UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Continuation Method for Total Variation Denoising Problems

Tony F. Chan
H.M. Zhou
Raymond H. Chan

June 1995

CAM Report 95-28

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90024-1555

Continuation Method for Total Variation Denoising Problems

Tony F. Chan* H.M. Zhou† Raymond H. Chan‡

April 11, 1995

ABSTRACT

The denoising problem can be solved by posing it as a constrained minimization problem. The objective function is the TV norm of the denoised image whereas the constraint is the requirement that the denoised image does not deviate too much from the observed image. The Euler-Lagrangian equation corresponding to the minimization problem is a nonlinear equation. The Newton method for such equation is known to have a very small domain of convergence. In this paper, we propose to couple the Newton method with the continuation method. Using the Newton-Kantorovich theorem, we give a bound on the domain of convergence. Numerical results are given to illustrate the convergence.

Key Words: Denoising, Total-variation, Newton method, Fixed-point method.

1 Introduction

Noises are introduced in images in the formation, transmission or recording process. In this paper, we concern with the removal of noises in an image. Consider the model equation

$$u_0(x,y) = u(x,y) + \eta(x,y) \tag{1}$$

where $\eta(x,y)$ is a Gaussian white noise, $u_0(x,y)$ is the observed intensity function of the image and $u(x,y)$ is the original image. Our objective is to get a reasonable approximation of $u(x,y)$.

There are many different methods proposed to obtain an estimate of $u(x,y)$. In Rudin, Osher and

*Department of Mathematics, University of California, Los Angeles, CA, 90024. Research supported by grants NSF-ASC 92-01266 and ONR-N00014-92-J-1980.

†Department of Mathematics, Chinese University of Hong Kong, Hong Kong.

‡Department of Mathematics, Chinese University of Hong Kong, Hong Kong. Research supported by HKRGC grant no. CUHK 316/94E.

Fatemi,[1] they considered the constrained minimization problem

$$\min_u \int_\Omega |\nabla u| dx, \tag{2}$$

subject to

$$||u - u_0||^2 = \sigma^2. \tag{3}$$

Here Ω is a bounded convex region in the d-dimensional space, $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$, $||\cdot||$ denotes the norm in $\mathbf{L}^2(\Omega)$ and σ is the standard derivation of the noise $\eta(x, y)$.

Applying the Lagrange multiplier method to (2) and (3), we transform the problem to the following TV-penalized minimization problem:

$$\min_u \left\{ \int_\Omega |\nabla u| dx + \frac{\lambda}{2} ||u - u_0||^2 \right\}, \tag{4}$$

where $\lambda/2$ is the Lagrange multiplier. We note that $2/\lambda$ is the regularization parameter that controls the trade-off between the goodness of fit (3) and the variation or smoothness of the solution as required by (2). The Euler-Lagrange equation of (4) is given by:

$$\nabla \left(\frac{\nabla u}{|\nabla u|} \right) - \lambda(u - u_0) = 0 \tag{5}$$

with boundary condition

$$\frac{\partial u}{\partial n} = 0, \qquad x \in \partial\Omega. \tag{6}$$

Here $\partial\Omega$ is the boundary of Ω and n is the normal vector of $\partial\Omega$.

In actual computation, because of the singularity at $\nabla u = 0$, a small positive parameter β is added to the denominator of first term in (5). More precisely, we solve

$$\mathcal{L}(u) \equiv \nabla \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta}} \right) - \lambda(u - u_0) = 0 \tag{7}$$

with boundary condition (6). The least square functional (4) to be minimized is changed accordingly to

$$\mathcal{F}(u) = \int_\Omega \sqrt{|\nabla u|^2 + \beta} dx + \frac{\lambda}{2} ||u - u_0||^2. \tag{8}$$

To solve (7), Rudin *et. al.*[1] used artificial time marching technique. In particular, $u(x, y)$ is given as the steady-state solution to the following parabolic equation:

$$\frac{\partial u}{\partial t} = \mathcal{L}(u)$$

with boundary condition (6). The initial condition at time $t = 0$ is given by $u = u_0$, the observed image. In contrast, Vogel and Omen[3] used a relaxed fixed-point iteration to solve (7). More precisely, with the kth iterant u_k given, we compute the $(k+1)$th iterant by solving:

$$\nabla \left(\frac{\nabla u_{k+1}}{\sqrt{|\nabla u_k|^2 + \beta}} \right) - \lambda u_{k+1} = -\lambda u_0. \tag{9}$$

We note that this fixed-point iteration can be viewed as the semi-implicit time-marching scheme with infinite time step.

In Vogel and Omen,[3] the Newton method approach for the Euler-Lagrange equation (7) has also been considered. Let us denote $\mathcal{J}(u)$ to be the Jacobian of $\mathcal{L}(u)$. Then for any smooth function p, it is easy to show that

$$\mathcal{J}(u)p = \nabla\left(\frac{\nabla p}{\sqrt{|\nabla u|^2+\beta}}\right) - \nabla\left(\frac{\nabla p\cdot\nabla u}{(\sqrt{|\nabla u|^2+\beta})^3}\nabla u\right) - \lambda p. \tag{10}$$

The Newton method for (7) is

$$u_{k+1} = u_k - \mathcal{J}^{-1}(u_k)\mathcal{L}(u_k), \quad k = 0,1,\ldots. \tag{11}$$

Thus in each iteration, we have to compute $\mathcal{J}^{-1}(u_k)\mathcal{L}(u_k)$. Although we have no close-form formula for $\mathcal{J}^{-1}(u_k)$, we can still compute $\mathcal{J}^{-1}(u_k)\mathcal{L}(u_k)$ by using iterative methods such as the conjugate gradient (CG) method. This just requires the action of $\mathcal{J}(u_k)$ onto arbitrary smooth function p, which can be computed by (10).

We note that in (10) if we drop the higher order term (i.e. the middle term in the right hand side), and approximate the Jacobian by the first and third terms, then the resulting quasi-Newton method is the same as the relaxed fixed point method (9).

The numerical results in Vogel and Omen[3] suggests that the full Newton method (11) is divergent for β small. In this paper, we will see that the Newton method fails because it has a very small domain of convergence when β is small. Hence the initial guess should be closed to the true solution in order that the method converges. In order to acheive that, we employ the continuation method. In essence, we start the method with a large β, then we obtain the solution corresponding to this large β. For large β, the domain of convergence is large. Therefore the Newton method converges quadratically for reasonable initial guess such as the observed image. We use the solution for large β as the initial guess for the method with smaller β. We will see that this method can give convergence for arbitrarily small β. Using Newton-Kantorovich theorem, we will establish sufficient conditions that guarantee the convergence.

The outline of the paper is as follows. In §2, we give our Newton continuation algorithm. In §3, we analyze the dependence of the domain of convergence of the method on the parameters λ and β. Numerical results are given in §4.

2 The Method

Let us motivate our method by considering the following 1-dimensional problem: finding the zeros for

$$y = \frac{x}{\sqrt{x^2+\beta}} + \lambda(x - x_0), \tag{12}$$

when β is small, *cf.* (7). Figure 1 shows the graphs of y for different values of β when $\lambda = 0.5$ and $x_0 = 0.4$. We see that for β small, the domain of convergence of the Newton method will be very small. Thus if we start the Newton method for a small β, the method is divergent. However, we can easily find

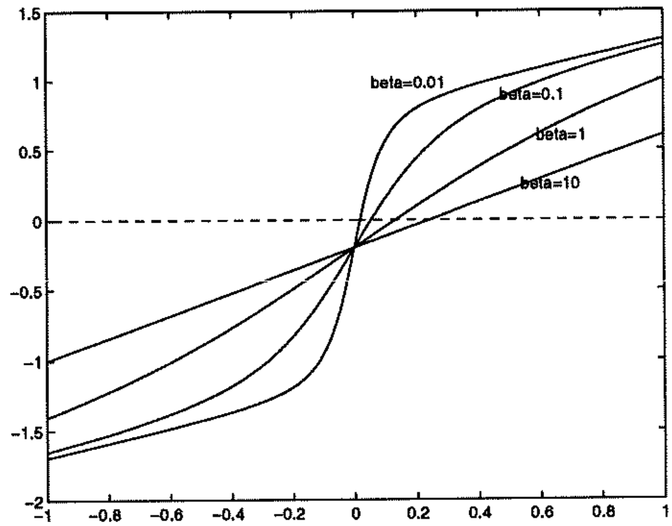

Figure 1. Graphs of function (12) with $\beta = 10, 1, 0.1$ and 0.01.

the root for any small β by using the continuation method. More precisely, we start the Newton method with $\beta = 10$ say. The method is convergent even when we are very far away from its solution. The solution to this problem should be a good initial guess for the Newton method for $\beta = 1$. The solution of the Newton method for $\beta = 1$ in turn will be a good initial guess for the problem with $\beta = 0.1$ and so on.

Our method for solving the modified Euler-Lagrange equation (7) is based on the same idea. Let β^* and λ^* be given positive numbers at which we want to solve (7). Our approach is to choose suitable large λ_1 and β_1, then fix one of them, say β_1 and decrease the other i.e. λ, towards the given value λ^*. Once we get the solution corresponding to β_1 and λ^*, then we fix λ^* and decrease β from β_1 towards β^*. For each given λ and β, we solve the solution by the Newton method (11). The update $\mathcal{J}^{-1}(u)\mathcal{L}(u)$ is obtained by the conjugate gradient method. This only requires the computation of $\mathcal{J}(u^k)p$ which can be evaluated by (10). To be more specific, we write down the algorithm below:

The Algorithm

(i) Set $k = 1$ and choose suitable large λ_1 and β_1. Use u_0, the observed image as the initial guess $u_{1,1}$.

(ii) While $\lambda_k \geq \lambda^*$, do

(a) Use $u_{k,1}$ as the initial guess.

(b) Use Newton method to find the solution of

$$\nabla \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta_1}} \right) - \lambda_k (u - u_0) = 0.$$

Denote the solution as $u_{k+1,1}$.

(c) Choose $\lambda_{k+1} < \lambda_k$.

(d) Set $k = k + 1$.

(iii) Let the final solution of Step (ii) be denoted by $u_{*,1}$. Set $l = 1$.

(iv) While $\beta_l \geq \beta^*$, do

(a) Use $u_{*,l}$ as the initial guess.

(b) Use Newton method to find the solution of

$$\nabla\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta_l}}\right) - \lambda^*(u - u_0) = 0.$$

Denote the solution as $u_{*,l+1}$.

(c) Choose $\beta_{l+1} < \beta_l$.

(d) Set $l = l + 1$.

(v) The final solution of Step (iv), denoted by $u_{*,*}$ will be the solution to the Euler-Lagrange equation (7) with the prescribed parameters λ^* and β^*.

Clearly one can construct another similar method: fixed λ first and varies β. Our numerical results indicate that it is not as good as fixing β first.

3 Convergence Results

Before we begin, let us introduce some notations. Let $\mathbb{R}^{n\times n}$ denote the set of all n-by-n matrices. For all $v, w \in \mathbb{R}^{n\times n}$, we define the discrete $\mathbf{L}^2$ inner product of v and w as

$$\langle v, w\rangle_0 = \frac{1}{n^2}\sum_{i,j=1}^{n} v_{ij}w_{ij}.$$

We will use ∇_1 and ∇_2 to denote the central difference operators in two different directions. More precisely,

$$\nabla_1 v_{ij} = \frac{v_{(i+1)j} - v_{(i-1)j}}{h}, \quad 1 \leq i \leq n-1,$$

and

$$\nabla_2 v_{ij} = \frac{v_{i(j+1)} - v_{i(j-1)}}{h}, \quad 1 \leq j \leq n-1.$$

Here $h = 1/n$. We set $\nabla_1 v_{1j} = \nabla_n v_{nj} = 0$ for all $1 \leq j \leq n$ and $\nabla_2 v_{i1} = \nabla_2 v_{in} = 0$ for all $1 \leq i \leq n$. We define the gradient of the matrix v at the (i, j)th entry to be

$$\nabla v_{ij} = (\nabla_1 v_{ij}, \nabla_2 v_{ij}), \quad 1 \leq i, j \leq n.$$

In our proof, we also require the discrete Sobolev $\mathbf{H}^1$ norm of v, which is defined as

$$||v||_1^2 \equiv ||v||_0^2 + ||\nabla_1 v||_0^2 + ||\nabla_2 v||_0^2.$$

In this matrix setting, the discretized Euler-Lagrange equation (7) becomes a matrix equation $L(u) = 0$, where the (i,j)th entry of the matrix $L(u)$ is given by

$$[L(u)]_{ij} = \nabla \left(\frac{\nabla u_{ij}}{\sqrt{|\nabla u_{ij}|^2 + \beta}} \right) - \lambda(u - u_0)_{ij} = 0. \tag{13}$$

We note that $L(u)$ can be viewed as a functional L_u from $\mathbb{R}^{n\times n}$ into $\mathbb{R}$. In fact, given any matrix $p \in \mathbb{R}^{n\times n}$, by the summation by part formula,

$$L_u(p) \equiv \langle L(u), p\rangle_0 = \frac{1}{n^2} \sum_{i,j=1}^{n} \left\{ -\frac{\nabla u_{ij} \cdot \nabla p_{ij}}{\sqrt{|\nabla u_{ij}|^2 + \beta}} - \lambda(u - u_0)_{ij} p_{ij} \right\},$$

where

$$\nabla u_{ij} \cdot \nabla p_{ij} \equiv \nabla_1 u_{ij} \nabla_1 p_{ij} + \nabla_2 u_{ij} \nabla_2 p_{ij}, \quad 1 \le i, j \le n.$$

Now we discretize the Jacobian operator $\mathcal{J}(u)$ in (10). Given the matrices $u, p \in \mathbb{R}^{n\times n}$, the discrete Jacobian operator $J(u)$ acting on p is a matrix defined by

$$[J(u)p]_{ij} \equiv \nabla \left(\frac{\nabla p_{ij}}{\sqrt{|\nabla u_{ij}|^2 + \beta}} \right) - \nabla \left(\frac{\nabla p_{ij} \cdot \nabla u_{ij}}{(\sqrt{|\nabla u_{ij}|^2 + \beta})^3} \nabla u_{ij} \right) - \lambda p_{ij}. \tag{14}$$

Similar to above, we can view $J(u)$ as a bilinear form on $\mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ into $\mathbb{R}$. More precisely, for all $p, q \in \mathbb{R}^{n\times n}$, $J_u(p,q)$ is defined as

$$J_u(p,q) \equiv \langle J(u)p, q\rangle_0 = \frac{1}{n^2} \sum_{i,j=1}^{n} \left\{ -\frac{\nabla p_{ij} \cdot \nabla q_{ij}}{\sqrt{|\nabla u_{ij}|^2 + \beta}} + \frac{(\nabla u_{ij} \cdot \nabla p_{ij})(\nabla u_{ij} \cdot \nabla q_{ij})}{\sqrt{|\nabla u_{ij}|^2 + \beta}^3} - \lambda p_{ij} q_{ij} \right\}.$$

We note that the operator norms of L_u and J_u are defined by:

$$|||L_u||| = \sup_{p \neq 0} \frac{|L_u(p)|}{||p||_1}$$

and

$$|||J_u||| = \sup_{p\neq 0, q \neq 0} \frac{|J_u(p,q)|}{||p||_1 ||q||_1}.$$

Clearly, L_u and J_u are bounded operators. In the following, for simplicity, we write

$$\diamond u_{ij} = \sqrt{|\nabla u_{ij}|^2 + \beta}.$$

The Newton method for the discrete Euler-Lagrange equation (13) is

$$u_{k+1} = u_k - J(u_k)^{-1} L(u_k), \quad k = 1, 2, \ldots \tag{15}$$

where $J(u)^{-1}$ is the inverse of the operator $J(u)$ as defined in (14). We note that we do not have a formula for $J(u_k)^{-1}$. However, the action of $J(u_k)$ onto any matrix $p \in \mathbb{R}^{n\times n}$ can easily be evaluated by (14). In our numerical tests, $J(u_k)^{-1}L(u_k)$ is computed by the conjugate gradient (CG) method. We now claim that the operator $J(u)$ is negative definite, hence the CG method is applicable.

LEMMA 1. *The Jacobian operator J_u is symmetric and negative definite with*

$$|||{J_u}^{-1}||| \le \frac{1}{\mu}$$

where

$$\mu = \min_{1\le i,j\le n} \left\{ \frac{\beta}{\diamond {u_{ij}}^3}, \lambda \right\}. \tag{16}$$

Proof: For all $p \in \mathbb{R}^{n\times n}$, we have

$$\begin{aligned} J_u(p,p) &= \frac{1}{n^2}\sum_{i,j=1}^{n} \left\{ -\frac{\nabla p_{ij}\cdot\nabla p_{ij}}{\diamond u_{ij}} + \frac{(\nabla p_{ij}\cdot\nabla u_{ij})^2}{\diamond {u_{ij}}^3} - \lambda|p_{ij}|^2 \right\} \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n} \left\{ \frac{-|\nabla p_{ij}|^2(|\nabla u_{ij}|^2+\beta) + (\nabla u_{ij}\cdot\nabla p_{ij})^2}{\diamond {u_{ij}}^3} - \lambda|p_{ij}|^2 \right\} \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n} \left\{ (-\frac{\beta|\nabla p_{ij}|^2}{\diamond {u_{ij}}^3} - \lambda|p_{ij}|^2) + \frac{-(|\nabla p_{ij}|^2|\nabla u_{ij}|^2) + (\nabla p_{ij}\cdot\nabla u_{ij})^2}{\diamond {u_{ij}}^3} \right\} \\ &\equiv A_1 + A_2. \end{aligned}$$

Obviously, $A_1, A_2 \le 0$. Hence

$$|J_u(p,p)| \ge -A_1 = \frac{1}{n^2}\sum_{i,j=1}^{n} \left\{ \frac{\beta|\nabla p_{ij}|^2}{\diamond {u_{ij}}^3} + \lambda|p_{i,j}|^2 \right\} \ge \mu||p||_1^2.$$

Therefore

$$|||J_u||| \le \frac{1}{\mu}. \qquad \square$$

We will use the following Newton-Kantorovich theorem to determine the domain of convergence for our method.

THEOREM 1. **(Newton-Kantorovich[2])** *Given a mapping $f : \mathbb{R}^m \to \mathbb{R}^m$ and a convex set $\mathbf{C} \in \mathbb{R}^m$, let D_f be the Jacobian of f and satisfy the conditions:*

(a) $||D_f(x) - D_f(y)|| \le \gamma||x-y||$ *for all* $x, y \in \mathbf{C}$,

(b) $||D_f^{-1}(x_0)f(x_0)|| \le \alpha$,

(c) $||D_f^{-1}(x_0)|| \le \eta$,

for some $x_0 \in \mathbf{C}$. *Consider the quantities*

$$\delta = \alpha\gamma\eta,$$

and

$$\rho = \frac{1-\sqrt{1-2\delta}}{\delta}\alpha.$$

If $\delta \leq 1/2$ *and the ball* $\mathbf{S}_\rho(x_0)$ *centered at* x_0 *with radius* ρ *is a subset of* $\mathbf{C}$, *then the sequence* $\{x_k\}$ *defined by the Newton method:*

$$x_{k+1} = x_k - D_f(x_k)^{-1} f(x_k), \quad k = 0, 1, 2, \ldots,$$

remains in $\mathbf{S}_\rho(x_0)$ *and converges to the unique zero of* $f(x)$ *in* $\mathbf{S}_\rho(x_0) \cap \mathbf{C}$.

In view of Lemma 1, we already have the bounds for the second and third conditions in Theorem 1. We now establish a bound for the first condition in Theorem 1.

LEMMA 2. *Let* $u, v \in \mathbb{R}^{n\times n}$, *then*

$$|||J_u - J_v||| \leq 2n\nu ||u-v||_1,$$

where

$$\nu = \max_{1\leq i,j\leq n}\left\{\frac{1}{\diamond u_{ij}\diamond v_{ij}}\right\} + \max_{1\leq i,j\leq n}\left\{\frac{1}{\diamond u_{ij}{}^2}\right\} + \max_{1\leq i,j\leq n}\left\{\frac{1}{\diamond v_{ij}{}^2}\right\}. \tag{17}$$

Proof: For all $p, q \in \mathbb{R}^{n\times n}$,

$$\begin{aligned}
& (J_u - J_v)(p,q) \\
= \; & \frac{1}{n^2}\sum_{i,j=1}^{n}\left\{-\nabla p_{ij}\nabla q_{ij}\left(\frac{1}{\diamond u_{ij}} - \frac{1}{\diamond v_{ij}}\right) + \left[\frac{\nabla u_{ij}\nabla p_{ij}\nabla u_{ij}\nabla q_{ij}}{\diamond u_{ij}{}^3} - \frac{\nabla v_{ij}\nabla p_{ij}\nabla v_{ij}\nabla q_{ij}}{\diamond v_{ij}{}^3}\right]\right\} \\
= \; & \frac{1}{n^2}\sum_{i,j=1}^{n}\left\{(\nabla p_{ij}\cdot\nabla q_{ij})\left[\frac{\nabla(u+v)_{ij}\cdot\nabla(u-v)_{ij}}{\diamond u_{ij}\diamond v_{ij}(\diamond u_{ij} + \diamond v_{ij})}\right]\right. \\
& \left. + \left[\frac{(\nabla u_{ij}\cdot\nabla q_{ij})(\nabla u_{ij}\cdot\nabla p_{ij})}{\diamond u_{ij}{}^3} - \frac{(\nabla v_{ij}\cdot\nabla q_{ij})(\nabla v_{ij}\cdot\nabla p_{ij})}{\diamond v_{ij}{}^3}\right]\right\} \\
\equiv \; & \frac{1}{n^2}\sum_{i,j=1}^{n}(B_1 + B_2).
\end{aligned}$$

We first note that $|\nabla(u+v)_{ij}| \leq |\diamond u_{ij} + \diamond v_{ij}|$ for all $1 \leq i,j \leq n$ and

$$\max_{1\leq i,j\leq n} \nabla(u-v)_{ij} \leq n||u-v||_1.$$

Therefore it is straightforward to show that

$$\frac{1}{n^2}\sum_{i,j=1}^{n}|B_1| \leq \max_{1\leq i,j\leq n}\left\{\frac{1}{\diamond u_{ij}\diamond v_{ij}}\right\} n||u-v||_1||p||_1||q||_1.$$

For B_2, we note that

$$B_2 = \frac{(\nabla u_{ij}\cdot\nabla p_{ij})}{\diamond u_{ij}{}^3}(\nabla q_{ij}\cdot\nabla(u-v)_{ij}) + \frac{(\nabla v_{ij}\cdot\nabla q_{ij})}{\diamond v_{ij}{}^3}(\nabla p_{ij}\cdot\nabla(u-v)_{ij}) + (\nabla u_{ij}\cdot\nabla p_{ij})(\nabla v_{ij}\cdot\nabla q_{ij})(\frac{1}{\diamond u_{ij}{}^3} - \frac{1}{\diamond v_{ij}{}^3}),$$

where the last term is equal to

$$-(\nabla v_{ij}\cdot\nabla q_{ij})\left\{(\nabla u_{ij}\cdot\nabla p_{ij})(\frac{(\nabla(u+v)_{ij}\cdot\nabla(u-v)_{ij})(\diamond u_{ij}\diamond v_{ij} + \diamond u_{ij}{}^2 + \diamond v_{ij}{}^2)}{(\diamond u_{ij}\diamond v_{ij})^3(\diamond u_{ij}+\diamond v_{ij})}\right\}.$$

Thus we have

$$|(J_u - J_v)(p,q)| \le 2n\nu\|u-v\|_1\|p\|_1\|q\|_1. \qquad \square$$

Combining Lemmas 1, 2 and Theorem 1, we obtain our main theorem.

THEOREM 2. *Let $L(u)$ be defined as in (13). For fixed λ and β, let u_1 be the initial guess and $\{u_k\}$ be the corresponding Newton sequence for solving $L(u) = 0$. Let $\|u_2 - u_1\|_1 = \alpha$ and μ, ν be given by (16) and (17) respectively. If*

$$\delta = \frac{\alpha n\nu}{\mu} \le \frac{1}{2},$$

then $u_k \in \mathbf{S}_\rho(u_1)$ for all k and converges to the unique solution of $L(u) = 0$ in $\mathbf{S}_\rho(u_1)$, where the radius of the ball $\mathbf{S}_\rho(u_1)$ is given by

$$\rho = \frac{1-\sqrt{1-2\delta}}{\delta\mu}.$$

4 Numerical Examples

In this section we present results of our denoising algorithms on two test images. The noisy image u_0 is obtained by adding random noise of level σ to the true image u. More precisely, we add random error to each pixel of the true image such that $\|u-u_0\|_0/\|u\|_0 = \sigma$. In the examples, we choose $\sigma = 0.3$.

Our first test example is a 32-by-32 pixel image. The original image and the noisy image are shown in Figure 2. In Table 1, the first two columns indicate the path we took in getting to the final prescribed λ and β. The initial guess for the first set of λ and β is chosen to be the observed image u_0. The initial guess u_1 at the other sets of λ and β is given by the optimal solution of its previous set of λ and β. In the table, we also give the numbers of Newton iterations (N) and the average numbers of the inner CG iterations (C) for different set of parameters λ and β. We note that the cost per one inner CG loop is of $O(n^2)$ operations, where $n = 32$ here. The tolerance for the Newton and CG methods are 10^{-7} and 10^{-10} respectively. In Table 1, we also give the residual $\|L(u_k)\|_0$, where u_k is the optimal solution for the given set of λ and β. The last column gives the difference between the initial guess u_1 at the current set of λ and β and the second iterant u_2 obtained after one Newton iteration (*cf.* Theorem 2).

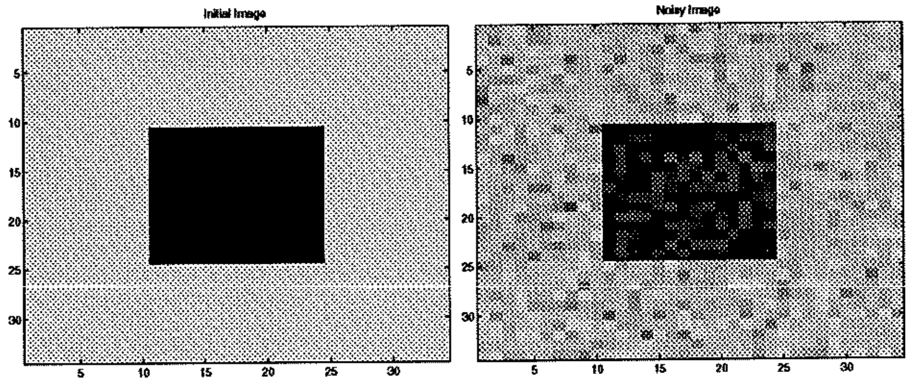

Figure 2. Original image (left) and noisy image with noise level $\sigma = 0.3$ (right).

λ	β	N	C	residual	$\|\|u_1 - u_2\|\|_0$
1000	20	4	9	2.0e-9	0.0468
400	20	5	13	1.0e-12	0.0412
150	20	5	23	2.1e-10	0.0446
75	20	4	29	4.0e-10	0.0367
75	10	4	32	1.6e-12	0.0078
75	2	5	54	2.1e-11	0.0189
75	0.5	5	77	1.6e-9	0.0132
75	0.1	6	119	2.0e-9	0.0121

Table 1. Number of iterations for Example 1.

In Figure 3, we show the denoised image with $\beta = 0.1$ and $\lambda = 75$. To emphasize that our method always converges for arbitarily small positive β and λ, we also show in Figure 3 the denoised image for $\beta = 10^{-6}$ and $\lambda = 75$. We observe that the denoised image is often quite good visually even for reasonably large β and λ. However, it will still be good to have a quadratically convergent method which allows us to get fast convergence for arbitrarily small β and λ.

In Figure 4, we show the original image and the noisy image of our second example, which is a 64-by-64-pixel image. The convergence results are listed in Table 2. The denoised image is in Figure 5.

We finally remark that the linear solves in the inner loops are done with CG with no preconditioning. However, our continuation method can be used with any linear solvers or with any preconditioners. The use of specific preconditioners to speed up the inner loop CG method will be discussed in our future work.

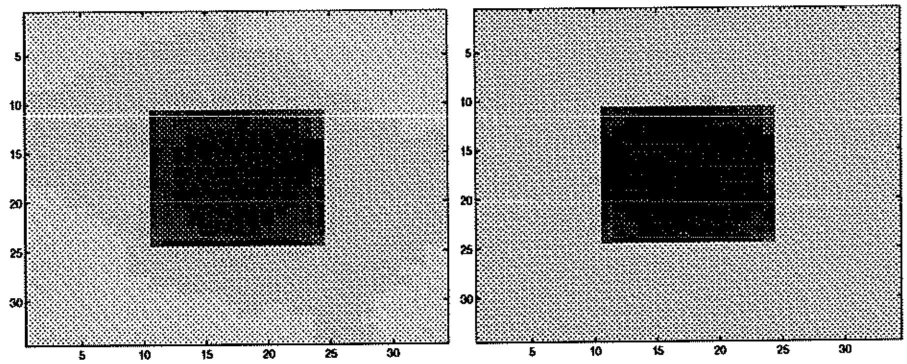

Figure 3. Restored image with $\lambda = 75$ and $\beta = 0.1$ (left) and $\beta = 10^{-6}$ (right).

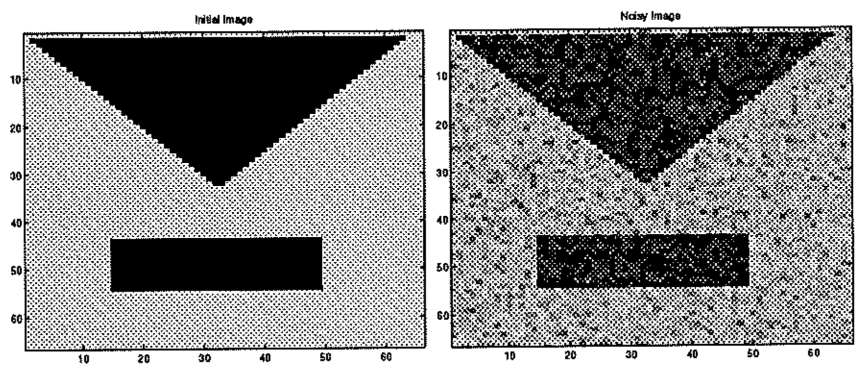

Figure 4. Original image (left) and noisy image with noise level $\sigma = 0.3$ (right).

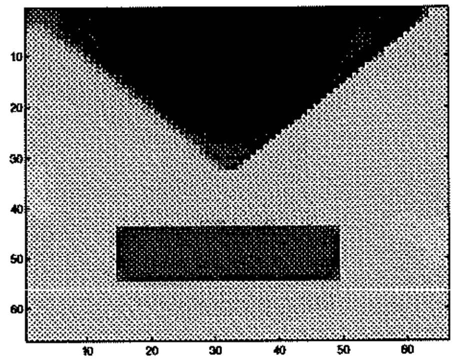

Figure 5. Restored image with $\lambda = 100$ and $\beta = 0.1$.

λ	β	N	C	residual	$\|\|u_1 - u_2\|\|_0$
1200	20	7	16	1.0e-12	0.0929
600	20	6	20	2.3e-10	0.0554
400	20	6	22	1.0e-12	0.0261
250	20	5	37	6.2e-8	0.0231
150	20	5	43	3.0e-12	0.0259
100	20	4	59	1.3e-9	0.0269
100	10	4	66	1.6e-10	0.0081
100	2	6	110	6.1e-10	0.0184
100	0.5	6	157	8.7e-9	0.0121
100	0.1	12	289	1.3e-11	0.0113

Table 2. Number of iterations for Example 2.

5 Acknowledgment

The first author would like to acknowledge the hospitality of the Department of Mathematics at the Chinese University of Hong Kong where this work was initiated during a visit.

6 REFERENCES

[1] L. Rudin, S. Osher and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, 60 (1992) 259–268.

[2] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1980.

[3] C. Vogel and M. Oman, *Iterative Methods for Total Variation Denoising*, Preprint.