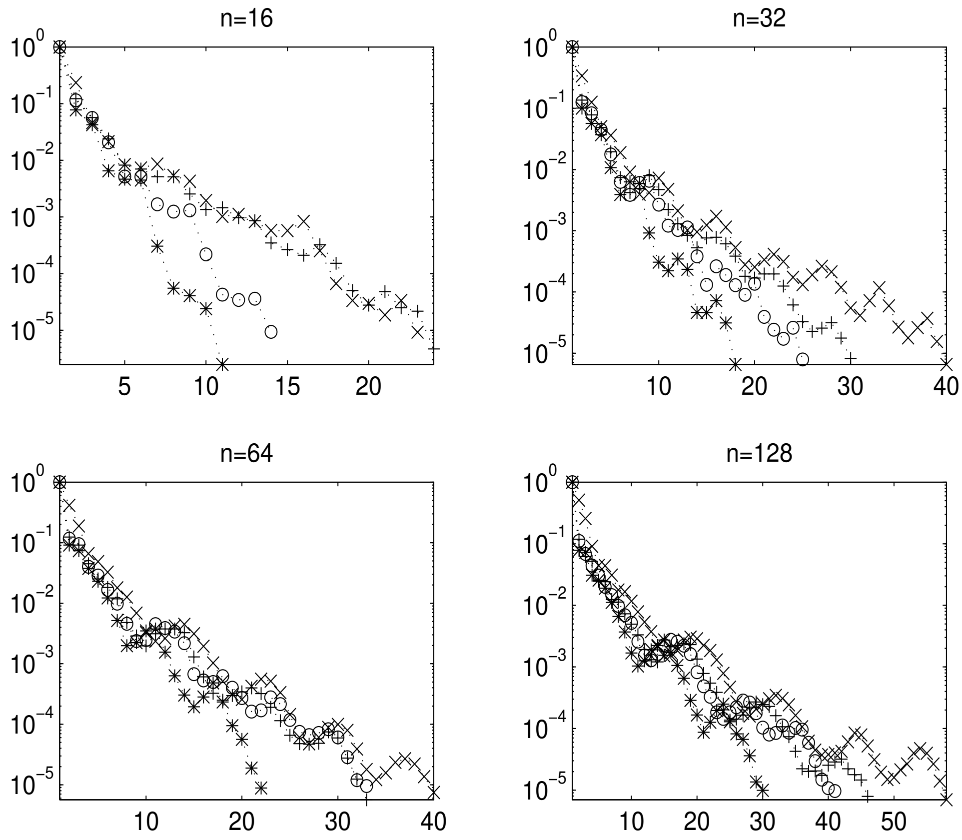

FIG. 4.2. *Shown here are convergence histories for the preconditioned conjugate gradient method for the example described in Section 4.1. The L^2 norms of the residuals are plotted versus iteration count. The symbols $\star$, x, + and $\circ$ represent the various methods: $\star$ indicates MG, x indicates HB, + indicates AWM-HB, and $\circ$ indicates AWM-HB with weighted norm.*

method, the AWM-HB method, and the "weighted" AWM-HB method. The latter is the AWM-HM method as described in Section 3.1.3, except that in place of the standard mass matrix $G^{(k)}$ we used the weighted mass matrix

$$G_w^{(k)} \stackrel{\mathrm{def}}{=} \left\{ \left(\kappa \varphi_j^{(k)},\ \varphi_i^{(k)} \right) \right\}_{x_i, x_j \in \mathcal{N}_k}$$

instead of $G^{(k)}$ as given by (3.6). The results of these computations are presented in Figure 4.2. Our experience, as reported here, indicates the clear superiority of the multigrid preconditioner for this problem.

4.2. Computational solution of an inverse problem. Finally, we performed the numerical minimization of $\Phi^{\alpha,\beta}$ as described in (2.10) of Section 2.5, via the fixed point iteration (3.1), for the deconvolution problem with noisy data as described in Section 2.3.1. Because the results of Section 4.1 sugguest a clear superiority of the multigrid preconditioner for this problem, we used it in the inverse problem.

We performed the minimization on the level 6 (64×64) mesh. The parameters α and β were set to 10^{-6} and 10^{-1}, respectively, The stopping criterion for the fixed point iterations

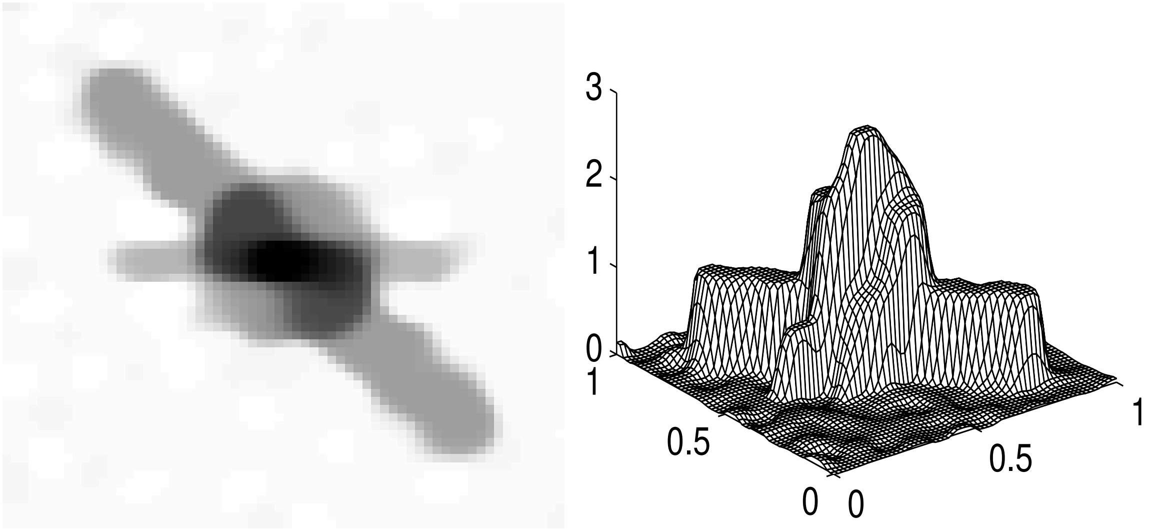

FIG. 4.3. *Computational results of the inverse problem described in Section 4.2.*

TABLE 4.1
For the example described in Section 4.2, shown here are, for each of the fixed point iterations, the number of preconditioned conjugate gradient (PCG) iterations taken in solving (3.1) (the maximimum allowed was 40), the L^2 norm of the residual at the end of the PCG iterations, and an estimate of the relative L^1 norm of the resulting step δu.

FP iter. #	# PCG iter's	$\|\text{residual}\|_2$	Relative F.P. stepsize
1	14	3.271e-06	1.121
2	40	2.467e-06	0.527
3	40	1.495e-06	0.1904
4	40	1.834e-06	0.3213
5	40	1.262e-06	0.3524
6	40	7.913e-07	0.1646
7	40	3.743e-07	0.07962
8	40	1.418e-07	0.02539
9	40	5.58e-08	0.01049
10	40	2.349e-08	0.006164

was when the approximation of relative L^1 norm $\|\delta u^{(m)}\|_1/\|u^{(m)}\|_1$ of the fixed-point step $\delta u^{(m)}$ fell below 10^{-2}. (The approximate L^1 norms were computed simply by taking the l_1 norm of the nodal coefficients.) On each fixed point iteration, at most 40 PCG iterations were allowed in the approximate solution of (3.1). The performance of this scheme is reported in Table 4.1; the resulting recontruction of u is shown in Figure 4.3.

5. Summary. We have given a rather detailed motivation for the use of total variation regularization for distributed parameter inverse problems. While having the advantage of allowing discontinuities in the reconstructions, this method is computationally intensive. Specifically, the use of this regularization for minimization problems results in the objective functional (2.8) with nonlinear Euler equation (2.11). Fixed point iteration for this equation leads to a sequence of linear problems of the form (3.2), $[\mathcal{K}^*\mathcal{K} + \alpha A]w = f$. We have examined the numerical performance of the preconditioned conjugate gradient method for this equation, using as the preconditioner various multilevel approximations of the A^{-1}.

As indicated in §4.1, of the preconditioners we studied, the standard multigrid method was clearly superior. Further, the numerical results of §4.2 demonstrate that this approach is reasonable effective for the minimization problem with total variation regularization, at least for the deconvolution problem.

However, the results of that section also suggest a need for further improvement. In particular, Table 4.1 shows that on each of the fixed point iterations (except the first), the preconditioned conjugate gradient method terminated after the maximimum prescribed number of iterations (forty). This suggests that good approximations of A^{-1} are not necessarily good approximations for $[\mathcal{K}^*\mathcal{K} + \alpha A]^{-1}$. Since $\mathcal{K}^*\mathcal{K}$ is compact and its eigenvalues converge rapidly to zero, a possible strategy for approximating $[\mathcal{K}^*\mathcal{K} + \alpha A]^{-1}$ would be to use a multigrid approximation of A^{-1} together with a low-rank approximation for $\mathcal{K}^*\mathcal{K}$, and the Sherman-Morrison formula for perturbations of matrix inverses (see, e.g., [8]). This idea will be pursued in future work.

REFERENCES

[1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] J. H. BRAMBLE, *Multigrid Methods*, Second Ed., Pitman Research Notes in Mathematics Series, no. 234, Longman Scientific & Technical, UK, 1995.

[3] T. F. CHAN, H. M. ZHOU, AND R. H. CHAN, *Continuation method for total variation denoising problems*, UCLA Report CAM 95-18, 1995.

[4] DAVID C. DOBSON, AND FADIL SANTOSA, *An image-enhancement technique for electrical impedance tomography*, Inverse Problems, 10 (1994), pp. 317–334.

[5] ———, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1198.

[6] DAVID C. DOBSON AND CURTIS R. VOGEL, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer Anal., 34 (1997), pp. 1779–1791.

[7] ENRICO GIUSTI, *Minimal surfaces and functions of bounded variation*, Birkhauser, Boston, 1984.

[8] W. W. HAGER, *Updating the inverse of a matrix*, SIAM Review, 31 (1989), pp. 221–239.

[9] *M. Hanke and J. G. Nagy*, *Restoration of atmosherically blurred images by symmetric indefinite conjugate gradient techniques*, Inverse problems, 12 (1996), pp. 157–173.

[10] P.L. LIONS, S. OSHER, AND L. RUDIN, *Denoising and deblurring algorithms with constrained nonlinear PDE's*, SIAM J. Numer. Analysis, submitted.

[11] M.E. OMAN, *Fast multigrid techniques in total variation-based image reconstruction*, in 7th Copper Mountain Conf. on Multigrid Methods, NASA Conference Publication 3339, April 1995. Please supply page numbers.

[12] P. OSWALD, *Multilevel Finite Element Approximation. Theory and Applications*, Teubner Skripten zur Numerik, Teubner, Stuttgart, 1994.

[13] L.I. RUDIN, S. OSHER, AND C. FU, *Total variation based restoration of noisy, blurred images*, preprint, Cognitech, Inc., 2800 - 28th St., Suite 101, Santa Monica, CA 90405, SIAM J. Numer. Analysis, submitted.

[14] P. S. VASSILEVSKI, On two ways of stabilizing the hierarchical basis multilevel methods, *SIAM Review* 39 (1997). Please supply page numbers.

[15] P. S. VASSILEVSKI AND J. WANG, *Wavelet–like methods in the design of efficient multilevel preconditioners for elliptic PDEs*, in Multiscale Wavelet Methods for PDEs, W. Dahmen, A. Kurdila and P. Oswald, eds., Academic Press, 1997, to appear.

[16] ———, *Stabilizing the hierarchical basis by approximate wavelets, I: theory*, Numer. Linear Alg. Appl. (1997), to appear.

[17] ———, *Stabilizing the hierarchical basis by approximate wavelets, II: implementation and numerical experiments*, SIAM J. Sci. Comput., submitted.

[18] C. R. Vogel. *Sparse matrix equations arising in distributed parameter identification*, SIAM J. Matrix Analysis, submitted.

[19] ———, Publicly-available software,
`http://www.math.montana.edu/~vogel/deconv`,
`http://www.math.montana.edu/~vogel/deconv/Source/FFT/`

[20] C. R. VOGEL AND M. E. OMAN, *Iterative methods for total variation denoising*, SIAM J. Sci. Comp. 17 (1996), pp. 227–238.

ETNA
Kent State University
etna@mcs.kent.edu

[21] ———, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Trans. on Image Proc., submitted.

[22] J. G. WADE, K. SENIOR. AND S. SEUBERT, *Convergence of derivative approximations in the inverse conductivity problem*, BGSU Tech Rep. #96-14; SIAM J. Appl. Math, submitted.

[23] H. YSERENTANT, *On the multilevel splitting of finite element spaces*, Numer. Math. 49 (1986), pp. 379–412.