Predictor-Corrector Method for Total Variation Based Image Denoising *

Weiguo Li [+] and Lei Wang

School of Mathematics & Computational Sciences, Petroleum University of China, Dongying 257061, China

Abstract. Since their introduction in a classic paper by Rudin, Osher and Fetemi [1], total variation minimizing models have become one of the most popular and successful methodology for image restoration. More recently, there has been a resurgence of interest and exciting new developments, some extending the applicabilities to inpainting, blind deconvolution and vector-valued images, while others offer improvements in better preservation of contrast, geometry and textures, in ameliorating the staircasing effect, and in exploiting the multiscale nature of the models. In addition, new computational methods have been proposed with improved computational speed and robustness. In this paper, a predictor-Corrector techniques are pointed out and applied in to the total variation-based image denoising.The numerical experiments shows the improvement are fairly valid.

KEY WORDS: Predictor-Correct, total variation, image denoising, PDE.

1. Introduction

Variational model have been extremely successful in a wide variety of restoration problems, and remain one of the most active areas of research in mathematical image processing and computer vision. By now, their scope encompasses not only the fundamental problem of image denoising, but also other restoration tasks such as deblurring, blind deconvolution, and inpainting. Variational models exhibit the solution of these problems as minimizers of appropriately chosen functionals. The minimization technique of choice for such models routinely involves the solution of nonlinear partial differential equations (PDEs) derived as necessary optimality conditions.

Perhaps the most basic (fundamental) image restoration problem is denoising. It forms a significant preliminary step in many machine vision tasks, such as object detection and recognition. It is also one of the mathematically most intriguing problems in vision. A major concern in designing image denoising models is to preserve important image features, such as those easily detected by the human visual system, while removing noise. One such important image feature are the edges; these are places in an image where there is sharp change in image properties, which happens for instance at object boundaries. A great deal of research has going into designing models for removing noise while preserving edges; recently there has also been a lot of effort in preserving other fine scale image features, such as texture. All successful denoising models take advantage of the fact that there is an inherent regularity found in natural images; this is how they attempt to tell apart noise and actual image information. Variational and PDE based models make it particularly easy to impose geometric regularity on the solutions obtained as denoised images, such as smoothness of boundaries. This is one of the main reasons behind their success.

Total variation based image restoration models were first introduced by Rudin, Osher, and Fatemi (ROF) in their pioneering work [1] on edge preserving image denoising. It is one of the earliest and best known examples of PDE based edge preserving denoising. It was designed with the explicit goal of preserving sharp discontinuities (edges) in images while removing noise and other unwanted fine scale detail. Being convex, the ROF model is one of the simplest variational models having this most desirable property. The revolutionary aspect of this model is its regularization term that allows for discontinuities but at the same time disfavors oscillations. It was originally formulated in [1] for gray-scale imagery in the following form:

$$\inf_{\int_{\Omega}(u-f)^2 dx = \sigma^2} \int_{\Omega} |\nabla u| dx \tag{1.1}$$

Here, Ω denotes the image domain (for instance, the computer screen), and is usually a rectangle. The function $f(x): \Omega \to R$ represents the given observed image, which is assumed to be corrupted by Gaussian noise of variance σ^2. The constraint of the optimization forces the minimization to take place over images that are consistent with this known noise level. The objective functional itself is called the total variation (TV) of the function $u(x)$; for smooth images it is equivalent to the L^1 norm of the derivative, and hence is some measure of the amount of oscillation found in the function $u(x)$. Optimization problem (1) is equivalent to the following unconstrained optimization, which was also first introduced in [1]:

$$\inf_{u \in L^2(\Omega)} \int_{\Omega} |\nabla u| dx + \lambda \int_{\Omega} (u-f)^2 dx \tag{1.2}$$

Here, $\lambda \geq 0$ is a Lagrange multiplier. The equivalent of problem (1.1) and (1.2) has been established in [2]. In the original paper [1] there is an iterative numerical procedure given for choosing $u(x)$ so that the solution $u(x)$ obtained solves (1.1).

We point out that total variation based energies appear, and have been previously studied in, many different areas of pure and applied mathematics. For instance, the notion of total variation of a function and functions of bounded variables appear in the theory of minimal surfaces. In applied mathematics, total variation based models and analysis appear in more classical applications such as elasticity and fluid dynamics. Due to ROF, this notion has now become central also in image processing.

Over the years, the ROF model has been extended to many other image restoration tasks, and has been modified in a variety of ways to improve its performance. In this paper, we will concentrate on numerical algorithms proposed for minimizing the ROF objective. Recently, most of algorithm in this topic fall into the three main approaches, namely, direct optimization, solving the associated Euler-Lagrange equations and using the dual variable explicitly in the solution process to overcome some computational difficulties encountered in the primal problem. We will focus on the last approach in section 2. And a predictor-corrector algorithm is given in section 3. In last section, some numerical examples are given.

2. A Primal-Dual Method for TV Image Restoration

In their original paper [1], Rudin et al. proposed the use of artificial time marching to solve the Euler-Lagrange equations which is equivalent to the steepest descent of the energy function. More precisely, consider the images as a function of space and time and seek the steady state of the equation

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|_{\beta}} \right) - 2\lambda(u - f) \tag{2.1}$$

Here, $|\nabla u|_{\beta} := \sqrt{|\nabla u| + \beta}$ is a regularized version of $|\nabla u|$ to reduce degeneracies in flat regions, where $|\nabla u| \approx 0$. In numerical implementation, an explicit time marching scheme with time step Δt and space step size Δx is used. Under this method, objective value of the ROF model is guaranteed to be decreasing and the solution will tend to the unique minimizer as time increases. However, the convergence is usually slow due to the Courant-Friedrichs-Lewy (CFL) condition, $\Delta t \leq c\Delta x^2 |\nabla u|$ for some constant $c > 0$, imposed on the size of the time step, especially in flat regions where $|\nabla u| \approx 0$.

To completely get rid of CFL conditions, Vogel and Oman proposed in [3] a fixed point iteration scheme (FP) which solves the stationary Euler-Lagrange directly. The Euler-Lagrange equation is linearized by lagging the diffusion coefficient and thus the $i+1-th$ iterate is obtained by solving the sparse linear equation:

$$\nabla \cdot \left(\frac{\nabla u^{i+1}}{|\nabla u^i|_{\beta}} \right) - \lambda(u^{i+1} - f) = 0 \tag{2.2}$$

While this method converges only linearly, empirically, only a few iterations are needed to achieve visual accuracy. In practice, one typically employs specifically designed solvers to solve (2.2) in each iteration.

Although regularization by $1/\left|\nabla u\right|_\beta$ avoids the coefficient of the parabolic term becoming arbitrarily large, the use of a large enough β for effective regularization will reduce the ability of the ROF model to preserve edges.

Due to the presence of the highly nonlinear term $\nabla \cdot \left(\dfrac{\nabla u}{\sqrt{\left|\nabla u\right|^2 + \beta}} \right)$, Newton's method does not work satisfactorily in the sense that its domain of convergence is very small. This is specially true if the regularizing parameter β is small. On the other hand, if β is relatively large, then this term is well behaved. So it is natural to use a continuation procedure stating with a large value of β and gradually reducing it to the desired value. Chan et al. proposed such an approach in [4]. Although the choice of the sequence of subproblem to solve is crucial for its efficiency, this method is locally quadratically.

Chan et al. in [5], Carter in [6] and Chambolle in [7] exploit the dual formulation of the ROF model by using the identity $\left\|x\right\| \equiv \sup_{\|g\|\leq 1} x \cdot g$ for vectors in Euclidean spaces and treating g as the dual variable, one arrives at the dual formulation:

$$\sup_{g \in C_C^1(\Omega, B^2)} \int_\Omega f \nabla \cdot g \, dx - \frac{1}{2\lambda}\int_\Omega (\nabla \cdot g)^2 \, dx \tag{2.3}$$

Where B^2 is the unit disk in R^2. Once g is obtained, the primal variable can be recovered by $u = f - \lambda^{-1}\nabla \cdot g$. A promise of dual formulation is that the objective function is differentiable in g, unlike the primal problem which is badly behaved when $\nabla u = 0$. However, the optimization problem becomes a constrained one which requires additional complexity to solve. The approach used in [5] solves for u and g simultaneously. Its derivation starts by treating the term $\nabla u/\left|\nabla u\right|$ in the primal Euler-Lagrange equation as an independent variable g, leading to the system:

$$\begin{cases} -\nabla \cdot g + \lambda(u - f) = 0 \\ g\left|\nabla u\right|_\beta - \nabla u = 0 \end{cases} \tag{2.4}$$

The above system of nonlinear equations is solved by Newton's method and quadratic convergence rate is almost always achieved. In the Newton updates, one may combine the two equations to eliminate the fixed point iteration (2.2). Empirically, this primal-dual method is much more robust than applying Newton's method directly to the primal problem in u only.

3. Predictor-Corrector method for total variation image denoising

United the primal-dual method and continuation method, we point out a predictor-corrector method for total variation-based image denoising. Because for a big β, the Newton method converges very fast we can use the solution for the larger β as the initial guess for the method with the smaller β. Our strategy is to decrease β gradually as to the given positive number β^* at which we want to solve the equation (2.4).

In fact, it's easily to get the whole procedure.

Algorithm 1. (Predictor-Corrector)

(i) Set β_0 and use $u_0 = f$ the observed images as the initial guess;

(ii) Use Primal-Dual method to find the solution u_0^*;

(iii) Do i=1,2,…,

    (a) if $\beta_{i+1} > \max(\sigma\beta_i)$, set $\beta_{i+1} = \beta_i$, tol=ε_0,

else $\beta_{i+1} = \beta^*$, tol $= \varepsilon_1$;

(b) use Primal-Dual method to find the solution u_{i+1} .

Generally, we choice ε_0 is bigger than ε . Such as $\varepsilon_0 = 10^{-2}$ and $\varepsilon = 10^{-5}$, because our aim is to provide a predictor for next iteration. The last step is real corrector. But the choice of $\sigma < 1$ depends on the problems that we want to solve. Such as $\sigma = 10^{-5}$. Experiments show this algorithm has a better behavior than original Primal-Dual method [5] whatever computing time or effect of denoising. Of course, this predictor-corrector algorithm has made a great progress compared to original continuation method [4].

If we would like to improve the corrector, we can also use the following corrector model:

$$\inf_{u \in L^2(\Omega)} \alpha \int_\Omega |\nabla u|_{\beta_1} dx + \mu \int_\Omega |\nabla(u - f)|_{\beta_2} dx + \frac{1}{2}\int_\Omega (u - f)^2 dx \tag{3.1}$$

This model retains the properties of ROF model-- recover the edges of the original image and regularize the geometry of level sets without penalizing discontinuities. And it allows us to reduce the error of edges between original and restoration images. Its corresponding Euler-Lagrange equation is

$$g(u) \equiv -\alpha \nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta_1}} \right) - \mu \nabla \cdot \left(\frac{\nabla(u - f)}{\sqrt{|\nabla(u - f)|^2 + \beta_2}} \right) + u - f = 0 \tag{3.2}$$

The idea to solve (3.2) is to introduce new variables w_1 and w_2 :

$$w_1 = \frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta_1}}, \qquad w_2 = \frac{\nabla(u - f)}{\sqrt{|\nabla(u - f)|^2 + \beta_2}} \tag{3.3}$$

and replace (3.2) by the following equivalent system of nonlinear PDEs

$$\begin{cases} h_1(u, w_1, w_2) \equiv -\alpha \nabla \cdot w_1 - \mu \nabla \cdot w_2 + u - f = 0, \\ h_2(u, w_1, w_2) \equiv w_1 \sqrt{|\nabla u|^2 + \beta_1} - \nabla u = 0, \\ h_3(u, w_1, w_2) \equiv w_2 \sqrt{|\nabla(u - f)|^2 + \beta_2} - \nabla(u - f) = 0. \end{cases} \tag{3.4}$$

Now we linearize this (u, w_1, w_2) system by Newton method.

$$\begin{pmatrix} I_N & -\alpha\nabla\cdot & -\mu\nabla\cdot \\ -\left(I_{2N} - \dfrac{w_1\nabla u^T}{\sqrt{|\nabla u|^2 + \beta_1}}\right)\nabla & \sqrt{|\nabla u|^2 + \beta_1} & 0 \\ -\left(I_{2N} - \dfrac{w_2\nabla(u - f)^T}{\sqrt{|\nabla(u - f)|^2 + \beta_2}}\right)\nabla & 0 & \sqrt{|\nabla(u - f)|^2 + \beta_2} \end{pmatrix} \begin{pmatrix} \delta u \\ \delta w_1 \\ \delta w_2 \end{pmatrix} = -\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \tag{3.5}$$

Equation (3.5) can be solved by elimination $\delta w_1, \delta w_2$ and solving the result equation for δu

$$\left[I_N - \alpha\nabla \cdot \left(\frac{1}{\sqrt{|\nabla u|^2 + \beta_1}} \left(I_{2N} - \frac{w_1\nabla u^T}{\sqrt{|\nabla u|^2 + \beta_1}} \right)\nabla \right) \right.$$

$$\left. -u\nabla \cdot \left(\frac{1}{\sqrt{|\nabla(u - f)|^2 + \beta_2}} \left(I_{2N} - \frac{w_1\nabla(u - f)^T}{\sqrt{|\nabla(u - f)|^2 + \beta_2}} \right)\nabla \right) \right]\delta u = -g(u) \tag{3.6}$$

After δu is obtained we can compute $\delta w_1, \delta w_2$ by the following equations

$$\delta w_1 = \frac{1}{\sqrt{|\nabla u|^2 + \beta_1}}\left(I_{2N} - \frac{w_1 \nabla u^T}{\sqrt{|\nabla u|^2 + \beta_1}}\right)\nabla \delta u - w_1 + \frac{w_1 \nabla u^T}{\sqrt{|\nabla u|^2 + \beta_1}}; \qquad (3.7)$$

$$\delta w_2 = \frac{1}{\sqrt{|\nabla(u-f)|^2 + \beta_2}}\left(I_{2N} - \frac{w_2 \nabla(u-f)^T}{\sqrt{|\nabla(u-f)|^2 + \beta_2}}\right)\nabla \delta u - w_2 + \frac{w_2 \nabla(u-f)^T}{\sqrt{|\nabla(u-f)|^2 + \beta_2}}. \qquad (3.8)$$

In our experiments, μ is often smaller than λ.

Algorithm 2. (Corrector Algorithm)

(i) Choose α and μ and make sure the predictor value u_1 of image u.

(ii) Use the formula (3.6), (3.7) and (3.8) to compute δu, δw_1, δw_2 respectively.

(iii) If $\|h_1\|_{new}/\|h_1\|_{old} < tol$, stop, else go to (ii).

4. Numerical Examples

In this section we present results of our denoising algorithms on three images. The noise image f 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 $\|u-f\|_F/\|u\|_F = \sigma$. In our examples, we choose $\sigma = 0.4$.

Our first image is a gray-scale image with 256×256 pixels and range [0,255] to which Gaussian white noise is added. The original and noisy images are shown below (Figure 1). In this experiment we take Ω to be unit square. We take values of 30 and 10^{-20} for α and β, and $tol = 10^{-4}$. The processed images from Primal-Dual method and Predictor-Corrector method are displayed in Figure 1 and computing time are 1203.6 seconds (53 steps) and 758.8 seconds (38 steps) separately.
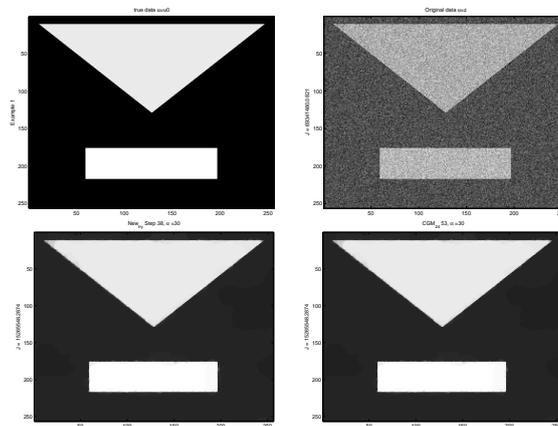

Figure 1 Top left is original image and top right is noise image.
$\alpha = 30, \beta = 10^{-20}, tol = 10^{-4}$. Bottom left is from P-D and bottom right from P-C.

In our second experiment we compare the performance of P-D and our P-C for various sizes of the image. We apply both methods to various sizes of the image shown in Figure 1. As before we have $\sigma = 0.4$. We take values of 30 and 10^{-4} for the parameters α and tol in all experiments. The inner linear steps in the primal-dual method are done using preconditioned conjugate gradient with incomplete Cholesky pre-conditioner, on step k the stop criteria for these inner steps is a relative decrease of the linear residual by a factor of $\min(0.1, 0.9\|r_{k-1}\|/\|r_{k-2}\|)$ where r is the nonlinear residual, as specified in [4]. The initial guess for all methods is the noisy image f and the stopping criteria for all methods is a decrease in the relative

residual by a factor of 10^{-4}. All the results for comparison can be found in Table 1.

Table 1.

n	P-D		P-C	
	steps	cpu time(s)	steps	cpu time(s)
$256(\beta = 10^{-2})$	12	77.6	11	62.2
$256(\beta = 10^{-3})$	14	114	12	99.9
$512(\beta = 10^{-2})$	13	390.2	12	298.9
$512(\beta = 10^{-3})$	15	674.9	14	540.3
$1024(\beta = 10^{-2})$	13	1617	12	1299
$1024(\beta = 10^{-3})$	16	3037	14	2347

Next we test our method on 2 more realistic images, the Lenna image and the satellite image with $SNR = 3$ and $SNR = 1$. All images are 256×256 in size and have range [0,255]. $\beta = 10^{-2}$ in all cases, $\alpha = 30$ and 50 for Lenna and 50 for the Satellite. As before the initial guess is the noisy image and the stopping criteria is a decrease in the relative residual by a factor of 10^{-4} and 10^{-8}. The results are given by Table 2 below.

Table 2.

Image	P-D		P-C	
	steps	Cpu time(s)	steps	cpu time(s)
Lenna (SNR 3)	14	91	13	75
Lenna (SNR 1)	19	156	18	128
Satellite(SNR 4)	16	123.1	14	89.4
Satellite(SNR 1)	16	122.7	14	89.9

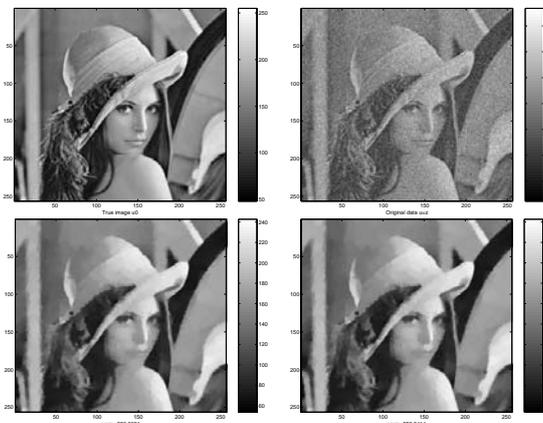

Figure 2 (Lenna) Top left is original image and top right is noise image.
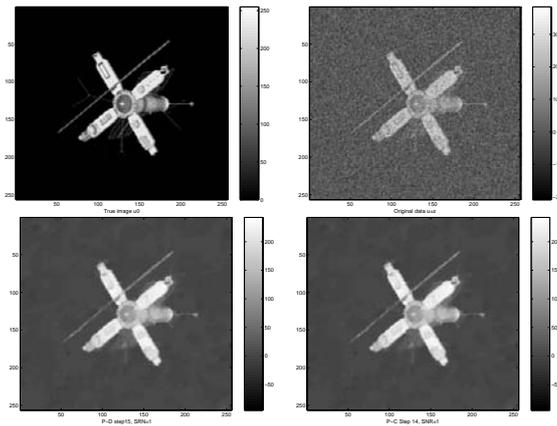$\alpha = 30, \beta = 10^{-20}, tol = 10^{-4}$. Bottom left is from P-D and bottom right from P-C.

Figure 3 (Satellite) Top left is original image and top right is noise image.
$\alpha = 50, \beta = 10^{-20}, tol = 10^{-4}$. Bottom left is from P-D and bottom right from P-C.

5. Conclusions

The standard Primal-Dual method worked well in denoising problems but we want to improve its performance in the convergence. With our recommended Predictor-Corrector methods we observe it gives about a 20% improvement. So our method is efficient and robust. Furthermore, we suggest that the predictor-corrector Newton method can be used directly if β is small (such as $\beta < 10^{-4}$), or the truncated versions of Newton (predictor-corrector) algorithm based on the conjugate gradient method with incomplete Cholesky as pre-conditioner. We have known the multi-grid method with the use of the Krylov acceleration procedure made the convergence fast and it is the same as our work in some extent. To improve the effect of image denoising, other models must be considered future.

6. References

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

[2]　A. Chambolle and P. Lions, Image Recovery via Total Variation Minimization and Related Problems, Numer. Math., 76(1997), 167-188.

[3]　C. Vogel and M. Oman, Iteration Methods for Total Variation Denoising, SIAM J. Sci. Comp., 17(1996), 227-238.

[4]　R. H. Chan, T. F. Chan, and H. M. Zhou, Advanced Signal Processing Algorithms, in Proceedings of the International Society of Photo-Optical Instrumentation Engeers, F. T. Luk, ed., SPIE, 1995, 314-325.

[5]　T. Chan, G. Golub, and P. Mulet, A Nonlinear Primal-Dual Method for Total Variation-based Image Restoration. SIAM J. Sci. Comp., 20(1999), 1964-1977.

[6]　J. Carter, Dual Methods for Total Variation-based Image Retoration, PhD thesis, UCLA, Los Angeles, CA, 2001.

[7]　A. Chambolle, An Algorithm for Total Variation Minimization and Applications, J. Math. Imaging Vision, 20(2004), 89-97.

[8]　C. R. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.

[9]　Qianshun Chang and I-Lian Chern, Acceleration Method for Total Variation-Based Image Denoising, SIAM J. Sci. Comput., 25(2003), 982-994.

[10]　D. Strong and T. Chan, Edge-preserving and Scale-dependent Properties of Total Variation Regularization, Inv. Probl., 19(2003), 165-187.