Iterative Parameter-choice and Algebraic Multigrid for Anisotropic Diffusion Denoising

Donghui Chen, Misha Kilmer, Scott MacLachlan

January 18, 2010

Abstract

Anisotropic diffusion methods have shown good qualitative results for image denoising. This paper gives a review of the anisotropic diffusion methodology and its application to image restoration. We investigate an AMG algorithm to solve a regularized anisotropic diffusion equation, which is not only well-posed, but also has a nontrivial steady-state solution. A new regularization parameter-choice method, the Brent-NCP algorithm is introduced, combining Brent's method and the normalized cumulative periodogram (NCP) information about the residual. We test our algorithm on three standard test images. The experimental results demonstrate the effectiveness of the AMG approach and the broad applications of the Brent-NCP parameter-choice algorithm.

1 Introduction

It is well known that during the formation, transmission, and recording processes, images deteriorate with various types of noise. Therefore, it is important to eliminate the noise automatically and efficiently. Many image denoising techniques have been proposed over the years, and a good review of them can be found in [5]. In particular, with developments in computer technology, methods based on partial differential equations (PDEs) have been extensively studied as approaches to image denoising.

Anisotropic diffusion, first introduced by Perona and Malik [17], has been widely accepted as a method for reducing noise while preserving and enhancing edges [2]. Many papers have proposed different techniques for solving the diffusion equations. However, most of these techniques are based on simple relaxation methods. As with any diffusion process, they are slow to resolve low-frequency artifacts, and the rate of convergence decreases sharply with increasing image size.

On the other hand, multigrid approaches eliminate both the high and low frequency errors rapidly [3, 4, 23]. The optimality of multigrid methods suggests that they are potentially good solvers for anisotropic diffusion problems. Previous research has shown this; geometric multigrid (GMG) methods are used to solve the anisotropic diffusion equations at each step in a fixed point iteration in [1, 22, 24]. Unfortunately, these results show that the GMG method converges slowly and is not robust with respect to discontinuous diffusivities, which is the case in image denoising applications.

One way to improve on the performance of GMG is to use algebraic multigrid (AMG) methods. AMG, first proposed in [3], is designed to utilize classical multigrid principles to obtain a fast solution algorithm for a wide range of problems. It has many improvements over GMG in its robustness and ease of use, especially when the coefficients of the PDE are discontinuous and vary widely [19, 23]. The concept of using AMG to solve anisotropic diffusion equations is not new. In [10], Sapiro et al. use AMG to denoise and segment hyperspectral images. In [9], Chen and Savage try to accelerate a standard AMG approach by repeatedly using the setup phase information within the AMG algorithm.

Besides finding a fast algorithm used to solve the linearized anisotropic diffusion equation, choosing an optimal regularization parameter is also critical in anisotropic diffusion denoising. Traditionally, these parameters are chosen based on either experience or the norm of the residual vector, with techniques such as Generalized Cross-Validation (GCV), L-curve fitting, etc. [14]. Generally, these methods require solving the problem many times in order to find the optimal regularization parameter, which is inefficient for a large problem. In this paper, we propose a new inexpensive parameter-choice method based on Brent's method and the normalized cumulative periodogram (NCP) information about the residual [13, 20]. Together with AMG using standard (Ruge-Stüben) coarsening [19], the proposed algorithm can find the optimal regularization parameter for the anisotropic diffusion equation quickly and efficiently. Our approach differs from previous work in the following two aspects:

1. We consider a different objective diffusion equation, which is not only well-posed, but also has a nontrivial steady-state solution. In practice, by searching for the steady-state solution directly, the proposed AMG solver does not require as many AMG iterations as in previous work [9].

2. By introducing the new Brent-NCP parameter-choice method, the proposed AMG denoising algorithm chooses the regularization parameters automatically. This makes our algorithm more adaptive and efficient. Furthermore, the application of the parameter-choice algorithm is not limited to anisotropic-diffusion based denoising. It can be applied to a very broad range of parameter-dependent denoising algorithms.

The remainder of this paper is organized as follows. Section 2 presents a review of the anisotropic diffusion equation in image restoration and introduces our cost functional. A semi-implicit discretization technique and fixed-point AMG algorithm are also discussed here. Section 3 introduces the new Brent-NCP regularization parameter-choice algorithm. Section 4 presents the experimental results. The conclusions of this paper are presented in Section 5.

2 Anisotropic Diffusion Review

Given a typical noisy image (where the noise is known to be white noise), the noise dominates the high-frequency components of the image [12]. Intuitively, we can smooth out the noise by convolving the image with a Gaussian kernel. This is equivalent to solving the heat equation. By forming the physical phenomena in terms of mathematical language, PDEs, we can understand the denoising problem more clearly and deeply. However, linear Gaussian smoothing not only removes high-frequency noise but also blurs edges and destroys finer textures. Anisotropic diffusion denoising is based on the idea of applying a smoothing process that depends on local properties of the image [2, 26].

2.1 Perona-Malik Diffusion Model

In [17], Perona and Malik introduced the anisotropic diffusion approach to replace classical isotropic diffusion. This approach can avoid the excessive smoothing effect that occurs with isotropic diffusion procedures, such as Gaussian filters. On the continuous domain, the anisotropic diffusion equation for an image, I, is given by

$$\begin{cases} I_t = \operatorname{div}(c(|\nabla I|^2)\nabla I) & \text{in } \Omega \times (0,T) \\ \frac{\partial I}{\partial N} = 0 & \text{on } \partial\Omega \times (0,T) \\ I(0,x) = I^0(x) & \text{in } \Omega \end{cases} \tag{1}$$

where ∇ is the gradient operator, div is the divergence operator, and

$$c(s^2) : [0,\infty] \to [0,\infty]$$

describes the diffusivity. The differential equation has initial condition $I^0(x)$, which is the noisy image.

Defining the flux function

$$\Phi(s) := c(s^2)s,$$

it is shown in [2] that the blurring-enhancing process depends on the sign of derivative of flux function, $b(s^2) := \Phi'(s) = c(s^2) + 2s^2c'(s^2)$. If $b(s^2) > 0$, edges are blurred when the Perona-Malik (PM) model is a forward parabolic equation, while if $b(s^2) < 0$, edges are sharpened when the PM model is a backward parabolic equation. Given the threshold K, PM model shows the desirable result of blurring small discontinuities in image I where $|\nabla I| < K$, and sharpening edges in image I where $|\nabla I| > K$. This gives rise to the following assumptions on $c(s^2)$.

$$\begin{cases} c(s^2) : [0,\infty] \to [0,\infty] \ \text{ decreasing}, \\ c(0) = 1, \\ b(s^2) = c(s^2) + 2s^2c'(s^2) < 0 \ \text{ for } \ s \geq K, \end{cases} \tag{2}$$

In their original paper, Perona and Malik choose the diffusivity to be $c(s^2) = \frac{1}{1+s^2/K^2}$, where K is a threshold determined by the noise level [6].

It is interesting to note that (1) can be regarded as a gradient descent method for the energy functional

$$E(I) := \frac{1}{2}\int_\Omega \Psi(\nabla I)dx,$$

where $\Psi(\nabla I)$ is a potential function whose gradient is the flux function $\Phi(\nabla I)$. The gradient descent method for the above energy functional results in the PM model (1).

While numerical results with the PM model (1) are quite impressive, the forward-backward diffusion process itself is not well-posed. This is the so-called Perona-Malik paradox. In [15], Kichenassamy proves that if the initial image, $I^0(x)$, is not infinitely differentiable, there is no weak solution of (1). Consequently, the notion of a "generalized solution", which is piecewise linear and contains jumps, is introduced. However, one should neither expect uniqueness nor stability with respect to the initial image. Examples of significantly differing solutions with nearly identical initial data have been reported [15].

2.2 Regularization

Although the ill-posedness of the PM model can be handled by applying an implicit spatial discretization [28], in order to make the numerical implementation more predictable, it is more natural to introduce regularization into the continuous PM equation. Catt et al. [7] introduce a spatial regularization that makes the forward-backward diffusion process become well posed. The idea is to use smoothed version, $G_\sigma * \nabla I$, of the image gradient ∇I in the diffusivity $c(|\nabla I|^2)$. Here, G_σ can be any "low-pass filter". In this paper, we assume that G_σ is a 5×5 Gaussian kernel with the standard deviation $\sigma = 1.5$. Since $G_\sigma * \nabla I = \nabla G_\sigma * I$, the spatial regularized PM model becomes

$$I_t = \operatorname{div}(c(|(\nabla G_\sigma * I)|^2)\nabla I) \tag{3}$$

Catt et al. [7] prove that there exists a unique solution for the regularized PM equation (3) with corresponding initial and boundary values. Furthermore, this spatial regularization makes the filter insensitive to noise at scales smaller than σ. This avoids the shortcoming of the original PM model, which cannot distinguish between "true" edges and "false" edges created by the noise.

While the regularized (3) becomes well-posed, it still leads to a process where the solution converges to a constant steady-state solution [27] (just as in the PM model (1)). In order to get nontrivial results, it is required to specify a stopping time, T_0. Sometimes, it is attempted to circumvent this task by adding an additional reaction term, which keeps the steady-state solution close to the original image [16]

$$I_t = \operatorname{div}(c(|(\nabla I)|^2)\nabla I) + \lambda(I^0 - I). \tag{4}$$

The reaction term, $(I^0 - I)$, has the effect of locally moderating the diffusion, as the diffused image, I, diffuses further away from the original image, I^0. In practice, such a modification shifts the problem of specifying a stopping time, T_0, to the problem of determining regularization parameter λ.

Combining the spatial regularization (3) and reaction anisotropic diffusion (4), we get the anisotropic diffusion equation, $I_t = \operatorname{div}(c(|(\nabla G_\sigma * I)|^2)\nabla I) + \lambda(I^0 - I)$. As discussed above, this PDE is not only well-posed, but also has a non-trivial steady-state solution satisfying

$$0 = \operatorname{div}(c(|(\nabla G_\sigma * I)|^2)\nabla I) + \lambda(I^0 - I). \tag{5}$$

In this paper, we solve the above nonlinear equation (5) to get the denoised image. Given a fixed G_σ, finding the optimal regularization parameter λ_{opt} is critical to solving (5), which is discussed in Section 3.

2.3 Discretization

Assuming regularization parameter λ is known, because semi-implicit discretization has numerical stability property [21], we use this scheme to find the solution of (5),

$$\lambda(I^{n+1} - I^0) = \operatorname{div}(c(|(\nabla G_\sigma * I^n)|^2)\nabla I^{n+1}), \tag{6}$$

where the superscript, n, denotes a numerical approximation taken at the nth iteration. Writing the divergence term as

$$\operatorname{div}(c\nabla I) = \frac{\partial}{\partial x}\left(c\frac{\partial I}{\partial x}\right) + \frac{\partial}{\partial y}\left(c\frac{\partial I}{\partial y}\right), \tag{7}$$

we use forward and central differences to approximate the derivatives of the image, I. In digital images, we take the distance between adjacent grid points, h, to be 1. Therefore, for simplicity, we omit the distance h in the following discretization formulas. The value of the divergence operator at grid point (i, j) can then be written as

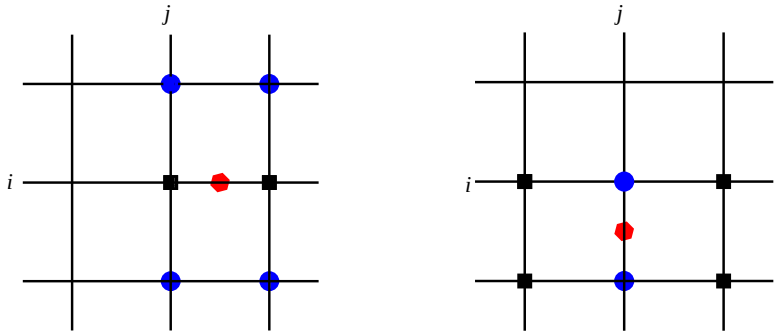
$$\begin{aligned}\operatorname{div}(c\nabla I)\,|_{i,j} &= c_{i+\frac{1}{2},j}(I_{i+1,j} - I_{i,j}) - c_{i-\frac{1}{2},j}(I_{i,j} - I_{i-1,j}) + c_{i,j+\frac{1}{2}}(I_{i,j+1} - I_{i,j}) - c_{i,j-\frac{1}{2}}(I_{i,j} - I_{i,j-1}) \\ &= c_{i+\frac{1}{2},j}I_{i+1,j} + c_{i-\frac{1}{2},j}I_{i-1,j} + c_{i,j+\frac{1}{2}}I_{i,j+1} + c_{i,j-\frac{1}{2}}I_{i,j-1} - (c_{i+\frac{1}{2},j} + c_{i-\frac{1}{2},j} + c_{i,j+\frac{1}{2}} + c_{i,j-\frac{1}{2}})I_{i,j}\end{aligned}$$

Notice that interpolation is needed to evaluate the diffusivity, c, at locations $(i \pm \frac{1}{2}, j)$ and $(i, j \pm \frac{1}{2})$. This can be done as follows, see also Figure 1. Denoting $\hat{I} = \nabla G_\sigma * I^n$, we use central differences and linear interpolation to compute

$$c_{i,j+\frac{1}{2}} := c\left((\hat{I}_{i,j+1} - \hat{I}_{i,j})^2 + \left(\frac{\hat{I}_{i+1,j+1} - \hat{I}_{i-1,j+1} + \hat{I}_{i+1,j} - \hat{I}_{i-1,j}}{4}\right)^2\right),$$

and

$$c_{i+\frac{1}{2},j} := c\left(\left(\frac{\hat{I}_{i+1,j+1} - \hat{I}_{i+1,j-1} + \hat{I}_{i,j+1} - \hat{I}_{i,j-1}}{4}\right)^2 + (\hat{I}_{i+1,j} - \hat{I}_{i,j})^2\right).$$



(a) $c_{i,j+\frac{1}{2}}$ (b) $c_{i+\frac{1}{2},j}$

Figure 1: Vertices involved in the approximation of the diffusivities $c_{i,j+\frac{1}{2}}$ and $c_{i+\frac{1}{2},j}$ (marked by hexagons). The grid points represented by dots are used to compute derivatives in the y-direction, while the grid points represented by squares are used to compute derivatives in the x-direction.

The advantage of the above discretization is that it is good for rotation-invariance properties, and, practically, the result is also less sensitive to noise [2].

The discretization of (6) is then

$$a_{i,j}I^{n+1}_{i,j} - \left(c_{i+\frac{1}{2},j}I^{n+1}_{i+1,j} + c_{i-\frac{1}{2},j}I^{n+1}_{i-1,j} + c_{i,j+\frac{1}{2}}I^{n+1}_{i,j+1} + c_{i,j-\frac{1}{2}}I^{n+1}_{i,j-1}\right) = \lambda I^0_{i,j},$$

where $a_{i,j} = \left(\lambda + \left(c_{i+\frac{1}{2},j} + c_{i-\frac{1}{2},j} + c_{i,j+\frac{1}{2}} + c_{i,j-\frac{1}{2}}\right)\right)$. In matrix-vector notation, the above discrete form can be written as

$$A(I^n, \lambda, \sigma)\text{vec}(I^{n+1}) = \lambda \text{vec}(I^0), \tag{8}$$

where $\text{vec}(I)$ denotes the vector obtained by stacking all of the columns of image I into a vector. The matrix $A(I^n, \lambda, \sigma)$ is a symmetric, positive-definite, and banded sparse matrix with positive diagonal entries and negative off-diagonal entries. The sum of each row is λ.

For fixed λ, σ, (8) represents a fixed-point linearization of the nonlinear PDE described by (5). This leads to the fixed-point AMG algorithm given as Algorithm 1. For each iteration, I^{n+1} in (8) is computed by standard Ruge-Stüben AMG approach, see detail in [4, 19, 23]. In the third step of Algorithm 1, instead of recomputing the AMG setup phase information [19], we recycle the old AMG setup data but change the fine-scale matrix $A_n = A(I^n, \lambda, \sigma)$. This appears to lead to speedup of Algorithm 1, however further study of the optimal setup strategy is still needed.

Algorithm 1 Fixed-Point AMG Algorithm for fixed λ and σ

1: Initialize the matrix $A_0 = A(I^0, \lambda, \sigma)$ in (8) with initial guess I^0, compute the the AMG setup data [19]
2: **while** norm$(I^{n+1} - I^n) >$ tol **do**
3: Compute the solution I^{n+1} of $A_n\text{vec}(I^{n+1}) = \lambda\text{vec}(I^0)$ by AMG method using the setup data computed in Step 1.
4: Initialize the matrix $A_{n+1} = A(I^{n+1}, \lambda, \sigma)$ in (8) with initial guess I^{n+1}
5: **end while**
6: **return** I^i

3 The Brent-NCP parameter choice method

The remaining outstanding issue in solving (5) is the choice of the optimal regularization parameter, λ. Most algorithms for choosing the regularization parameter are based on the norm of the residual vector, such as Generalized Cross-Validation (GCV), L-curve, etc. [14]. In [13], Hansen, Kilmer and Kjeldsen propose a different rule that seeks to use all of the information available in the residual vector. Assuming the noise is white noise, the key idea of this method is to choose the regularization parameter for which the residual vector changes from being dominated by the remaining signal to being white-noise like. By employing statistical tools and fast Fourier transforms, this method leads to a parameter-choice rule based on the normalized cumulative periodogram (NCP), which is particularly well-suited for large-scale problems. For more on using the cumulative periodograms for regularization parameter selection, see [20] and the references therein.

3.1 Normalized cumulative periodogram

Given a $n \times n$ 2D signal X, denote $q = \lfloor n/2 \rfloor + 1$, and let the $q \times q$ matrix $P = |\mathcal{F}(X)|^2$ be the power spectrum of X, where $\mathcal{F}(X)$ denotes the 2D Fourier transform of X. After reordering these elements in the order of increasing spatial frequency, $\hat{p} = \text{perm}(\text{vec}(P))$, the NCP for X is defined as a vector $\text{ncp}(X)$ of length $q^2 - 1$ with elements

$$\text{ncp}(X)_k = \frac{||\hat{p}(2 : k+1)||_1}{||\hat{p}(2 : q^2)||_1}, \quad k = 1, \cdots, q^2 - 1. \tag{9}$$

If the residual is white noise, the expected values of the elements in the NCP lie on a straight line between $(0, 0)$ and $(q^2, 1)$. Therefore, the theoretical NCP is denoted by the vector v s.t. $v_k = \frac{k}{q^2-1}$. Given the computed residual, r, if $\text{ncp}(r)$ lies within the Kolmogorov-Smirnov limits of the line v, the conclusion is the acceptance of the hypothesis that the residual is white noise [11].

In practice, there is often more than one regularization parameter satisfying the Kolmogorov-Smirnov test. In order to find a unique optimal regularization parameter, Hansen et al. suggest to compute the largest difference in absolute value between the estimated and the theoretical NCP, and choose the regularization parameter that minimizes this difference:

$$\underset{\lambda}{\text{argmin}}\, \mathcal{N}(\lambda) := ||v - \text{ncp}(r_\lambda)||_1, \tag{10}$$

where r_λ is denoted as the computed residual given regularization parameter λ.

3.2 Brent-NCP algorithm

Efficiently solving the minimization problem (10) is essential to the parameter-choice algorithm. In general, solving minimization problem (10) requires computing multiple solutions of (5) for different values of λ, which is time-consuming part. The proposed Brent-NCP method is designed to minimize the number of solves, which makes it more efficient.

Brent's method combines some parts of a golden section search method and some of a parabolic interpolation method [18]. This method is characterized by quadratic convergence in case of smooth functions and guaranteed linear convergence in case of nonsmooth or oscillatory functions. For convenience, we summarize Brent's method applied to minimize $\mathcal{N}(\lambda)$ in (10) below.

Algorithm 2 Brent-NCP Algorithm

1: Set $ax < cx$
2: $bx \leftarrow ax + \frac{3-\sqrt{5}}{2} \times (cx - ax)$
3: $\lambda_{opt} \leftarrow (ax + cx)/2$
4: Compute the residual $r_{\lambda_{opt}} := I(\lambda_{opt}) - I^0$, where $I(\lambda_{opt})$ is the solution image computed by fixed-point AMG Algorithm 1 applied to (5) given regularization parameter λ_{opt}
5: Compute $\mathcal{N}(\lambda_{opt}) = ||v - \text{ncp}(r_{\lambda_{opt}})||_1$
6: **while** $|\lambda_{opt} - bx| > \text{tol}$ **do**
7: Construct a trial parabolic fit [18]
8: **if** parabolic fit is acceptable **then**
9: Take the parabolic step
10: **else**
11: Take a golden section step
12: **end if**
13: Update the values $ax, bx, cx, \lambda_{opt}$ [18], compute $\mathcal{N}(\lambda_{opt}) = ||v - \text{ncp}(r_{\lambda_{opt}})||_1$, where $r_{\lambda_{opt}} := I(\lambda_{opt}) - I^0$ is the residual, $I(\lambda_{opt})$ is computed by Algorithm 1 applied to (5) given the updated regularization paramter λ_{opt}
14: **end while**
15: **return** λ_{opt}

In practice, the above Brent-NCP algorithm converges very quickly, see Table 2 in Section 4. In addition, it has very broad application areas. By switching the method for computing $I(\lambda_{opt})$ from the fixed-point AMG Algorithm 1 to other algorithms in the fourth and thirteenth step of Algorithm 2, we can find the optimal regularization parameters for those algorithms, see Section 4 for detail.

4 Experimental results

This section is devoted to presenting the results obtained with the proposed algorithm. A comparison with the conventional linear Gaussian filter, Wiener filter [12], and total variation (TV) denoising [8] is also presented.

Since the exact, noise-free images were known, the distance between restored image I and the ground truth image I^0 was computed using the Mean Structure Similarity (MSSIM) [25], which is designed to improve on traditional methods like peak signal-to-noise ratio (PSNR) and mean squared error (MSE). Given any two images $\mathbf{x}$ and $\mathbf{y}$, the Structure Similarity (SSIM) is defined as

$$SSIM(\mathbf{x},\mathbf{y}) = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + c_1)(2\text{cov}(\mathbf{x},\mathbf{y}) + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + c_2)} \tag{11}$$

where $\mu_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$ are the means of images $\mathbf{x}$ and $\mathbf{y}$ respectively, $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{y}}$ are the variances of images $\mathbf{x}$ and $\mathbf{y}$, $\text{cov}(\mathbf{x},\mathbf{y})$ is the covariance of the two images, and c_1 and c_2 are two variables to stabilize the division with small denominators, the defaults are $c_1 = 0.0001, c_2 = 0.0009$.

In practice, SSIM is calculated on local windows rather than over the whole image. As in [25], we use a normalized 11×11 circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_i \mid i = 1, 2, \cdots, N = 121\}$, with standard deviation of 1.5 in our experiments. As a result, $\mu_{\mathbf{x}}$, $\sigma_{\mathbf{x}}$ and $\text{cov}(\mathbf{x},\mathbf{y})$ in the SSIM measure (11) are modified as $\mu_{\mathbf{x}} = \sum_{i=1}^{N} w_i x_i, \sigma_{\mathbf{x}} = \left(\sum_{i=1}^{N} w_i (x_i - \mu_{\mathbf{x}})^2\right), \text{cov}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{N} w_i (x_i - \mu_{\mathbf{x}})(y_i - \mu_{\mathbf{y}})$. In order to get a single overall similarity measure of the two images, the MSSIM is computed by choosing the local window pixel-by-pixel

$$MSSIM(I, I^0) = \frac{1}{mn} \sum_{i=1}^{mn} SSIM(\mathbf{x}_i, \mathbf{y}_i), \tag{12}$$

where $\mathbf{x}_i$ and $\mathbf{y}_i$ are the ith local window, and mn is size of the two images, I and I^0.

Three common test images and a noisy variant of each of them are shown in the first and second row of Figure 2. White Gaussian noise has been added. Its variance was chosen to be 50%, 50% and 25% times the images value range, respectively. We applied the proposed fixed-point AMG Algorithm 1, as well as TV, Wiener filters and linear Gaussian filters, to the three noisy test images.

First, we test the applicability of Brent-NCP Algorithm 2. Table 1 shows the computed optimal regularization parameters, λ, using Algorithm 2. In the fourth and thirteenth step of Algorithm 2, we use four different methods to compute the denoised image $I(\lambda_{opt})$. The first two methods use Algorithm 1 to solve the anisotropic diffusion equation (5), the difference is a) using only a single V-cycle to compute the solution I^{n+1} in Step 3 of Algorithm 1, while b) solving the system in Step 3 of Algorithm 1 to a residual-reduction tolerance of 0.01. Besides anisotropic diffusion denosing, we also use c) TV and d) linear Guassian filters [14] to get the denoised image $I(\lambda_{opt})$. For TV denoising, the cost functional is defined the as

$$\min_I TV(I) + \frac{\lambda}{2}\|I - I^0\|_2^2, \tag{13}$$

where $TV(I) = \int_\Omega |\nabla I|^2$, and λ is the regularization parameter. We use the algorithm proposed by Chambolle in [8] to solve the above TV minimization problem. For linear Gaussian filters, the regularization parameters are the convolution kernel, G_σ. In the experiments, we set the size of kernel to be 5×5 by trial and error. The variance σ is computed using the Brent-NCP Algorithm 2.

Denoting the above 4 different methods as $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$, respectively, the first two columns in Table 1 show that the optimal regularization parameters computed using method $\mathbf{A}$ are very close to those using method $\mathbf{B}$. The computational work for these two methods are shown in Table 2. We can see that the total number of V-cycles needed in method $\mathbf{A}$ is much less than that of needed in method $\mathbf{B}$. We can further speedup the NCP-Brent Algorithm 2 by using a single V-cycle in Step 3 of Algorithm 1, however this may lead to poorer-quality solutions, further investigation is still needed.

Table 1: Computed regularization parameters λ using Brent-NCP Algorithm 2.

	AMG Algorithm 1		TV	Gaussian filters
images	Method $\mathbf{A}$	Method $\mathbf{B}$	λ	σ
Head phantom	5.9052e-02	6.4438e-02	6.2442	1.2461
Satellite	1.4755e-01	1.2705e-01	8.3027	2.0163
Cameraman	1.3868e-01	1.0966e-01	7.6393	1.1559

Table 2: Comparison of the computational work in Brent-NCP Algorithm 2.

	# of Brent steps		# of linearizations		# of V-cycles	
images	Method **A**	Method **B**	Method **A**	Method **B**	Method **A**	Method **B**
Head phantom	14	14	15	15	77	3263
Satellite	15	15	16	16	96	2112
Cameraman	12	14	13	15	48	1739

Given the computed optimal regularization λ_{opt} in Table 1, we compute the denoised image using four different methods: a) fixed-point AMG Algorithm 1 with anisotropic diffusion equation 5 (in Step 3 of Algorithm 1, we use a single V-cycle), b) TV, c) Wiener filters, and d) linear Guassian filters. The MSSIM of the final denoised results are shown in Table 3. The result images are shown in the last two rows in Figure 2 and Figure 3. The results using AMG with anisotropic diffusion after 1 V-cycle are similar to the results achieved with a linear Gaussian filter shown in the last row of Figure 3. The quality of final restored images using the AMG algorithm with anisotropic diffusion are between those results from Wiener filters and those from TV denoising.

Table 3: Best denoised results using different methods.

	MSSIM of the results using different denosing methods				
images	noisy image	Fixed-point AMG Algorithm 1	Gaussian filters	Wiener filters	TV
Head phantom	0.16	0.76	0.49	0.58	0.83
Satellite	0.16	0.62	0.45	0.50	0.76
Cameraman	0.26	0.74	0.59	0.65	0.76

5 Conclusion

The goal of this paper is to introduce a new regularization parameter-choice approach and investigate the application of the AMG algorithm to image restoration. After reviewing the anisotropic diffusion equation methodology and its application to image restoration, we solve a regularized anisotropic diffusion equation (5), which is not only well-posed, but also has a nontrivial steady-state solution, using AMG Algorithm 1. In order to make the algorithm more adaptive and efficient, we introduce a new automatic regularization parameter-choice method, which combines Brent's method and the NCP information of residual. The application of the Brent-NCP parameter-choice Algorithm 2 is not limited to anisotropic-diffusion based denoising, and is shown to have applicability to a very broad range of parameter-dependent denoising algorithms. The experiments show that the denoised images quality using AMG Algorithm 1 are similar as those results using TV, and better than linear Gaussian filter and Wiener filter.

References

[1] Scott T. Acton. Multigrid anisotropic diffusion. *IEEE Trans. Image Processing*, 7(3):280–290, 1998.

[2] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2006.

[3] A. Brandt, S. McCormick, and J. Ruge. Algebraic multigrid (AMG) for sparse matrix equations. In D. J. Evans, editor, *Sparsity and its Applications*, pages 257–284. CAMBRIDGE, 1984.

[4] William L. Briggs, Van Emden Henson, and Steve F. McCormick. *A multigrid tutorial (2nd ed.)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[5] Antoni Buades, Bartomeu Coll, and Jean Michel Morel. On image denoising methods. Technical report, Technical Note, CMLA (Centre de Mathematiques et de Leurs Applications), 2004.

[6] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[7] F. Catte, P. L. Lions, J. M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.

[8] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[9] Ke Chen and Joseph Savage. An accelerated algebraic multigrid algorithm for total-variation denoising. *Bit Numerical Mathematics*, 47(2):277–296, 6 2007.

[10] J.M. Duarte-Carvajalino, G. Sapiro, M. Velez-Reyes, and P.E. Castillo. Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods. *Geoscience and Remote Sensing, IEEE Transactions on*, 46(8):2418–2434, Aug. 2008.

[11] W. A. Fuller. *Introduction to Statistical Time Series, Second Edition.* John Wiley, New York, NY, 1996.

[12] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing, Second Edition.* Prentice Hall, 2002.

[13] P. C. Hansen, M. Kilmer, and R. H. Kjeldsen. Exploiting residual information in the parameter choice for discrete ill-posed problems. *BIT*, 46:41–59, 2006.

[14] Per Christian Hansen, James G. Nagy, and Dianne P. O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering (Fundamentals of Algorithms 3) (Fundamentals of Algorithms).* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.

[15] Satyanad Kichenassamy. The Perona-Malik paradox. *SIAM J. Appl. Math.*, 57(5):1328–1342, 1997.

[16] K. Niklas Nordström. Biased anisotropic diffusion–a unified regularization and diffusion approach to edge detection. Technical Report UCB/CSD-89-514, EECS Department, University of California, Berkeley, 5 1989.

[17] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *on Computer Vision, Proceedings of IEEE Computer Society workshop*, pages 16–22, 1987.

[18] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 2 edition, October 1992.

[19] J. W. Ruge and K. Stuben. Algebraic multigrid. In S. F. Mccormick, editor, *Multigrid Methods*, pages 73–130, Philidelphia, Pennsylvania, 1987. SIAM.

[20] Bert W Rust and Dianne P O'Leary. Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Problems*, 24(3):034005 (30pp), 2008.

[21] John C. Strikwerda. *Finite Differences And Partial Differential Equations.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2004.

[22] T. Tasdizen and R.T. Whitaker. An efficient geometric multigrid solver for the anisotropic diffusion equation in two and three dimensions. SCI Institute Technical Report UUSCI-2004-002, University of Utah, June 2004.

[23] U. Trottenberg, C.W. Oosterlee, and A. Schüller. *Multigrid.* Academic Press, San Diego, U.S.A., 2001.

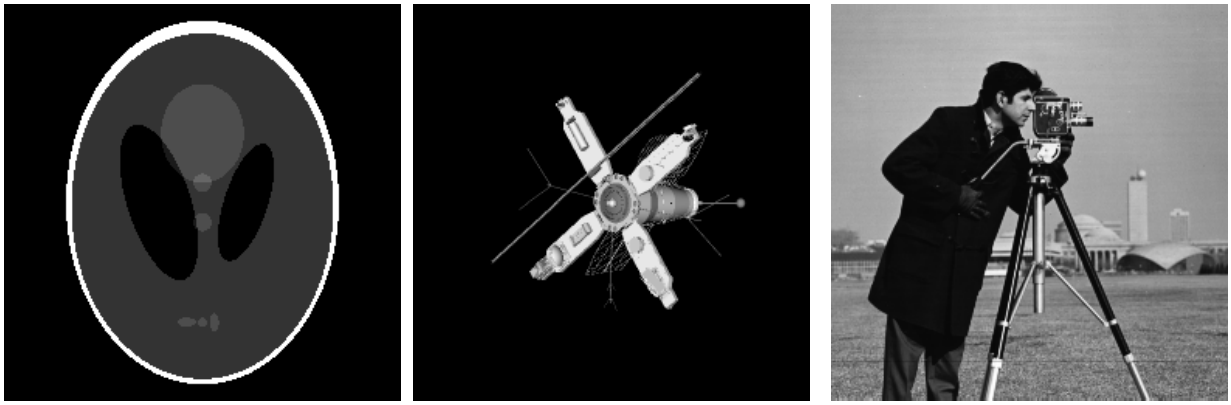
[24] C. R. Vogel. A multigrid method for total variation-based image denoising. *in Computation and Control IV, conference proceedings*, 20, 1995.

[25] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004.

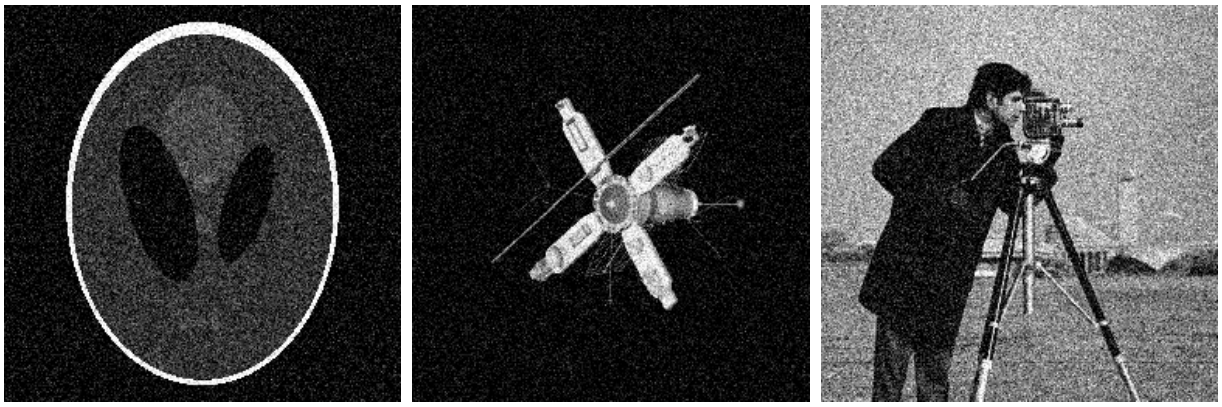
[26] Joachim Weickert. *Anisotropic Diffusion in Image Processing.* PhD thesis, Universität Kaiserslautern, Germany, 1996.

[27] Joachim Weickert. A review of nonlinear diffusion filtering. In *SCALE-SPACE '97: Proceedings of the First International Conference on Scale-Space Theory in Computer Vision*, pages 3–28, London, UK, 1997. Springer-Verlag.

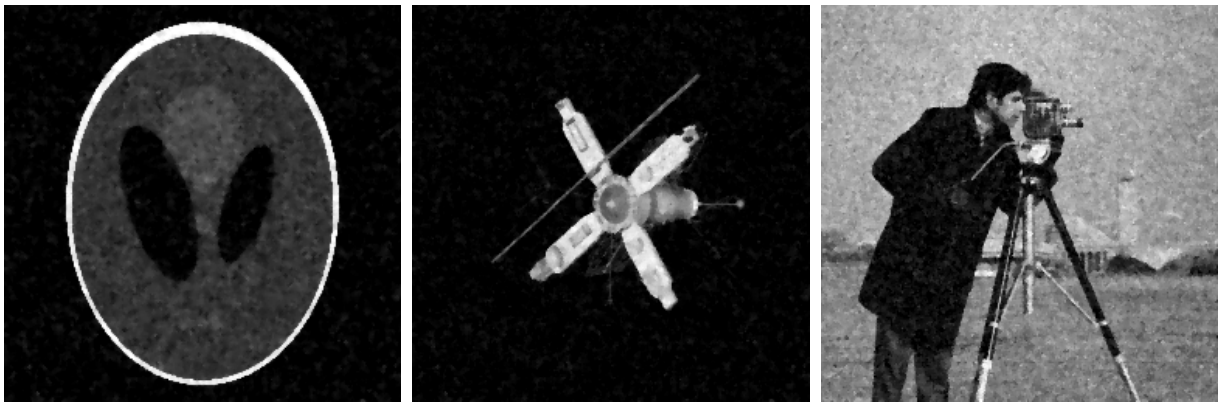
[28] Joachim Weickert and Brahim Benhamouda. A semidiscrete nonlinear scale-space theory and its relation to the perona-malik paradox. In *F. Solina (Ed.), Advances in computer vision*, pages 1–10. Springer, 1997.



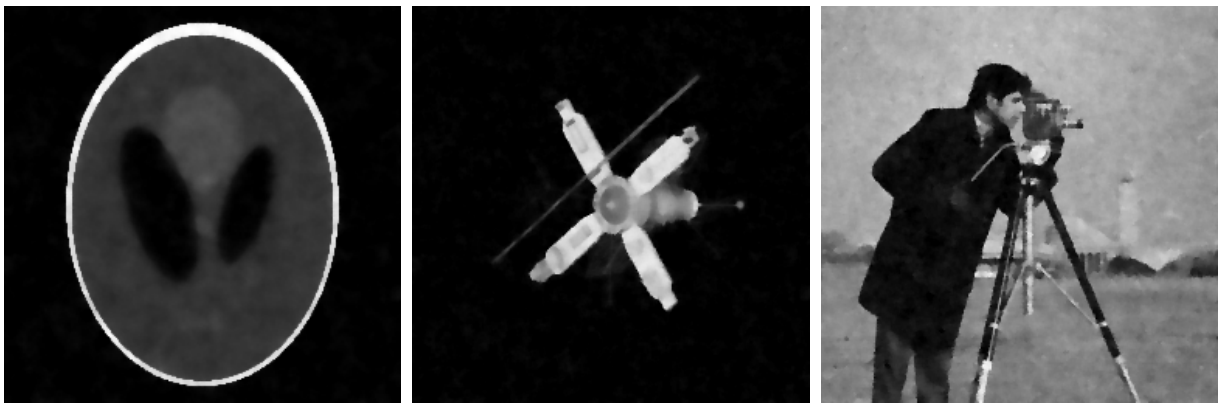
(a) phantom (b) satellite (c) cameraman



(d) noisy phantom, MSSIM = 0.16 (e) noisy satellite, MSSIM = 0.16 (f) noisy cameraman, MSSIM = 0.26

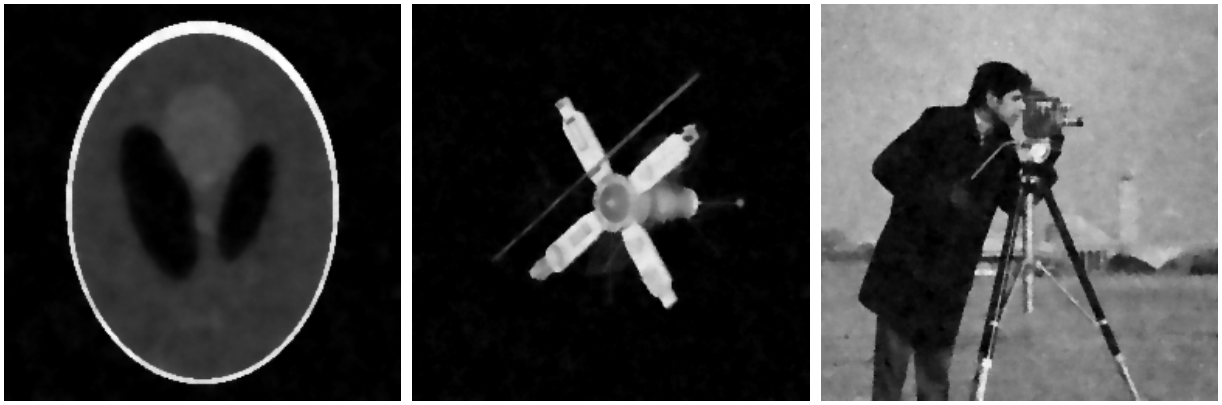


(g) 1 V-cycle, MSSIM = 0.51 (h) 1 V-cycle, MSSIM = 0.40 (i) 1 V-cycle, MSSIM = 0.53

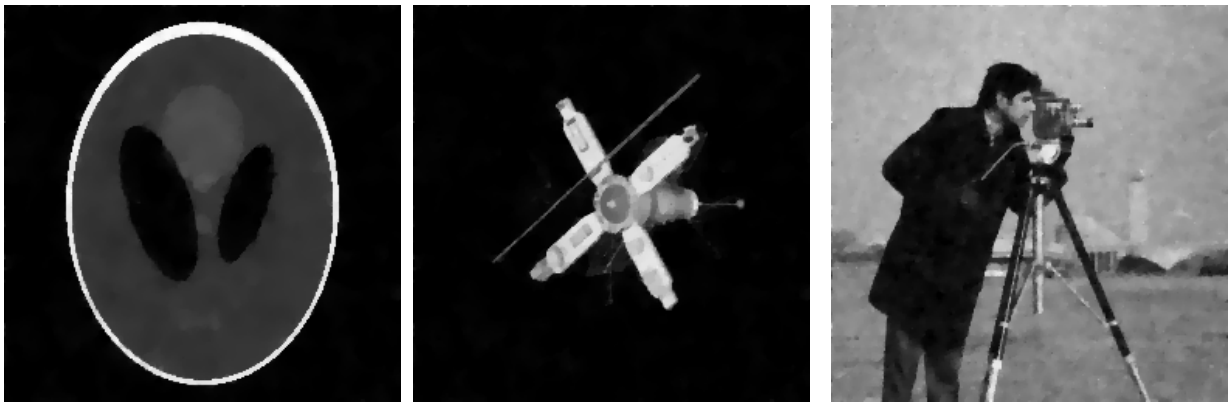


(j) 5 V-cycle, MSSIM = 0.76 (k) 5 V-cycle, MSSIM = 0.62 (l) 3 V-cycle, MSSIM = 0.74

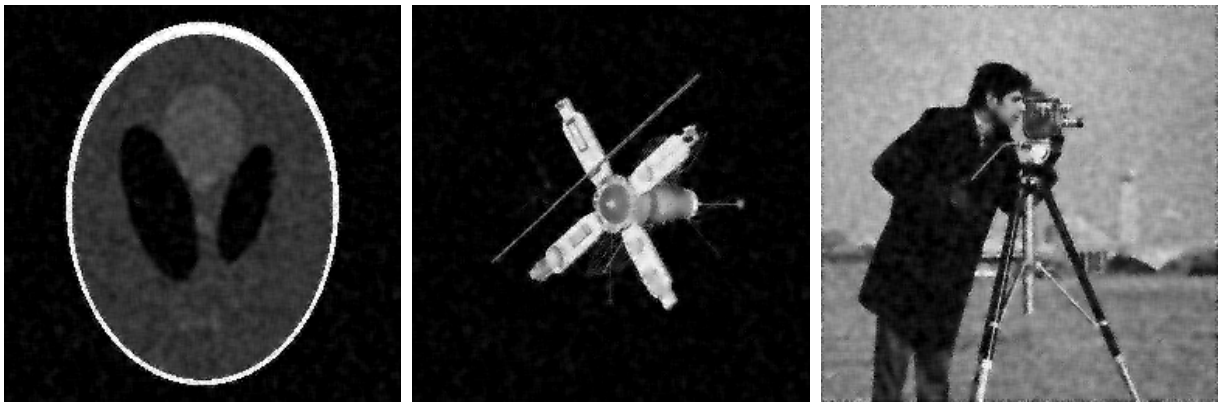
Figure 2: Test images and results by AMG with anisotropic diffusion. First row: original noise-free images. Second row: noisy image. Third row: results by AMG with 1 V-cycle. Last row: final AMG results.



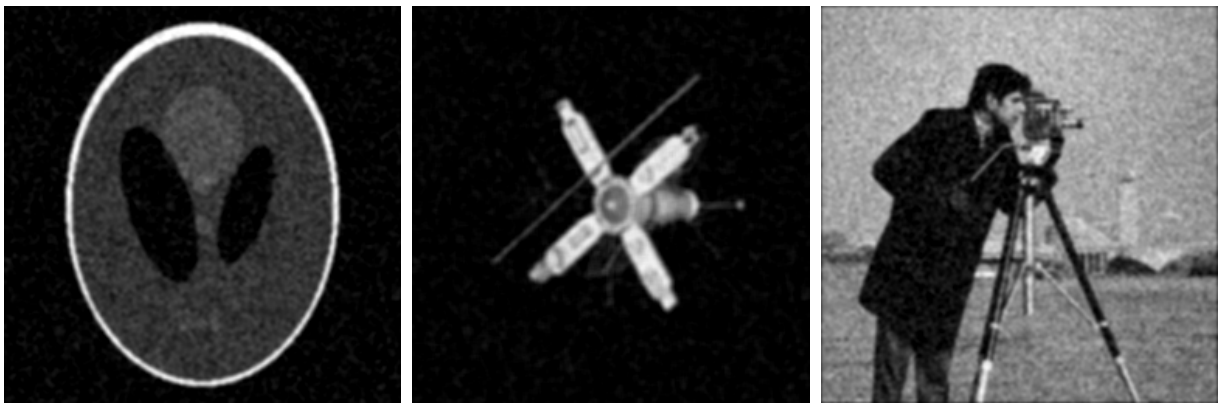
(a) 5 V-cycle, MSSIM = 0.76 (b) 5 V-cycle, MSSIM = 0.62 (c) 3 V-cycle, MSSIM = 0.74



(d) TV, MSSIM = 0.83 (e) TV, MSSIM = 0.76 (f) TV, MSSIM = 0.76



(g) Wiener filter, MSSIM = 0.58 (h) Wiener filter, MSSIM = 0.50 (i) Wiener filter, MSSIM = 0.65



(j) Gaussian filter, MSSIM = 0.49 (k) Gaussian filter, MSSIM = 0.45 (l) Gaussian filter, MSSIM = 0.59

Figure 3: Optimal denoised results for AMG with anisotropic diffusion, TV, Wiener and linear Gaussian filter. First row: final AMG results. Second row: TV results. Third row: Wiener filter results. Last row: linear Gaussian filter results.