A NONLINEAR PRIMAL-DUAL METHOD FOR TOTAL VARIATION-BASED IMAGE RESTORATION*

TONY F. CHAN†, GENE H. GOLUB‡ AND PEP MULET§

Abstract. We present a new method for solving total variation (TV) minimization problems in image restoration. The main idea is to remove some of the singularity caused by the non-differentiability of the quantity $|\nabla u|$ in the definition of the TV-norm *before* we apply a linearization technique such as Newton's method. This is accomplished by introducing an additional variable for the *flux* quantity appearing in the gradient of the objective function, which can be interpreted as the normal vector to the level sets of the image u. Our method can be viewed as a primal-dual method as proposed by Conn and Overton [9] and Andersen [3] for the minimization of a sum of Euclidean norms. In addition to possessing local quadratic convergence, experimental results show that the new method seems to be globally convergent.

1. Introduction. During some phases of the manipulation of an image some random noise and blurring is usually introduced. The presence of this noise and blurring makes difficult and inaccurate the latter phases of the image processing.

The algorithms for noise removal and deblurring have been mainly based on least squares. The output of these $\mathcal{L}^2$-based algorithms will be a continuous function, which cannot obviously be a good approximation to our original image if it contains edges. To overcome this difficulty a technique based on the minimization of the Total Variation norm subject to some noise constraints is proposed in [19], where a time marching scheme was also proposed to solve the associated Euler-Lagrange equations. Since this method can be slow due to stability constraints in the time step size, a number of alternative methods have been proposed in the literature, [23], [7], [16].

One of the difficulties in solving the Euler-Lagrange equations is the presence of a highly nonlinear and non-differentiable term, which causes convergence difficulties for Newton's method even when combined with a globalization technique such as a line search. The idea of our new algorithm is to remove some of the singularity caused by the non-differentiability of the objective function *before* we apply a linearization technique such as Newton's method. This is accomplished by introducing an additional variable for the *flux* quantity appearing in the gradient of the objective function, which can be interpreted as the unit normal to the level sets of the image function. Our method can be viewed as a primal-dual method as proposed by Conn and Overton [9] and Andersen [3] for the minimization of a sum of Euclidean norms. We caution that our use of the name *primal-dual* is based on a duality principle applied to the TV-norm (to be explained in more detail in Section 4) and should not be confused with the popular algorithms with the same name in linear and non-linear programming. Experimental results show that the new method is globally convergent, whereas the

†Department of Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1555. E-mail address: chan@math.ucla.edu. Supported by grants NSF ASC-92-01266 and ONR-N00014-96-1-0277.

‡Computer Science Department, Stanford University. E-mail: golub@sccm.stanford.edu.The work of this author was in part supported by the NSF :Grant # CCR-9505393.

§Department of Mathematics, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1555 and Departament de Matemàtica Aplicada, Universitat de València, Dr. Moliner, 50, 46100 Burjassot, Spain. E-mail address: mulet@math.ucla.edu. Supported by DGICYT grants EX94 28990695 and PB94-0987.

*This is a revised version of the earlier UCLA Computational and Applied Mathematics technical report 95-43 with the same title.

primal Newton method has a small domain of convergence. It is hoped that the new approach can be applied to other geometry-based PDE methods in image restoration, such as anisotropic diffusion [18], affine invariant flows [20] and mean curvature flows [2], since the same singularity caused by $|\nabla u|$ occurs in these formulations as well.

The organization of this paper is as follows: in section 2 we introduce the problem, the nonlinear equations associated to it and discuss how to solve them. In section 3 we present our new linearization technique for the (unconstrained) Tikhonov regularization form of the problem. In section 4 we show that the new linearization corresponds to a classical primal-dual formulation for some convex problems. Finally, in section 5 we present some numerical results for the denoising case.

2. Total Variation Regularization. An image can be interpreted as either a real function defined on Ω, a bounded and open domain of $\mathbb{R}^2$, (for simplicity we will assume Ω to be a rectangle henceforth) or as a suitable discretization of this continuous image.

Our interest is to restore an image which is contaminated with noise and/or blur. The restoration process should recover the edges of the image. Let us denote by z the observed image and u the real image. The model of degradation we assume is $Ku+n=z$, where n is a Gaussian white noise and K is a (known) linear *blur operator* (usually a convolution operator).

In general, the problem $Ku=z$, with K a compact operator, is ill-posed, so it is not worth solving this equation (or a discretization of it), for the data in the usual applications is inexact, and the solution would be highly oscillatory. But if we impose a certain *regularity* condition on the solution u, then the problem becomes stable. We can consider two related techniques of regularization: Tikhonov regularization and noise level constrained regularization.

Tikhonov regularization consists in solving the unconstrained optimization problem:

$$\min_u \alpha R(u) + \frac{1}{2}||Ku - z||^2_{\mathcal{L}^2}, \tag{2.1}$$

for some functional R which measures the irregularity of u in a certain sense and a suitably chosen coefficient α which will balance the tradeoff between a good fit to the data and a *regular* solution.

Another approach consists in solving the following constrained optimization problem:

$$\begin{aligned} &\min_u \; R(u) \\ &\text{subject to } ||Ku - z||^2_{\mathcal{L}^2} = \sigma^2, \end{aligned} \tag{2.2}$$

Here we seek a solution with minimum irregularity from all candidates which match the known noise level.

Examples of regularization functionals that can be found in the literature are, $R(u) = ||u||, ||\Delta u||, ||\nabla u \cdot \nabla u||$, where ∇ is the gradient and Δ is the Laplacian, see [21] and [14]. The drawback of using these functionals is that they do not allow discontinuities in the solution, and since we are interested in recovering features of the image, they are not suitable for our purposes.

In [19], it is proposed to use as regularization functional the so-called *Total Variation norm* or *TV-norm*:

$$TV(u) = \int_\Omega |\nabla u| \, dx\, dy = \int_\Omega \sqrt{u_x^2 + u_y^2} \, dx\, dy. \tag{2.3}$$

The TV norm does not penalize discontinuities in u, and thus allows us to recover the edges of the original image. For simplicity we use in this section the Tikhonov formulation of the problem. Treatment of the constrained problem (2.2) can be found in [8]. Hence the restoration problem can be written as:

$$\min_u \int_\Omega \left(\alpha\sqrt{u_x^2+u_y^2}+\frac{1}{2}(Ku-z)^2\right)dx\,dy, \tag{2.4}$$

that is:

$$\min_u \alpha|| \, |\nabla u| \, ||_{\mathcal{L}^1}+\frac{1}{2}||Ku-z||_{\mathcal{L}^2}^2, \quad |\nabla u|=\sqrt{u_x^2+u_y^2}. \tag{2.5}$$

The Euler-Lagrange equation for this problem, assuming homogeneous Neumann boundary conditions, is:

$$0=-\alpha\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right)+K^*(Ku-z). \tag{2.6}$$

where K^* is the adjoint operator of K with respect to the L^2 inner product. This equation is not well defined at points where $\nabla u=0$, due to the presence of the term $1/|\nabla u|$. A commonly used technique to overcome this difficulty is to slightly perturb the Total Variation norm functional to become:

$$\int_\Omega \sqrt{|\nabla u|^2+\beta}\,dx\,dy, \tag{2.7}$$

where β is a small positive parameter. In [1] it is shown that the solutions of these perturbed problems converge to the solution of (2.4) when $\beta\to 0$. So now the problem is:

$$\min_u \alpha\int_\Omega \sqrt{|\nabla u|^2+\beta}\,dx\,dy+\frac{1}{2}||Ku-z||_{\mathcal{L}^2}^2, \tag{2.8}$$

and the corresponding Euler-Lagrange equation is:

$$0=-\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)+K^*(Ku-z)=g(u). \tag{2.9}$$

The main difficulty that this equation poses is the linearization of the highly nonlinear term $-\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)$.

A number of methods have been proposed to solve (2.9). L. Rudin, S. Osher and E. Fatemi [19] used a time marching scheme to reach the steady state of the parabolic equation $u_t=-g(u)$ with initial condition $u=z$:

$$u_t=\alpha\nabla\cdot\left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)-K^*(Ku-z), \quad u(x,0)=z(x). \tag{2.10}$$

This method can be slowly convergent due to stability constraints. C. Vogel and M. Oman [23] proposed the following fixed point iteration to solve the Euler-Lagrange equation:

$$-\alpha\nabla\cdot\left(\frac{\nabla u^{k+1}}{\sqrt{|\nabla u^k|^2+\beta}}\right)+K^*(Ku^{k+1}-z)=0. \tag{2.11}$$

Starting with $u_0 = z$, at each step u^{k+1} is obtained as the solution of the linear differential-convolution equation (2.11), whose coefficients are computed from u^k. This method is robust but only linearly convergent.

Due to the presence of the highly nonlinear term $\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2+\beta}}\right)$, Newton's method does not work satisfactorily, in the sense that its domain of convergence is very small. This is especially true if the regularizing parameter β is small. On the other hand, if β is relatively large then this term is well behaved. So it is natural to use a continuation procedure starting with a large value of β and gradually reducing it to the desired value. T. Chan, R. Chan and H. Zhou proposed in [7] such an approach. Although this method is locally quadratically convergent, the choice of the sequence of subproblems to solve is crucial for its efficiency. The authors have not succeeded in finding a fully satisfactory selection procedure, although some heuristics can be used.

3. A new linearization based on a dual variable. We propose here a better technique to linearize the term $\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right)$. This technique bears some similarity to techniques from some specific non-linear primal-dual optimization methods [3], [9] and gives a better global convergence behavior than that of the usual Newton's continuation method. In the next section we give a justification of the primal-dual nature of the algorithm that we present below.

The method is based on the following simple observation. While the singularity and non-differentiability of the term $w = \nabla u/|\nabla u|$ is the source of the numerical problems, w itself is usually *smooth* because it is in fact the unit normal vector to the level sets of u. The numerical difficulties arise only because we linearize it in the wrong way.

The idea of the new method is to introduce

$$w = \frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta}} \tag{3.1}$$

as a new variable and replace (2.9) by the following equivalent system of nonlinear partial differential equations:

$$\begin{aligned} -\alpha \nabla \cdot w + K^*(Ku - z) &= 0 \\ w\sqrt{|\nabla u|^2 + \beta} - \nabla u &= 0. \end{aligned} \tag{3.2}$$

We can then linearize this (u, w) system, for example by Newton's method. This approach is similar to the technique of introducing a *flux variable* in the *mixed finite element method* [4]. As we will see in the next section, these equation are, precisely, the equations of a saddle point problem, for which the problem (2.4) is the primal problem and w is the variable of the dual problem.

For completeness we compare the linearization of the u system:

$$\left[-\alpha \nabla \cdot \left(\frac{1}{|\nabla u|}(I - \frac{\nabla u \nabla u^T}{|\nabla u|^2})\nabla\right) + K^*K\right] \delta u = -g(u), \tag{3.3}$$

to the linearization of the (w, u)-system:

$$\begin{bmatrix} |\nabla u| & -(I - \frac{w \nabla u^T}{|\nabla u|})\nabla \\ -\alpha\nabla\cdot & K^*K \end{bmatrix} \begin{bmatrix} \delta w \\ \delta u \end{bmatrix} = -\begin{bmatrix} f(w, u) \\ g(w, u) \end{bmatrix}. \tag{3.4}$$

Equation (3.4) can be solved by first eliminating δw and solving the resulting equation for δu:

$$\left[-\alpha\nabla\cdot\left(\frac{1}{|\nabla u|}(I-\frac{w\nabla u^T}{|\nabla u|})\nabla\right)+K^*K\right]\delta u=-g(u) \tag{3.5}$$

After δu is obtained we can compute δw by:

$$\delta w=\frac{1}{|\nabla u|}(I-\frac{w\nabla u^T}{|\nabla u|})\nabla\delta u-w+\frac{\nabla u}{|\nabla u|} \tag{3.6}$$

We note that the cost per iteration of our new linearization technique is only slightly higher than for the standard Newton's method (3.3), because the main cost is the solution of the differential-convolution equations (3.3) and (3.5) for δu.

The motivation is that the (w,u) system is somehow better behaved than the u system. Although at this point we do not have a complete theory to support this, we will now give a scalar example that can explain the better convergence behavior of the new approach. We compare Newton's method applied to the equivalent equations $f(x)=a-\frac{x}{\sqrt{x^2+\beta}}=0$ (which resembles (3.1)) and $g(x)=a\sqrt{x^2+\beta}-x=0$ (which resembles $w\sqrt{|\nabla u|^2+\beta}-\nabla u=0$), where $a\approx 1$ and $\beta\approx 0$. In Fig 3.1 we can see that g looks more "linear" than f over much of the x-axis. In particular, Newton's method applied to f diverges when the initial guess is not close to the solution, whereas it converges when applied to g and for any positive initial guess. This is confirmed by the numerical results shown in Table 3.1. We believe that the reason why the algorithm presented here shows such a dramatic convergence improvement over the standard Newton's method is precisely this better linearization.

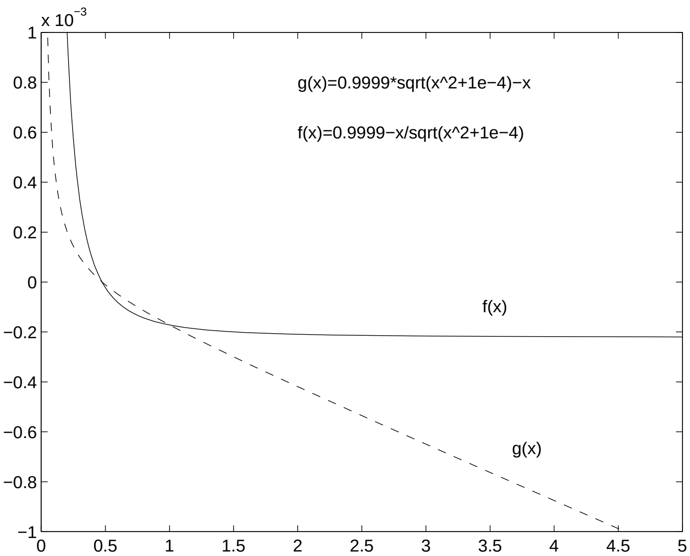

FIG. 3.1. *Plot of $f(x)$ and $g(x)$. The y-axis has been scaled by a factor of* 10^{-3}.

4. Duality. The goal of this section is to show that problem (2.4) is the primal problem for a corresponding dual problem and that the solution for both is characterized by a system of equations involving both the primal and the dual variables.

	Newton's iteration for $g(x)=0$					Newton's iteration for $f(x)=0$				
$x_0\backslash\beta$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}
1	7	6	4	3	5	*	9	6	6	*
2	6	5	2	4	6	*	8	4	*	*
3	6	4	3	5	6	10	7	5	*	*
4	5	4	3	5	7	9	6	*	*	*
5	5	3	4	6	7	8	5	*	*	*

TABLE 3.1

Comparison of the number of iterations required by Newton's method to solve $f(x)=0$ and $g(x)=0$, for $a=0.9999$, for different β (horizontally) and different initial guesses x_0 (vertically). A $*$ means that the corresponding iteration failed to converge.

As can be seen in [12], for instance, the TV-norm admits the following dual formulation:

$$TV(u)=\sup\{\int_\Omega -u\nabla\cdot w\,dx\,dy\colon w\in(C_0^\infty(\Omega))^2, \|w\|_\infty\le 1\}, \tag{4.1}$$

where $C_0^\infty(\Omega)$ is the set of smooth functions on Ω that vanish on the boundary $\partial\Omega$. It is fairly easy to see that this extends (2.3) to non-differentiable u. Actually, for u belonging to the Sobolev space $W^{1,1}(\Omega)$, an argument using integration by parts shows that the solution w to the problem

$$\sup_{|w|\le 1}\int_\Omega u\nabla\cdot w\,dx\,dy$$

satisfies in the limit

$$w(x)=\frac{\nabla u(x)}{|\nabla u(x)|},$$

for almost all $x\in\Omega$ such that $\nabla u(x)\ne 0$ or, equivalentyly,

$$|\nabla u|w-\nabla u=0, \tag{4.2}$$

almost everywhere.

With this formulation, problem (2.4) can be written as:

$$\min_u \sup_{\|w\|_\infty\le 1}\Phi(u,w),$$

where

$$\Phi(u,w)=\int_\Omega(\alpha u\nabla\cdot w+\frac{1}{2}(Ku-z)^2)\,dx\,dy. \tag{4.3}$$

By using arguments of duality theory for convex programming (see [11], for instance), we have that a solution (u^*,w^*) of (4.3) is a saddle point for $\Phi(u,w)$, that is:

$$\min_u \sup_{\|w\|_\infty\le 1}\Phi(u,w)=\Phi(u^*,w^*)=\sup_{\|w\|_\infty\le 1}\min_u\Phi(u,w), \tag{4.4}$$

and, moreover, w^* is solution of the *dual* problem:

$$\sup_{\|w\|_\infty \leq 1} \Psi(w), \tag{4.5}$$

where

$$\Psi(w) = \min_u \Phi(u, w). \tag{4.6}$$

Since Φ is quadratic in u, we can easily derive that this minimum must satisfy the following equation:

$$-\alpha \nabla \cdot w + K^*(Ku - z) = 0, \tag{4.7}$$

where K^* denotes the adjoint operator of K.

Furthermore, the solution u^* and w^* of (2.4) and (4.5), respectively (or (u^*, w^*) of (4.4)) can be characterized as the solution of the following set of *primal-dual* equations:

$$\left.\begin{aligned} -\alpha \nabla \cdot w + K^*(Ku - z) = 0 \\ |\nabla u| w - \nabla u = 0 \end{aligned}\right\}, \tag{4.8}$$

which are precisely the (w, u)-system of (3.2), the equations resulting from the introduction of the variable w.

5. Numerical results. Denote by C and $\overline{C}$ standard finite difference discretizations of the linear operators:

$$-\alpha \nabla \cdot \left(\frac{1}{|\nabla u|} (I - \frac{\nabla u \nabla u^T}{|\nabla u|^2}) \nabla \right) + K^* K \tag{5.1}$$

$$-\alpha \nabla \cdot \left(\frac{1}{|\nabla u|} (I - \frac{w \nabla u^T}{|\nabla u|}) \nabla \right) + K^* K, \tag{5.2}$$

that appear in equations (3.3) and (3.5), respectively. Due to the size of these matrices we use iterative methods to solve the discretization of (3.3) and (3.5).

The matrix C is symmetric and positive definite when the natural assumptions $\alpha > 0$ and K being invertible hold (we will assume so henceforth), thus the conjugate gradient can be applied to solve the discretization of (3.3). It can be shown that the matrix $\overline{C}$ is positive definite (i.e. its symmetrization $\tilde{C} = \frac{1}{2}(\overline{C} + \overline{C}^T)$ is positive definite) if $\alpha > 0$, K is invertible and $|w_i| \leq 1$ at all grid points (see [8]), but it is generally non-symmetric, thus the conjugate gradient method cannot be applied to it. Our approach is to replace $\overline{C}$ by its symmetrization $\tilde{C}$, i.e., use an approximate Newton's method. The symmetrized matrix $\tilde{C}$ can be regarded as a discretization of the operator:

$$-\alpha \nabla \cdot \left(\frac{1}{|\nabla u|} (I - \frac{1}{2} \frac{w \nabla u^T + \nabla u w^T}{|\nabla u|}) \nabla \right) + K^* K.$$

From the theory of the convergence of the approximate Newton's method, it follows that the convergence of the resulting iteration is quadratic, since $\tilde{C}$ converges locally to $\overline{C}$ (since $w \to \frac{\nabla u}{|\nabla u|}$), at least linearly, see [15, 5.4.1].

As mentioned above, the non-singularity of $\tilde{C}$ can be ensured if we impose the condition $|w_i| \leq 1 \; \forall i$ throughout the process. This can be achieved inexpensively by choosing a step length $s > 0$ such that:

$$s = \rho \sup\{\alpha\colon |w_i + \alpha \delta w_i| < 1 \, \forall i\}, \qquad \rho \in (0,1), \tag{5.3}$$

where it is assumed that $|w_i| < 1 \; \forall i$. Actually, our experience reveals that this choice of step length for w already produces a globally convergent algorithm, even for small β and without the use of a line search for the variable u. However, to ensure global convergence if (5.3) were not used, a step length procedure on the primal variables u must be used. Since $\tilde{C}$ is symmetric positive definite, $\delta u = -(\tilde{C})^{-1} g(u)$ is a descent direction for the (primal) objective function of (2.4) and also for the function $\|g(u)\|^2_{\mathcal{L}^2}$, where g is the gradient of the objective function as given in equation (2.9). It follows easily from the definition of $\tilde{C}$ that its smallest eigenvalue is the smallest eigenvalue of K^*K, which is strictly positive, by the assumption of the non-singularity of K, thus δu is a gradient related descent direction. Hence, a line search based on a sufficient descent of either one of the merit functions mentioned above could be used to ensure global convergence. We have found that using $\|g(u)\|^2_{\mathcal{L}^2}$ as merit function has given better results.

If we denote by ψ the chosen merit function, the line search procedure that we have used consists in a backtracking algorithm that selects as step length the first number s in the sequence of powers of 2 of decreasing exponent $1, \frac{1}{2}, \frac{1}{4}, \ldots$ that satisfies the inequality:

$$\psi(u + s\delta u) \leq \psi(u) + \eta s \nabla \psi(u)^T \delta u, \quad \eta \in (0,1). \tag{5.4}$$

In addition to the line search procedure, we may also use a continuation procedure on the parameter β to improve the global convergence behaviour of the algorithm. This procedure consists in selecting a sufficiently large value of β and gradually reducing it to the desired value β_{target}. The pseudo-codes for the continuation and the Newton's algorithms are given in Figures 5.1 and 5.2 respectively.

```
u = z, β_succ = β_curr = ||∇z||²_∞, ρ < 1
while β_curr > β_target
   [u_new,success]=newton(u, z, β_curr)
   if success, then
      β_succ = β_curr
      β_curr = β_curr * ρ % decrease β_curr
      ρ_old = ρ, choose ρ ∈ (0, ρ_old) % decrease ρ
      u = u_new
   else
      β_curr = β_succ % increase β_curr
      ρ_old = ρ, choose ρ ∈ (ρ_old, 1) % increase ρ
   end
end
```

FIG. 5.1. *Pseudo-code for continuation algorithm. β_{curr} is the parameter β for the problem to be solved at the current iteration, β_{succ} is the β used in the last succesful solution of a Newton's method, ρ is the reduction factor for β_{curr}. The initial choice of $\beta_{curr} = ||\nabla z||^2_\infty$ is based upon the heuristic that the domain of convergence of the Newton's method for a large β (relative to the term $|\nabla z|^2$ appearing in the formulation of the TV-norm) should be large.*

```
[u, success] = newton(u, z, β)
success=0
w = 0
for i=1:#iterations
   compute update direction (δu, δw)
   if line_search_on_u_requested, then
      compute s_p from (5.4)
   else
      s_p = 1
   end
   if line_search_on_w_requested, then
      compute s_d from (5.3)
   else
      s_d = 1
   end
   u = u + s_p δu
   w = w + s_d δw
   if stopping criterion is satisfied, then
      success=1
      return
   end
end
```

FIG. 5.2. *Pseudo-code for primal-dual Newton algorithm with possible line searches. The primal Newton algorithm is basically identical, with the exception that the dual variables w and dual updates δw do not appear. The variables line_search_on_u_requested and line_search_on_w_requested are global variables supplied by the user to use the corresponding line searches.*

Our first experiment consists in the comparison of the primal Newton and the primal-dual Newton methods under the following circumstances:

1. Continuation on β and no line search.
2. Continuation on β and line search on u.
3. No continuation on β and use line search on u.
4. No continuation and use line search on u and w for the primal-dual method.

The original image, which is 256×256 pixels and has dynamic range $[0, 255]$, appears in Fig 5.3. A Gaussian white noise with variance $\sigma^2 \approx 1200$ is added to it, resulting in the image displayed in Fig 5.4, with $SNR = \frac{||u-\overline{u}||_{\mathcal{L}^2}}{\sigma} \approx 1$. Fig 5.5 depicts the solution obtained by the primal-dual Newton method.

We have set the parameter α in the Tikhonov formulation to the inverse of the Lagrange multiplier yielded by a previous run of the constrained problem solver, in this case $\alpha = 1.18$. The parameter β has been set to 0.01. Furthermore, we have used *truncated* versions of Newton's algorithm, based on the conjugate gradient method with incomplete Cholesky as preconditioner. The stopping criterion for the (outer) Newton's iteration is a relative decrease of the nonlinear residual by a factor of 10^{-4}. The stopping criterion for the n-th inner linear iteration is a relative decrease of the linear residual by a factor of η_n, where we follow the suggestion of [15, Eq. 6.18] and set

$$\eta_n = \begin{cases} 0.1 \text{ if } n = 0, \\ \min(0.1, 0.9||g_n||^2/||g_{n-1}||^2) \end{cases} \tag{5.5}$$

where g_n denotes the gradient of the objective function, i.e. the right hand side of the discretization of (3.5) or (3.3), at the n-th iteration. In Table 5.1 we compare the primal-dual and the primal versions of Newton's method for the experiments described above.

The conclusions that can be drawn from this experiment are:

- The most crucial factor for the primal-dual method is controlling the dual variables via the step length algorithm appearing in (5.3). In fact, our experience is that this algorithm with the dual step length is globally convergent for the parameters α and β in a reasonable range. A line search for the primal variables almost always yields unit step lengths.
- The primal-dual method with the dual step length algorithm does not need continuation to converge, although using it might be slightly beneficial in terms of work.
- The primal-dual method with the dual step length has a much better convergence behavior than the primal method.

In our second experiment, we compare the primal-dual Newton, fixed point and time marching methods. We have implemented the fixed point ((2.11)) and the time marching ((2.10)) algorithms using the same standard discretization of the differential operators as for the primal-dual method. For this method, we use the step length algorithm for the dual variables, no continuation and the same parameters as in the previous experiment. These parameters are also used for the fixed point method, except that we have used a fixed linear relative residual decrease $\eta_n = 0.1$ (it is in this case *optimal* according to our experience). For the time marching method we have used a line search based on sufficient decrease of the objective function, in other words, an adaptive choice for the increment Δt to maintain the iteration stable. The stopping criterion for the time marching method is based on the iteration count since we have not been able to achieve the prescribed accuracy in a reasonable amount of time. In Figs. 5.6, 5.7 and 5.8 we plot the convergence history of this experiment.

The conclusions we draw from this experiment are:

- The primal-dual algorithm is quadratically convergent, whereas the others are at best linearly convergent.
- The primal-dual algorithm behaves similarly to the fixed point method in the early stages, but in a few iterations can attain high accuracy.
- The cost per iteration of the primal-dual method is between 30 and 50 per cent more than for the fixed point iteration. The memory requirements roughly satisfy this as well.

In the last experiment we have chosen the following parameters for the primal-dual method: same degraded image as in the previous experiments, $\alpha = 1$, $\beta = 0.01$ (no continuation), 10^{-8} as outer iteration relative tolerance, 10^{-4} as CG relative tolerance. We have then generated 50 random initial approximations z^i, $i = 1, \ldots, 50$, and obtained from the primal-dual algorithm with initial guess z^i a number of iterates $u_0^i = z^i, \ldots, u_{n_i}^i$. In all these applications of the algorithm we have obtained convergence and the number n_i of iterations has ranged from 14 to 17. Although we do not have a proof for the global convergence of this algorithm, these results strengthen our conjecture of the global convergence of the primal-dual method.

Acknowledgements. We would like to thank Andy Conn and Michael Overton for making their preprint [9] available to us, and Jun Zou for many helpful conversations on mixed finite elements with the first author. The third author wants to heartily thank Paco Arándiga, Vicente Candela, Rosa Donat and Antonio Marquina

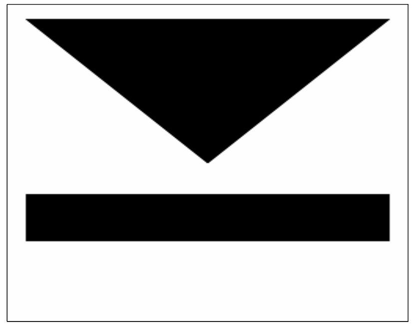

FIG. 5.3. *Original image,* 256×256 *pixels.*

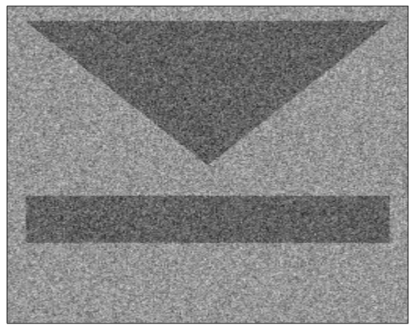

FIG. 5.4. *Noisy image, SNR*≈ 1, $||g|| = 2.07$.

for their continuous support.

REFERENCES

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

[2] L. ALVAREZ, P. LIONS, AND J. MOREL, *Image selective smoothing and edge detection by nonlinear diffusion II*, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.

[3] K. D. ANDERSEN, *Minimizing a sum of norms (large scale solution of symmetric positive definite linear systems)*, PhD thesis, Odense University, 1995.

[4] F. BREZZI, *A survey of mixed finite element methods*, in Finite elements, ICASE/NASA LaRC, Springer, 1988, pp. 34–49.

[5] P. H. CALAMAI AND A. R. CONN, *A stable algorithm for solving the multifacility localtion problem involving euclidean distances*, SIAM Journal on Scientific and Statistical Computing, 1 (1980), pp. 512–526.

[6] ———, *A second-order method for solving the continuous multifacility location problem*, in Numerical analysis: Proceedings of the Ninth Biennial Conference, Dundee, Scotland, G. A. Watson, ed., vol. 912 of Lecture Notes in Mathematics, Springer-Verlag, 1982, pp. 1–25.

[7] R. H. CHAN, T. F. CHAN, AND H. M. ZHOU, *Continuation method for total variation denoising problems*, Tech. Report 95-18, University of California, Los Angeles, 1995.

[8] T. CHAN, G. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, Tech. Report 95-43, UCLA, September 1995.

[9] A. R. CONN AND M. L. OVERTON, *A primal-dual interior point method for minimizing a sum of euclidean norms*. preprint, 1994.

[10] D. DOBSON AND F. SANTOSA, *Recovery of blocky images from noisy and blured data*, Tech. Report 94-7, Center for the Mathematics of Waves, University of Delaware, 1994.

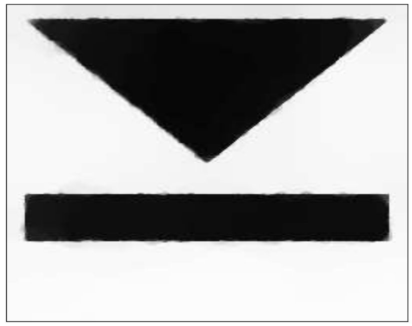

FIG. 5.5. *Denoised image, 13 Newton's iterations, 58 CG iterations,* $||g|| = 5.57 \times 10^{-5}$.

	primal-dual Newton		primal Newton	
	NWT	CG	NWT	CG
continuation, no line search	25	186	70	265
continuation, line search on u	25	187	53	210
continuation, line search on $u\&w$	13	51	53	210
no continuation, line search on $u\&w$	12	58	Not converged	

TABLE 5.1
Comparison of primal and primal-dual Newton methods.

[11] I. EKELAND AND R. TEMAM, *Analyse convexe et problèmes variationnels*, Dunod, Paris, 1974.
[12] E. GIUSTI, *Minimal Surfaces and Functions of Bounded Variations*, Birkhäuser, 1984.
[13] G. GOLUB AND C. VAN LOAN, *Matrix computations,* 2^{nd} *ed.*, The John Hopkins University Press, 1989.
[14] C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*, Pitman, Boston, 1984.
[15] C. T. KELLEY, *Iterative Methods for Linear and Nonlinear Equations*, vol. 16 of Frontiers in Applied Mathematics, SIAM, 1995.
[16] Y. LI AND F. SANTOSA, *An affine scaling algorithm for minimizing total variation in image enhancement*, Tech. Report 12/94, Center for theory and simulation in Science and Engineering, Cornell University, 1994.
[17] M. E. OMAN, *Fast multigrid techniques in total variation-based image reconstruction.* to appear in the Preliminary Proceedings of the 1995 Copper Mountain Conference on Multigrid Methods.
[18] P. PERONA AND J. MALIK, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intelligence, 12 (1990), pp. 629–639.
[19] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[20] G. SAPIRO AND A. TANNENBAUM, *Area and length preserving geometric invariant scale-space*, in Proc. 3rd European Conf. on Computer Vision, Stockholm, Sweden, May 1994, vol. LNCS, vol. 801, 1994, pp. 449–458.
[21] A. N. TIKHONOV AND V. Y. ARSENIN, *Solutions of Ill-Posed Problems*, John Wiley, New York, 1977.
[22] C. R. VOGEL, *A multigrid method for total variation-based image denoising.* in Computation and Control IV, conference proceedings to be published by Birkhauser.
[23] C. R. VOGEL AND M. E. OMAN, *Iterative methods for total variation denoising*, SIAM J. Sci. Statist. Comput., 17 (1996), pp. 227–238.

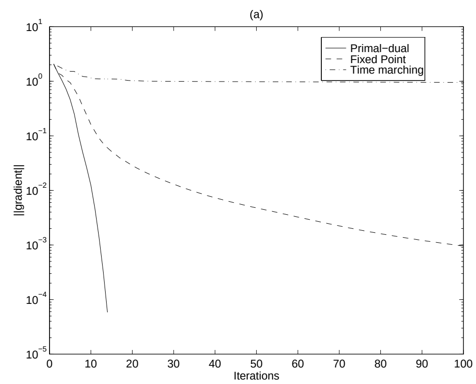

FIG. 5.6. *Plot of the $\mathcal{L}^2$-norm of the gradient $g(u)$ of the objective function versus iterations for the different methods.*

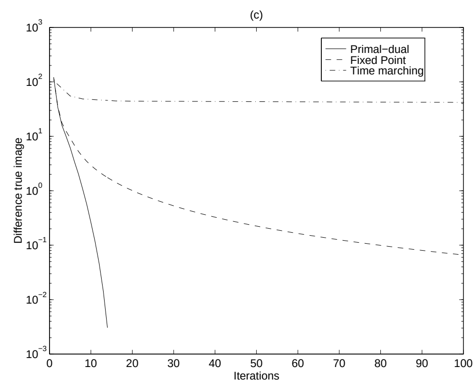

FIG. 5.7. *Plot of the $\mathcal{L}^2$-norm of the difference between the current iterate and the solution for the problem computed by Newton's method with high accuracy versus iterations for the different methods.*

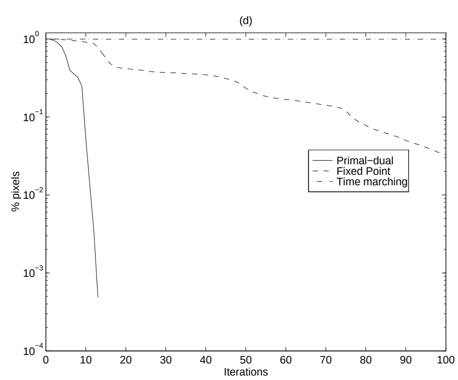

FIG. 5.8. *Plot of # pixels which differ more than .001 (relatively) from the solution for the problem computed by Newton's method with high accuracy versus iterations for the different methods.*