UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Iterative Methods for Total Variation Image Restoration

Tony F. Chan
Pep Mulet

October 1996
CAM Report 96-38

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

To appear in Scientific Computing, Springer-Verlag, 1996, Proc. of Winter School on Iterative Methods in Scientific Computing, Chinese University of Hong Kong, December 95.

Iterative Methods for Total Variation Image Restoration [1]

Tony F. Chan[2] and Pep Mulet[3]

1 Introduction

Image processing refers to the analysis and extraction of information from images. Some example of the tasks are: restoration, compression and segmentation [21]. Applications can be found in many areas, including medical diagnosis, satellite surveying and surveillance.

Since the number of pixels for even a two dimensional image with modest resolution often exceeds several hundred thousands, image processing involves many computationally intensive tasks. The requirements for real time response and multi-frame/band analysis add further to the need for fast and efficient computational algorithms.

Traditionally, the standard methods involve computation in the frequency domain [21], facilitated by efficient FFT (and more recently wavelet) algorithms. In the last few years, there has been a new movement towards a more PDE-based approach; see for example the recent review article by Alvarez and Morel [1] and the books by Morel and Solimini [28] and ter Haar Romeny [34]. The new models are motivated by a more systematic approach to restoring images with sharp edges, as well as for image segmentation. The image is diffused (denoised) according to a nonlinear anisotropic diffusion PDE, designed to diffuse less near edges [29]. The PDEs are often designed to possess certain desirable geometrical properties such as affine invariance and causality. Total variation based image restoration methods belong to this new class of models.

From a computational standpoint, the PDE formulations leads to computational problems which are often of a different nature from the traditional frequency domain and algebraic approach and calls for new computational techniques which can better exploit the fundamental nature of the governing nonlinear PDEs. As yet, the nonlinear diffusion models are considered somewhat expensive compared to traditional methods and naturally a part of the research in this area is directed towards making them more efficient while retaining and improving their desirable geometric properties.

In this chapter, we shall consider the use of iterative methods for total variation models for image restoration. Iterations are needed for solving the nonlin-

[1]This work has been supported in part by the ONR under contract N00014-96-1-0277 and the NSF under contract DMS-9626755.

[2]Dept. of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA. Email: chan@math.ucla.edu. URL: http://www.math.ucla.edu/~chan.

[3]Departament de Matemàtica Aplicada, Univ. de Valencia, Dr. Moliner, 50, 46100 Burjassot, Spain. Email: mulet@godella.matapl.uv.es. This author was also supported by Spanish DGICYT grants EX94 28990695 and PB94-0987.

ear problem, as well as for the linear algebra problems which arise at each step. Analogous to the situation for solving large discretized PDEs in several dimensions, the size of the problems are large enough that direct solution methods are too costly and iterative methods can be much more efficient.

The goal of this chapter is two fold. First, we would like to give a concise and self-contained overview of the total variation image restoration models and some existing numerical methods for solving them. Second, we would like to review briefly some of our own research in this area. Specifically, we will discuss two topics: the use of primal-dual Newton methods for solving a minimization problem involving the highly nonlinear total variation functional, and preconditioning techniques for differential-convolution type operators.

2 Image restoration

The recording of an image usually involves a degradation process: a *blurring*, due to atmosphere turbulence, camera misfocus or relative movement, followed by a random *noise*, due to errors of the physical sensors or to quantization. The actual degradation model will depend on many factors, but a commonly used model is that of a linear blurring operator and additive Gaussian white noise, which is the one we will consider in this paper. Others may involve nonlinear blurring operators, multiplicative noise, noise with more complicated distributions and with possible correlation with the image.

The aim of image restoration is the estimation of the ideal true image from the recorded one. The *direct problem* of computing the imaging system response (blurred image) from a given image is often assumed to be known and *well-posed*. The usual model for it is the convolution by a given kernel or *point spread function (PSF)*, which, in most of the cases, implies that the *inverse problem* of computing the true image from the observations is an *ill-posed problem*. A general principle for dealing with the instability of the inverse problem is that of *regularization*, which mainly consists in restricting the set of admissible solutions and including some *a priori information* (non negativity, smoothness, existence of edges, etc.) in the formulation of the problem. Both the accurate modeling of the imaging system and the choice of regularization will be essential for a satisfactory image restoration process.

Throughout this paper, the word *image* will refer to a real function u, defined on $\Omega = [0,1]^2$, with $u(x,y)$ being the intensity or grey level at pixel (x,y).

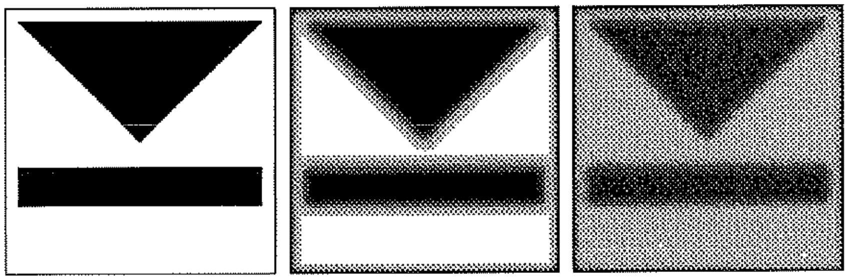


Figure 1: From left to right: original image, blurred image, observed image

2.1 Degradation model

Let u denote the ideal image to be estimated and z the observed image. We model the imaging system as

$$z = Ku + n,$$

where K is a linear operator (the blur operator) and n is a random field (noise), of which we may know some of its statistics (see Fig 1).

The operator K is usually a convolution by a known point spread function h, i.e.,

$$Ku(x,y) = \int_\Omega h(x-\overline{x}, y-\overline{y})u(\overline{x},\overline{y})\, d\overline{x}\, d\overline{y},$$

corresponding to a *spatially invariant* blurring process. Another interesting case, which can be seen as a special case of the latter, is when the operator K is the identity, which corresponds to *denoising*.

The noise n is assumed to be a *Gaussian white noise*, i.e., the values $u(x,y)$ are uncorrelated random variables with a normal distribution with mean 0 and variance σ^2, for all $(x,y) \in \Omega$.

2.2 Regularization

The inverse problem $Ku = z$ is ill-posed nature in most cases due to the compactness of the convolution operators with square integrable kernels. This means that the problem has no solution and/or this solution does not depend continuously on the data z. Since this data will contain measurements errors, direct solutions of the inverse problem, in case of existence, will be useless. We thus need to modify this inverse problem in order to turn it into a well-posed problem. This can be achieved by *regularizing* it. We can cite as regularization procedures the following (more references can be found in the survey paper [24]):

- **Direct Regularization Methods.**

  - Tikhonov Regularization.
  - Truncated matrix factorizations, e.g. SVD and RRQR.
  - Mollifier Methods.

- **Iterative Regularization Methods.**

  - Lanczos and Conjugate Gradient regularization.

Some of these methods require parameter specifications. We cite some parameter choice methods:

- Discrepancy Principle.
- Generalized Cross-Validation.
- L-curve.
- Methods based on Error Estimation.

2.3 Tikhonov Regularization

We will focus on the paradigm of Tikhonov-like regularization in this paper. We consider two variants:

1. **Noise level constrained formulation**: it consists in the solution of the constrained optimization problem

$$\min_u R(u)$$

$$\text{subject to} \quad \|Ku - z\|^2 = \sigma^2,$$

   where R is a functional, the *regularization functional*, that measures the *irregularity* of u. The quadratic constraint acts as the restriction of the solution set, since $\|Ku - z\|^2 = \|n\|^2 \approx \sigma^2$, and the functional R incorporates the *a priori* information of the image. This approach is equivalent to the discrepancy principle.

2. **Unconstrained formulation**: if there is no good estimate of the variance of the noise, then we may consider the unconstrained optimization problem

$$\min_u f(u) \equiv \frac{1}{2}\|Ku - z\|^2 + \alpha R(u),$$

   where α controls the tradeoff between a good fit to the data and an irregular solution. It can be shown that ideally α should be chosen to be the reciprocal of the Lagrange multiplier for the previous problem.

The regularization functionals used more often are quadratic functionals of the type $R(u) = ||Qu||^2$ where $Q = I$ (the identity operator) or $Q = \nabla$ (the H^1 semi-norm). This choice essentially gives a linear least squares problem, but has the drawback of penalizing discontinuities in u. Therefore, this is not a good choice if we are interested in edge restoration.

2.4 Algorithms for Tikhonov regularization with quadratic regularization functionals

If we consider a quadratic regularization functional $R(u) = ||Qu||_2^2$, with a linear functional Q, then we can write

$$f(u) = \alpha||Qu||_2^2 + ||Ku - z||_2^2 = \left\| \begin{bmatrix} K \\ \sqrt{\alpha}Q \end{bmatrix} u - \begin{bmatrix} z \\ 0 \end{bmatrix} \right\|_2^2 = ||Au - f||_2^2.$$

Thus the problem reduces to an ordinary least squares problem. In order to improve efficiency, one must try to exploit the Toeplitz (or block-Toeplitz) structure of $A^T A$. Since A can be very large (if the picture contains $n \times n$ pixels, then K is $n^2 \times n^2$) we need efficient algorithms.

2.4.1 Direct methods

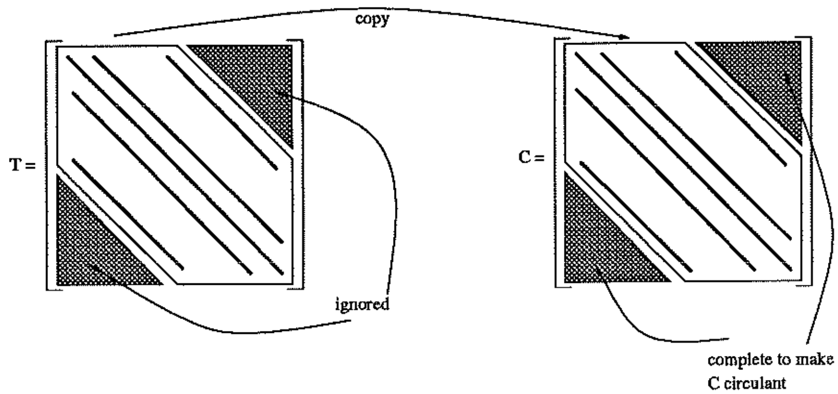
Fast direct methods for structured least squares problems is an actively researched area. The classical Levinson and Schur algorithms [20] can reduce the complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ or $\mathcal{O}(n \log^2 n)$. The stability of these fast and super-fast algorithms are still not fully understood [4]. More recently, fast algorithms based on Gaussian-elimination with partial pivoting for Cauchy matrices and matrices with displacement structures have been developed [25, 19].

2.4.2 Iterative methods: Circulant Preconditioners

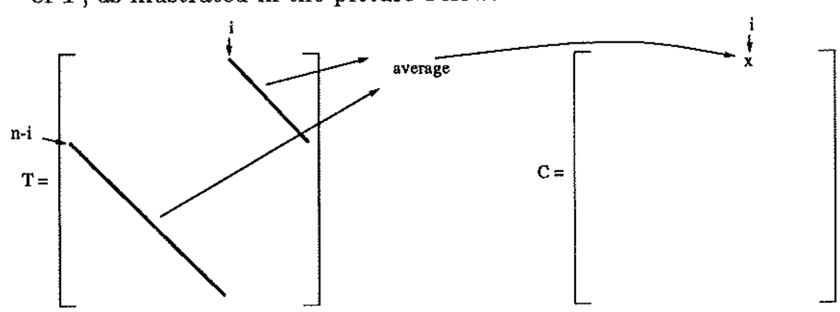
An alternative to direct methods is preconditioned iterative methods. This approach was first proposed by Strang [33]. The idea is the following. The conjugate gradient algorithm on the normal equations $A^T Au = A^T z$ has the advantage of only needing the products Av and $A^T v$. If K, Q are Toeplitz matrices then efficient algorithms can be employed to compute Av, $A^T v$, for instance, by embedding these matrices into circulant ones and then using FFT's. However, the convergence rate will be governed by the condition number of A and hence preconditioning will be essential for the feasibility of this approach.

Some preconditioners based on circulant matrices have been proposed:

- [33, 5]: It consists in using a circulant matrix C to approximate a Toeplitz matrix T by first copying the main diagonals of T to C and then completing C to form a circulant matrix. This is illustrated in the fashion outlined in the following picture.



- [14]: This is a variational approach. For a given matrix T (not necessarily Toeplitz), it consists in finding the circulant matrix C which minimizes the Frobenius norm of the difference $C-T$, i.e., $C = \underset{C \text{ circulant}}{\text{argmin}} \|T-C\|_F$. The solution amounts to averaging the entries in the corresponding diagonals of T, as illustrated in the picture below.



This preconditioner has the following advantages:

  - It preserves SPD matrices.
  - It is applicable to general A, not just Toeplitz T.

- Many other variants have been proposed (cf. the recent survey [13]). There is also a vast literature on convergence theory for fast transform preconditioners:

  - For Toeplitz matrices in the Wiener Class (i.e. whose entries in a row is absolutely summable), the spectrum of $M^{-1}A$ clusters around 1 [9], [26], [3].

- For elliptic operators, the condition number of the preconditioned matrix, $\mathcal{K}(M^{-1}A)$, can be bounded by $\mathcal{O}(1)$ independent of h, if the type of boundary conditions of the preconditioner matches those of the elliptic operator, and $\mathcal{O}(n)$ otherwise [6], [17], [10].

For applications of some of these preconditioning techniques to Toeplitz least squares problems, see [8, 7].

More recently, an idea that has received a lot of interest is to use early truncation of the conjugate gradient method applied to the normal equation for $Ku = z$ as a regularization procedure [23].

3 Total Variation Restoration

3.1 Total Variation Regularization

The main disadvantage of using quadratic regularization functionals such as the H^1 semi-norm of u is the inability to recover sharp discontinuities (edges). Mathematically, this is clear because discontinuous functions do not have bounded H^1 semi-norms. To remedy this, Rudin, Osher and Fatemi proposed in [31] the use of the *Total Variation*

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

as a regularization functional. The intuition for the use of this functional is that it measures the *jumps* of u, even if it is discontinuous. It is also connected to minimal surface problems.

The following result illustrates the advantage of the use of the Total Variation versus the quadratic functional given by the H^1-semi-norm for a simplified 1D problem (cf. Fig 2)

Lemma. Let $S = \{f | f(0) = a, f(1) = b\}$.

1. For the total variation norm, we have that $\min_{f \in S} TV(f) = b - a$ is achieved by any monotone, not necessarily continuous, $f \in S$.

2. For the H^1 semi-norm, we have that $\min\limits_{f \in S} \int_0^1 (f'(x))^2\, dx$ has the unique solution $f(x) = a(1-x) + bx$.

Thus, the total variation norm allows more functions, including discontinuous ones, to approximate a given noisy function, whereas the H^1 semi-norm "prefers" the linear function over others.

To illustrate the suitability of Total Variation denoising we depict a comparison of several signal denoising algorithms in Fig. 3 .

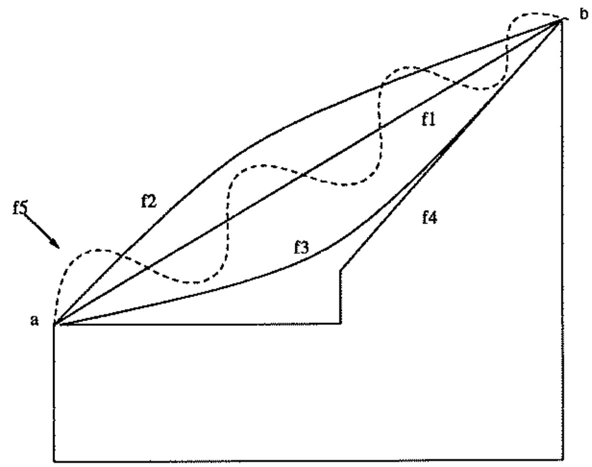


Figure 2: $TV(f_5) > TV(f_1) = TV(f_2) = TV(f_3) = TV(f_4)$

For simplicity, we will consider the unconstrained formulation of the Tikhonov regularization. The problem is then:

$$\min_u f(u) \equiv \alpha TV(u) + \frac{1}{2}\|Ku - z\|_2^2, \tag{1}$$

which is a strictly convex problem. To minimize $f(u)$, we will need its gradient. By using integration by parts with Neumann boundary conditions on (1), it can be shown that the gradient of the Total Variation functional is

$$\nabla TV(u) = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right).$$

Thus, the first order optimality condition for the problem (1) is given by:

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

This is the problem that we need to solve to find the restored image u.

3.2 Computational challenges

The solution of (2) poses several computational difficulties:

- The problem size can be very large, due to the large number of pixels, even for normal resolutions. This is even worse if 3D or video images are considered. The use of iterative methods for linear systems is necessary for most of the cases.

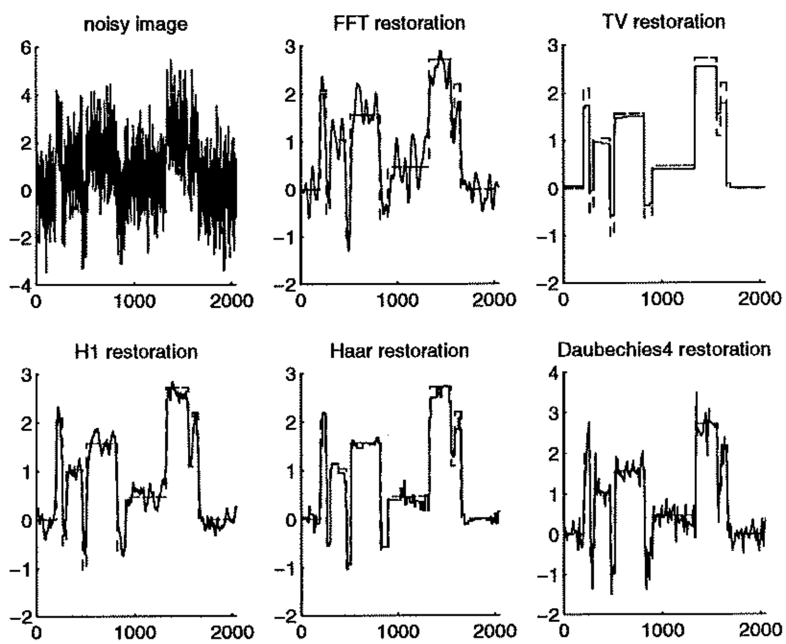


Figure 3: Original image plotted with dashed line. The Fourier, Haar and Daubechies plots correspond to truncated spectral decompositions and the H^1 restoration to Tikhonov regularization using the H^1-semi-norm as functional.

- The TV functional is non-differentiable at locations where $|\nabla u| = 0$, which often happens in real images. This makes necessary the use of numerical regularization, consisting in replacing the term $|\nabla u|$ by $\sqrt{|\nabla u|^2 + \beta}$ for a *small enough* positive artificial parameter β.
- The operator $\nabla \cdot \left(\frac{\nabla}{|\nabla u|}\right)$ is highly nonlinear.
- The ill-conditioning of the operators $\nabla \cdot \left(\frac{\nabla}{|\nabla u|}\right)$ (when linearized, it is a second order elliptic operator) and K^*K (it is a convolution operator) can cause numerical difficulties.
- When linearized, the elliptic operator $\nabla \cdot \left(\frac{\nabla}{|\nabla u|}\right)$ has highly varying coefficients, due to the presence of the term $\frac{1}{|\nabla u|}$.
- If $K \neq I$, then one needs to efficiently precondition the sum of an elliptic operator and a blurring operator. While good preconditioning techniques are available for the individual operators, preconditioning the sum is much more difficult. We will consider this issue in Section 5.

3.3 TV Restoration Algorithms

There are several iterative schemes which have been proposed in the literature for solving the optimality equation (2).

- The *Time Marching* scheme proposed in Rudin, Osher and Fatemi [31, 30] consists in finding the steady state of the parabolic PDE

$$u_t = -g(u), \ u(0) = z.$$

  It is equivalent to the steepest descent method. It can use the *projected gradient* method to handle noise constraints. The drawback of this method is that explicit methods can be slow due to stability constraints. Implicit methods can also be applied but then one has to deal with the nonlinearity and the solution of the resulting linear systems.

- In Vogel [35], the *Fixed Point Lagged Diffusivity Iteration* is introduced. This method consists in linearizing the nonlinear differential term in (2) by lagging the diffusion coefficient $1/|\nabla u|$ behind one iteration. Thus, u^{n+1} is obtained as the solution to the linear integro-differential equation

$$(-\alpha \nabla \cdot \left(\frac{\nabla}{|\nabla u^n|}\right) + \ K^*K)u^{n+1} = K^*z.$$

This algorithm can be interpreted within the framework of *generalized Weiszfeld's methods*, as introduced in [37]. As proved in that paper, this method is monotonically convergent, in the sense that the objective function evaluated at the iterates forms a monotonically descreasing sequence, and that the convergence rate is linear. In practice, this method is very robust.

- In Vogel and Oman [36] and Chan, Chan and Zhou [12], *Newton-type methods* are considered. Unfortunately, the domain of convergence of Newton's method applied to solve equation (2) is extremely small, specially when the regularizing parameter β is small. A *continuation method* is needed to make this method work efficiently. However, controlling the continuation process is not easy and many heuristic must be used to make it an overall efficient procedure.

4 A Primal-dual Method for Total Variation Minimization

The difficulty of the linearization of (2) is caused by the singularity of the square root that appears at the denominator. In this section, we briefly describe a special linearization technique that we have developed in [15] which has proven to be very efficient and robust compared to the methods mentioned earlier.

4.1 A linearization based on a dual variable

The key idea is to introduce a smooth auxiliary (dual) variable:

$$w \equiv \frac{\nabla u}{|\nabla u|}.$$

Note that $|w| = 1$. Then we rewrite (2) as a system of equations in the (w, u) variables:

$$\left.\begin{array}{rcl} f(w,u) = & w|\nabla u| - \nabla u & = 0 \\ g(w,u) = & -\alpha\nabla\cdot w + K^*Ku - K^*z & = 0 \end{array}\right\}$$

and linearze it instead of the u-system:

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

This approach is similar to the primal-dual method of [16, 2] for minimizing a sum of Euclidean norms.

We cite in the following some heuristic explanations for the advantages of this technique:

- The vector $\frac{\nabla u}{|\nabla u|}$ is the normal vector to the level sets of u (cf. Fig 4). Thus, w is smooth and well-defined if the level sets are smooth whereas if we linearize it we will lose this property. Moreover, $\nabla \cdot w = \kappa$ is the curvature of the level sets of u and the time marching method can be regarded as a curvature-driven flow. Hence, w is a natural variable to use in a geometry-based image restoration algorithm.

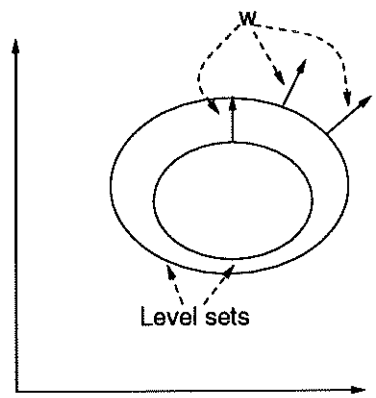


Figure 4: w variable is the normal to the level sets of u.

- The better global convergence behavior of Newton's method for (w, u)-system is due to the fact that this system is more *globally linear* than the u-system. To back up this assertion, we consider the following scalar example: we use Newton's method to find the solution of $f(x) = w - \frac{x}{\sqrt{x^2+\beta}} = 0 \equiv$ $g(x) = w\sqrt{x^2+\beta} - x = 0$ for $w = 0.9999, \beta = 1e-4, x \equiv \nabla u$ (cf. Fig 5). We find that Newton's method applied to $g(x) = 0$ converges for any initial guess > 0 and applied to $f(x)$ diverges for any initial guess > 2. Note also that the graph of $g(x)$ resembles a linear function in almost all the horizontal axis, whereas the graph of $f(x)$ resembles a rational function.

The linearization of the u-system yields

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

and that of the (w, u)-system yields:

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

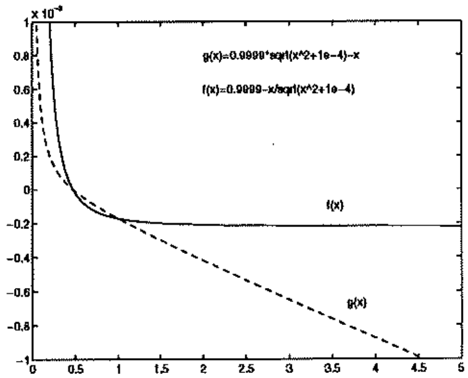


Figure 5: Plot of $f(x)$ and $g(x)$. The vertical axis has been scaled by 10^{-3}.

which after eliminating the δw variable yields:

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

Note that the cost of solving the (w,u)-system is comparable to that for the u-system. The main cost in solving the system in δu in both cases.

4.2 Dual definition for TV norm

The definition of the Total Variation as

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

is not well-defined if u is non-smooth. An equivalent dual definition that is well-defined for non-smooth u is the following (cf. [18]):

$$TV(u)=\max\{\int_\Omega u\,\nabla\cdot w\,dx\,dy\colon w=(w_1,w_2),\ w_i\in C^\infty(\Omega),\ |w|_\infty\le 1\}.$$

Hence, another formulation for the TV restoration problem is:

$$\min_u\max_{|w|\le 1}\ \alpha\int_\Omega u\,\nabla\cdot w\,dx\,dy+\frac{1}{2}||Ku-z||^2.$$

It can be shown that this leads to an equivalent primal-dual formulation.

4.3 Numerical Results for the Primal-Dual Method

We will now compare numerically the convergence behavior of the primal-dual method to that of the primal-Newton method under several circumstances. The continuation procedure used for both algorithms is based on a trial and error strategy, together with backtracking heuristics to minimize the work caused by possible errors. The line search on the u variable is based on sufficient decrease for the L^2-norm of the gradient of the objective function. The line search on the w-variable is a simple and inexpensive algorithm to ensure that $|w(x,y)| < 1$ throughout the iteration, by taking a fraction of $\inf_{(x,y)\in\Omega}(\sup\{s\colon |w(x,y) + s\delta w(x,y)| < 1\})$ (see [15] for details).

We show in Table 1 the work required by both algorithms to find the solution of the Total Variation restoration problem for the noisy image (K = identity) displayed in Fig 6, obtained from the true image, also displayed in Fig 6, by adding a random noise whose distribution is normal of zero mean and variance $\sigma^2 \approx 1200$, resulting in a SNR$\approx$ 1. The target parameter β we have used is 0.01×256^2. The parameter α has been set to 1.18, which has been computed from the Lagrange multiplier obtained from the solution of the constrained problem. The stopping criterion for the intermediate Newton's iterations, i.e., during the continuation, before solving the problem for the target β, has been to achieve a relative decrease of 10^{-2} for the L^2-norm of the gradient of the objective function. The relative tolerance for the last subproblem has been 10^{-4}.

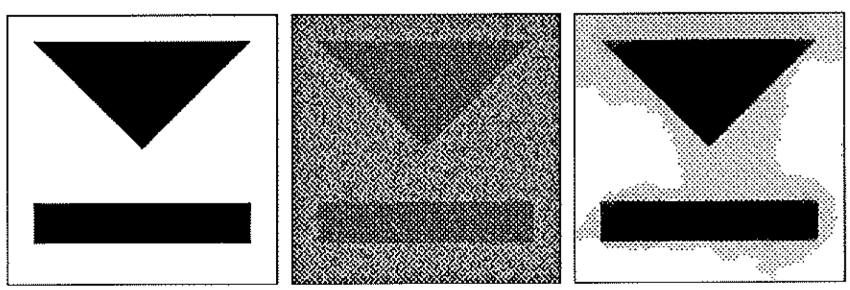


Figure 6: Original, noisy and denoised image, SNR$\approx$ 1.

We use the conjugate gradient method preconditioned by the incomplete Cholesky decomposition as linear solver. We have used an adaptive relative tolerance on the Euclidean norm of the residual as recommended in [27, Eq. 6.18]

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

In order to be able to apply the conjugate gradient method to the (w,u)-system, we have replaced its Jacobian by its symmetric part, which is an SPD matrix.

Furthermore, this Jacobian converges to a symmetric matrix and therefore the convergence is still locally quadratic by results in [27].

In Table 1, under the columns headed by NWT, we have the total number of Newton iterations required to solve the problem and under the columns headed by CG, we have the total number of (inner) CG iterations, i.e., matrix vector multiplications, including all iterations incurred by the backtracking procedure.

	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 1: NWT≡# nonlinear iterations,CG≡# CG iterations

The conclusions that can be drawn from this experiment are:

- The most crucial factor for the primal-dual method is controlling the dual variable w via the step length algorithm mentioned above. In fact, our experience is that this algorithm with the dual step length control 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 control 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 control 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 for a picture of a satellite contaminated with artificial noise [4]. For the primal-dual Newton method, we use the step length algorithm for the dual variables, no continuation and the same parameters as in the previous experiment. The same parameters are 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). We have used a line search based on sufficient decrease for the time marching method. The stopping criterion for the time marching method is based on the iteration count since we

[4] given to us by Prof. Bob Plemmons of Wake Forest University

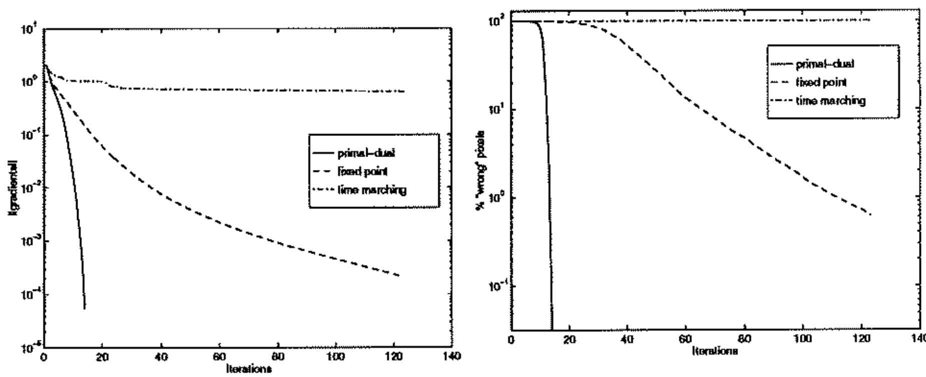


Figure 7: Convergence comparison of primal-dual, fixed point and time-marching: at the left we plot the Euclidean norm of the gradient of the Total Variation restoration functional versus iterations and at the right we plot the percentage of pixels that differ by more than 1 pixel value(the *pixel tolerance*, since the dynamic range of the image is 256) from a pre-computed high accuracy solution.

have not been able to achieve the prescribed accuracy in a reasonable amount of time. In Fig. 7 we plot the convergence history of this experiment and in Fig. 8 we display some of the intermediate results for the primal-dual and fixed point methods.

The conclusions we draw from this experiment are:

- The primal-dual algorithm is quadratically convergent, as can be seen from the left plot in Fig. 7, 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.
- Both the primal-dual and fixed point are able to restore the large scale features of the image after only a few iterations, but the fixed point method takes much longer to completely restore the finer scale features, as it can be seen from the right plot in Fig. 7 and the horizontal cross sections plots in Fig. 8.
- The cost per iteration, as well as the memory requirement, of the primal-dual method is between 30 and 50 per cent more than for the fixed point iteration. The cost of an iteration of the time marching method is roughly the same as that of an inner CG iteration for the primal-dual method, but both the primal-dual method and the fixed point method require far less (CG) iterations than the time marching method.

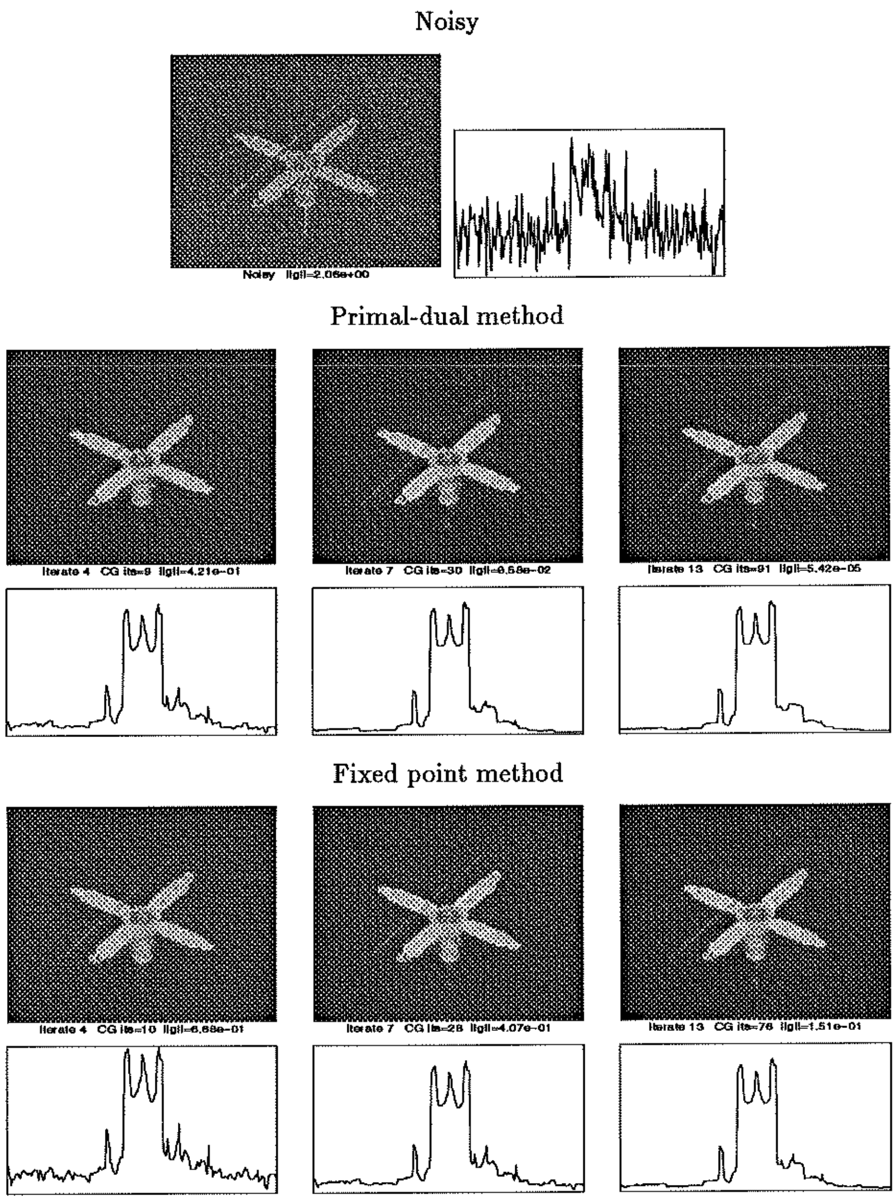


Figure 8: The size of the original image is 256×256 and the signal to noise ratio is approximately 1. The plots are horizontal cross sections of the corresponding images at pixel 130.

5 Cosine Transform Preconditioners For Total Variation Image Denoising and Deblurring

In this section, we will treat the issues regarding preconditioning the conjugate gradient algorithm used for the solution of the linear systems arising in Total Variation image restoration.

5.1 Preconditioners for Total Variation image restoration

Let us denote by $A = \alpha L + K^*K$, where L corresponds to the differential operator of either the fixed point iteration or Newton's method.

When $K = I$, the matrix A will correspond to an elliptic differential operator and there are many possible preconditioners for it, e.g. SSOR, ILU, MG, DD [22, 32]. We have satisfactorily used ILU preconditioning in our numerical experiments. In [35, 36] a multigrid preconditioner is used.

When $K \neq I$, there exist many preconditioners for each of the summands of A separately, e.g. the elliptic preconditioners mentioned above and circulant preconditioners [33] for the convolution operator K^*K. However, a preconditioner for the sum of both terms is not easy to find.

In Vogel and Oman [36], the following operator splitting/ADI type preconditioner is proposed:

$$M = (K^*K + \gamma_1 I)^{1/2}(\alpha L_\beta(u^n) + \gamma_2 I)(K^*K + \gamma_1 I)^{1/2}.$$

This approach works well when α is very large or very small, but the intermediate cases are often the ones which are of interest.

In the rest of this section, we shall briefly describe a preconditioning technique that we have developed in [11] which can be applied to both terms.

5.2 Optimal Discrete Cosine Transform Preconditioners

Recall that the matrix of the one-dimensional n-discrete cosine transform is

$$C_n = \sqrt{\frac{2-\delta_{i1}}{n}} \cos\left(\frac{(i-1)(2j-1)\pi}{2n}\right), \quad 1 \leq i, j \leq n,$$

which is an orthogonal matrix. Furthermore, if v is an n-vector, then the product $C_n v$ can be performed in $O(n \log n)$ operations.

We denote the set of matrices which can be diagonalized by the discrete cosine transform by

$$\mathcal{B}_{n\times n} = \{C_n^t \Lambda_n C_n \mid \Lambda_n \text{ is a real diagonal matrix}\}.$$

Given A_n, we define the optimal cosine transform preconditioner for A_n, $c(A_n)$, as the minimizer B_n of the Frobenius norm of the difference $B_n - A_n$, $||B_n - A_n||_F$, over all matrices $B_n \in \mathcal{B}_{n\times n}$. The matrix $c(A_n)$ is always symmetric and it can be shown that $c(A_n)$ is positive definite whenever A_n is so. More details can be found in [10].

The construction cost of $c(A_n)$, for different types of matrix A_n, is shown in Table 2.

A_n	cost of constructing $c(A_n)$
general	$O(n^2)$
Toeplitz (from K)	$O(n \log n)$
Diagonally sparse (from L)	$O(n \log n)$

Table 2: Construction cost for the optimal cosine transform preconditioner.

The matrix of the two-dimensional $n \times n$-discrete cosine transform is $C_n \otimes C_n$ and the optimal two-dimensional cosine transform preconditioner for a matrix $A_{n\times n}$ is similarly defined as the matrix $c_2(A_{n\times n}) = B_{n\times n}$ that can be diagonalized by the two-dimensional cosine transform and which minimizes the Frobenius norm $||B_{n\times n} - A_{n\times n}||_F$, over all those matrices that can be diagonalized by the two-dimensional cosine transform. The construction cost of $c_2(A_{n\times n})$, for different types of matrix $A_{n\times n}$, is shown in Table 3.

A_n	cost of constructing $c(A_n)$
general	$O(n^4)$
2d Toeplitz (from K)	$O(n^2 \log n)$
2d diagonally sparse (from L)	$O(n^2 \log n)$

Table 3: Construction cost for the two-dimensional optimal cosine transform preconditioner.

5.3 Cosine Transform for TV denoising and deblurring

Since the matrix A is a general dense matrix, from Table 3 we deduce that the cost of constructing the preconditioner $c_2(A)$ would be n^4, which can easily be

Preconditioner	#. CG iterations
Cosine + diag. scal.	204
Cosine only	348
diag. scal. only	462
no preconditioning	922

Table 4: Comparison of the work for the TV image restoration with different preconditioners

prohibitively large. Therefore some simplifications have to be done in order to reduce this cost. Since

$$c_2(A) = c_2(K^*K + \alpha L) = c_2(K^*K) + c_2(\alpha L) \approx c_2(K)^* c_2(K) + \alpha c_2(L),$$

we propose

$$M = c_2(K)^* c_2(K) + \alpha c_2(L)$$

as a preconditioner for A.

If $c_2(K) = C^t \Lambda_1 C$ and $c_2(L) = C^t \Lambda_2 C$, then

$$M = (C_n \otimes C_n)^t \ (\Lambda_1^* \Lambda_1 + \alpha \Lambda_2) \ (C_n \otimes C_n),$$

which is easily invertible in $O(n^2 \log n)$ operations using the FFT. We can also incorporate the technique of diagonal scaling for taking into account large variations of the coefficients of the differential operator. Note that the cost per conjugate gradient step is $O(n^2 \log n)$.

5.4 Numerical Results for Cosine Transform Preconditioners

In this section we compare the work used to perform 5 fixed point iterations using the conjugate gradient method with different preconditioners. This work is measured in inner CG iterations.

The test and original images are displayed in Fig 9 and the parameters that have been used are: $\beta = 0.01$, $\alpha = 10^{-6}$ and conjugate gradient relative tolerance $= 10^{-2}$. The results are shown in Table 4. It can be seen that the diagonally scaled cosine transform preconditioner can reduce the number of iterations by a factor of more than 4 compared with using no preconditioning at all. Other tests in [11] show that the performance is also insensitive to the value of α.

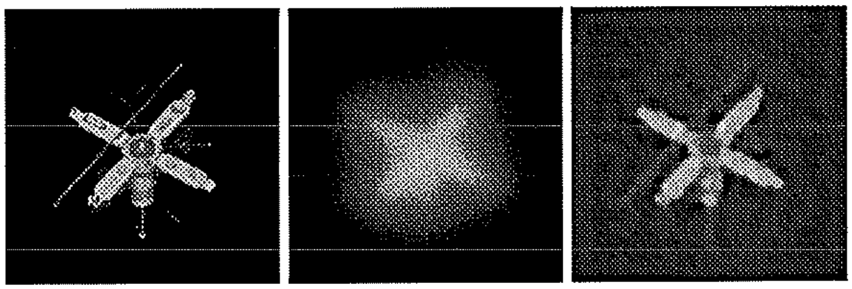


Figure 9: Original image, blurred and deblurred satellite image.

6 Conclusion

These nonlinear PDE based image restoration methods are very promising. Their primary advantage is a systematic way of preserving sharp edges. There is a nice mathematical framework and the techniques can be extended to more general image processing tasks such as segmentation and object recognition. There is also the possibility of combining them with the FFT/wavelet based methods as well. While these methods are still relatively expensive compared to traditional transform and frequency based methods, well established PDE and CFD numerical schemes can be used to make them more efficient. In particular, iterative methods can play a significant role and we have attempted to show some of the possibilities in this chapter.

References

[1] L. ALVAREZ AND J. MOREL, *Formalization and computational aspects of image analysis*, Acta Numerica, (1994), pp. 1–59.

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

[3] E. BOMAN AND I. KOLTRACHT, *Fast transform based preconditioners for Toeplitz equations*, SIAM J. Matrix Anal., 16 (1995), pp. 628–645.

[4] J. BUNCH, *Stability of methods for solving Toeplitz systems of equations*, SIAM J. Sci. Stat. Comp., 6 (1985), pp. 349–364.

[5] R. CHAN, *Iterative Methods for Overflow Queueing Models I* , Stud. Appl. Math., 74 (1987), pp. 171–176.

[6] R. Chan and T. Chan, *Circulant Preconditioners for Elliptic Problems*, J. Numer. Linear Algebra Appls., 1 (1992), pp. 77–101.

[7] R. Chan, J. Nagy, and R. Plemmons, *Block circulant preconditioners for 2-dimensional deconvolution problems*, in Proceedings to the Conference on Advanced Signal Processing Algorithms, Architectures and Implementation III, Vol. 1770, San Diego, CA, July, 1992, F. T. Luk, ed., SPIE, 1992.

[8] ——, *Circulant preconditioned toeplitz least squares iterations*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 80–97.

[9] R. Chan and G. Strang, *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 104–119.

[10] R. Chan and C. Wong, *Sine Transform Based Preconditioners for Elliptic Problems*. submitted.

[11] R. H. Chan, T. F. Chan, and C.-K. Wong, *Cosine transform based precontioners for total variation minimization problems in image processing*, in Proc. 2nd IMACS Symposium on Iterative Methods, Blagoevgrad, Bulgaria, June 1995, S. Margenov and P. Vassilevski, eds., IMACS, 1995, pp. 311–329.

[12] R. H. Chan, T. F. Chan, and H. M. Zhou, *Continuation method for total variation denoising problems*, in Advanced Signal processing algorithms, F. T. Luk, ed., vol. 2563, SPIE - The Int'l Society for Optical Eng., 1995, pp. 314–325.

[13] R. H. Chan and M. Ng, *Conjugate gradient methods for toeplitz systems*, Tech. Rep. 94-8(35), Dept. of Math., The Chinese University of Hong Kong, 1994. To appear in SIAM Review.

[14] T. Chan, *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 766–771.

[15] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for Total Variation-based image restoration*, in ICAOS'96, 12th Int'l Conf. on Analysis and Optimization of systems: Images, wavelets and PDE's, Paris, June 26-28, 1996, M. Berger, R. Deriche, I. Herlin, J. Jaffre, and J. Morel, eds., no. 219 in Lecture Notes in Control and Information Sciences, 1996, pp. 241–252.

[16] A. R. Conn and M. L. Overton, *A primal-dual interior point method for minimizing a sum of euclidean norms*. preprint, 1995.

[17] G. FIORENTINO AND S. SERRA, *Tau preconditioners for high order elliptic problems*, in Proc. 2nd IMACS Symposium on Iterative Methods, Blagoevgrad, Bulgaria, June 1995, IMACS, 1995, pp. 342–353.

[18] E. GIUSTI, *Minimal Surfaces and Functions of Bounded Variations*, Birkhäuser, 1984.

[19] I. GOHBERG, T. KAILATH, AND V. OLSHEVSKY, *Fast Gaussian elimination with partial pivoting for matrices with displacement structure*, Math. Comp., (1995). To appear.

[20] G. GOLUB AND C. V. LOAN, *Matrix Computations, 2nd Edition*, The Johns Hopkins Univ. Press, 1989.

[21] R. GONZALEZ AND R. WOODS, *Digital image processing*, Addison Wesley, 1993.

[22] W. HACKBUSCH, *Iterative solution of large linear systems of equations*, Springer Verlag, NY, 1994.

[23] M. HANKE, *Conjugate gradient type methods for ill-posed problems*, Notes in Math., No. 327, Longman, Harlow, 1995.

[24] M. HANKE AND P. C. HANSEN, *Regularization methods for large scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315.

[25] G. HEINIG, *Inversion of generalized cauchy matrices and other classes of structured matrices*, in Linear algebra in signal processing, vol. 69 of IMA volumes in mathmatics and its applications, 1994, pp. 95–114.

[26] T. HUCKLE, *Circulant and Skew-circulant Matrices for Solving Toeplitz Matrix Problems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 767–777.

[27] C. T. KELLEY, *Iterative Methods for Linear and Nonlinear Equations*, vol. 16 of Frontiers in Applied Mathematics, SIAM, 1995.

[28] J. M. MOREL AND S. SOLIMINI, *Variational methods for image segmentation*, Birkhauser, Boston, 1995.

[29] P. PERONA AND J. MALIK, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intelligence, 12 (1990), pp. 629–639.

[30] L. RUDIN AND S. OSHER, *Total variation based image restoration with free local constraints*, in Proc. Int. Conf. on Image Processing, Austin, TX, 1994, 1994, pp. 31–35.

[31] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[32] Y. SAAD, *Iterative methods for sparse linear systems*, PWS Publ. Co., Boston, 1996.

[33] G. STRANG, *A Proposal for Toeplitz Matrix Calculations*, Numer. Math., 51 (1987), pp. 143–180.

[34] B. M. TER HAAR ROMENY, *Geometry-Driven Diffusion in Computer Vision*, Kluwer Academic Publishers, 1994.

[35] C. R. VOGEL, *A multigrid method for total variation-based image denoising*, in Computation and Control IV, K. Bowers and J. Lund, eds., vol. 20 of Progess in Systems and Control Theory, Birkhauser, 1995.

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

[37] H. VOSS AND U. ECKHARDT, *Linear convergence of generalized Weiszfeld's method*, Computing, 25 (1980).