Nonlocal Convex Functionals for Image Regularization

Guy Gilboa[1] *, Jerome Darbon[1] **, Stanley Osher[1] *, and Tony Chan[1] **

Department of Mathematics, UCLA, Los Angeles, CA 90095, USA.

Abstract. We examine weighted nonlocal convex functionals. The weights determine the affinities between different regions in the image and are computed according to image features. The L^1 energy of this type can be viewed as a nonlocal extension of total-variation. Thus we obtain nonlocal versions of ROF, TV-flow, Bregman iterations and inverse-scale-space (based on nonlocal ROF). Constructing the weights using patch distances, similarly to the nonlocal-means of Buades-Coll-Morel results in very robust and powerful regularizations. The flows and minimizations are computed efficiently by extending some recently proposed graph-cuts techniques. Numerical results which illustrate the performance of such models are presented.

1 Introduction

The application of partial differential equations and variational methods to image and signal processing have shown to be very effective, allowing high quality regularization procedures. The energy functionals are well understood and a rigorous theory can often be established regarding their properties. There are, however, some limitations to functionals such as total variation [20]. These type of functionals are based solely on derivatives which are local features of the signal. Moreover, it is hard to adapt the process to more specific tasks or encode *a priori* information (without introducing additional energy terms). In this paper we aim to extend the variational framework to a nonlocal one and allow much more flexibility in the regularization, yet keeping a simple convex framework.

Our approach can be viewed as a continuous generalization of graphs and relates to concepts from spectral graph theory [9, 18]. A fundamental linear operator used in this field is the graph Laplacian. We give a more general nonlinear convex viewpoint and suggest a large family of operators which may be used. As the weights are computed according to noisy data they may have outliers which can be well handled by L^1-based regularizations. Our present study may be seen also as a generalization of an early work [8] regarding digital TV.

* Supported by grants from the NSF under contracts ITR ACI-0321917, DMS-0312222, and the NIH under contract P20 MH65166.

** Supported by grants from the ONR under contracts ONR N00014-06-1-0345, the NIH from the grant NIH U54-RR021813 and the NSF from grant NSF DMS-0610079.

The use of spectral graph theory for image segmentation was suggested by Shi and Malik [21] and influenced many studies (e.g. [23, 25, 14]). A harmonic analysis viewpoint for general high dimensional applications is given in [10].

In [4] the following nonlocal filter, referred to as *nonlocal means*, was suggested for image denoising:

$$NL(u)(x) = \frac{1}{c(x)} \int_{\Omega} e^{-d_a(u(x),u(y))/2h^2} u(y) dy \tag{1}$$

where

$$d_a(u(x), u(y)) = \int_{\Omega} G_a(t) |u(x+t) - u(y+t)|^2 dt \tag{2}$$

G_a is a Gaussian with standard deviation a, and $c(x)$ is a normalization factor: $c(x) = \int_{\Omega} e^{-d_a(u(x),u(y))/h^2} dy$. A variational understanding of the filter was first presented in [15] as a nonconvex functional and later in [13] as a convex quadratic functional. The work of [22] was the first to show how nonlocal means can be understood within the diffusion geometry framework of [10]. In our numerical experiments the weights of the functional are computed according to Eq. (2).

We first present the regularizing functonials before describing the numerical algorithms. In paper we focus on TV-based regularization and we show how such an optimization can be performed using a graph-cut based approach [3, 16]. In [12], the authors show how to solve classical ROF [20] minimization problem using graph-cuts. This method is shown to be very fast. In this paper we extend this approach to the proposed non-local model. We also describe how one can compute a Total Variation flow [1] and its non-local extension.

2 The Regularizing Functional

We begin by defining the following general convex nonlocal functional:

$$J(u) := \frac{1}{2} \int_{\Omega \times \Omega} \phi(|u(x) - u(y)|) w(x, y) dx dy, \tag{3}$$

where $\Omega \in \mathbb{R}^n$, $x = (x_1, \ldots, x_n) \in \Omega$ and $y = (y_1, \ldots, y_n) \in \Omega$. For images we have $n = 2$. The function ϕ is convex, positive and we assume $\phi(0) = 0$. For L^1 type functionals we further assume $\lim_{s \to \infty} \phi(s)/s = 1$. The weight function $w(x, y) \in \Omega \times \Omega$ is positive: $w(x, y) \geq 0$ and symmetric: $w(x, y) = w(y, x)$. For image processing tasks the weight function is based on image features and can be understood as the proximity between two points x and y, based on features in their neighborhood. For more details on the way to obtain such functions along with a few examples see [13].

Let $p(u)$ be a subgradient element of J:

$$p(u)(x) \in J'(u)(x) = \int_{\Omega} \phi'(|u(x) - u(y)|) \frac{u(x) - u(y)}{|u(x) - u(y)|} w(x, y) dy. \tag{4}$$

In [13] the quadratic form was investigated $\phi(s) = \frac{1}{2}s^2$, with the corresponding linear operator L:

$$Lu(x) = -p(u)(x)|_{\phi(s)=\frac{1}{2}s^2} = \int_\Omega (u(y) - u(x))w(x,y)dy. \tag{5}$$

We first define a flow for minimizing the general energy $J(u)$. Taking the input image f as the initial condition, a steepest descent for the functional defined in (3) is

$$u_t(x) = -p(u)(x), \quad u_{t=0} = f(x). \tag{6}$$

For $\phi(s) = s$ the flow may be interpreted as a nonlocal TV flow [1].

2.1 Variational Denoising

In the usual way, one can add a convex fidelity term to the convex functional J. For the L^2 fidelity, the denoised image u is the minimizer of

$$E(u,f) = J(u) + \frac{\lambda}{2}\|u - f\|_2^2, \tag{7}$$

and u satisfies the Euler-Lagrange equation

$$p(u) + \lambda(u - f) = 0. \tag{8}$$

Eq. (3) with $\phi(s) = s$ will be referred to as NL-ROF for the rest of this paper. As commonly done, one can also view this as a constrained problem:

$$u := \arg\min J(u), \quad \text{s.t. } \|u - f\|_2^2 = |\Omega|\sigma_n^2, \tag{9}$$

where σ_n^2 is the variance of an additive noise in a noisy image f. Then λ is viewed as a Lagrange multiplier and one can compute the constrained problem by initializing e.g. with $u|_{t=0} = f$, $\lambda = 1$ and iterating:

$$u_t = -p(u) + \lambda(f - u), \tag{10}$$

$$\lambda = \frac{1}{|\Omega|\sigma_n^2}\int_\Omega (f - u)p(u)dx, \tag{11}$$

using the gradient projection method, as in [20]. We also propose a dichotomic approach for computing λ in section 3.

2.2 Bregman Iteration and Inverse Scale Space

In [5],[7] we developed an inverse scale space (ISS) and relaxed inverse scale space approach to denoising. It was based on the continuous limit of Bregman iterations devised in [19], not only for denoising but for blind deconvolution [17] and other reconstruction tasks [24],[2] as well as being a useful theoretical device to obtain sharp estimates in standard reconstruction methods [6].

Briefly, to reconstruct an image from a given data f we begin with the variational problem

$$u_1 = \arg\min_u (J(u) + \lambda H(u, f))$$

(see [19] for more details). This leads to the sequence

$$u_k = \arg\min_u (J(u) - J(u_{k-1}) - < u - u_{k-1}, p_{k-1} > + \lambda H(u, f))$$

$k = 1, 2, \ldots$, with $u_0 = 0$ and $J(0) = p_0 = 0$, for $\lambda > 0$, where $p_{k-1} = p(u_{k-1})$. Under reasonable hypotheses it follows that $u_k \to \tilde{u}$ the minimizer of $H(u, f)$ monotonically, but, more importantly, the Bregman distance between u_k and g, g a "denoised version of $\tilde{u}$", which means $J(g) < \infty$, decreases until u_k gets too close to $\tilde{u}$. See [19] for the precise results.

The Bregman distance is defined by

$$D(g, u) = J(g) - J(u) - < g - u, p(u) > .$$

For $J(u)$ defined by (3) with $\phi(s) = s$, we have

$$D(g, u) = \frac{1}{2} \int_{\Omega \times \Omega} \left(|g(x) - g(y)| - \frac{(g(x) - g(y))(u(x) - u(y))}{|u(x) - u(y)|} \right) w(x, y) dx dy$$

The inverse scale space equation associated with (7) is

$$\frac{d}{dt} p(u) = (f - u)$$

with $u(0) = 0$ and the normalization $\int u = \int f = 0$ is required.

Solving this involves inverting p and evaluating it at every time step, which is computationally nontrivial.

Instead we use the relaxed inverse scale space approximation [5]:

$$\frac{\partial u}{\partial t} = -p(u) + \lambda(f - u + v)$$
$$\frac{\partial v}{\partial t} = \alpha(f - u)$$

for $\lambda, \alpha > 0,\ \alpha \leq \frac{\lambda}{4}$ with initial data $u(0) = v(0) = 0$.

2.3 Preliminary analysis

In the discussion below we require the following condition

$$\text{The associated linear operator } -L \text{ has a zero eigenvalue of multiplicity 1.} \tag{12}$$

Where L is defined in (5). This condition ensures connected weights (in a similar manner to graphs [9]):

Lemma 1. *Condition* (12) *holds if and only if for any two points x, y there exists a sequence: $z_1, \ldots, z_k$ such that $w(x, z_1)w(z_1, z_2)w(z_k, y) > 0$ (that is, every element in the sequence is strictly positive).*

For a proof see [13]. We analyze the simpler case of strictly convex functionals. Thus $\phi(s)$ is strictly convex in s, the subgradient is unique and $p(0) \equiv 0$. Many of the following results, however, can be extended to the general convex case.

Proposition 1 *The operator $p(u)$ as defined in Eq. (4) with strictly convex ϕ and satisfying condition (12) admits the following properties:*

(a) $u \equiv const$ *iff* $p(u) \equiv 0$.
(b) For any maximal point $u(x_0) \geq u(x)$, $\forall x \in \Omega$, $p(u)(x_0) \geq 0$. *Similarly for any minimal point* $u(x_1) \leq u(x)$, $\forall x \in \Omega$, $p(u)(x_1) \leq 0$.
(c) $p(u)$ *is a positive semidefinite operator, that is* $\langle p(u), u \rangle \geq 0$.
(d) $\int_\Omega p(u)(x)dx = 0$.

Proof. The first part of Property (a) is valid since for any constant u we have $s = |u(x) - u(y)| \equiv 0$. We get $p(u \equiv const) = p(0) = 0$. For the second part it is easy to see that for a given point x we have $p(u)(x) = 0$ if for all $y \in \Omega$ either $u(x) = u(y)$ or $w(x, y) = 0$. Using the connectivity condition (12) we reach the conclusion $u(x) = u(y)$, $\forall x, y$ (or else we get a contradiction $p(u) \neq 0$, for more details see the proof of Lemma 1).

Property (b) is seen using $\frac{\phi'(s)}{s} \geq 0$, $w(x, y) \geq 0$.

Property (c) can be validated (relying on the symmetry $w(x, y) = w(y, x)$ and $\frac{\phi'(s)}{s} \geq 0$) by:

$$\begin{aligned}
\langle p(u), u \rangle &= \int_{\Omega\times\Omega} \phi'(|u(x) - u(y)|) \tfrac{u(x)-u(y)}{|u(x)-u(y)|} w(x, y)u(x)dydx \\
&= \tfrac{1}{2} \int_{\Omega\times\Omega} [(u(x) - u(y))u(x) \tfrac{\phi'(|u(x)-u(y)|)}{|u(x)-u(y)|} w(x, y) + \\
&\quad + (u(y) - u(x))u(y) \tfrac{\phi'(|u(y)-u(x)|)}{|u(y)-u(x)|} w(y, x)]dydx \\
&= \tfrac{1}{2} \int_{\Omega\times\Omega} (u(x) - u(y))^2 \tfrac{\phi'(|u(x)-u(y)|)}{|u(x)-u(y)|} w(x, y)dydx \;\geq\; 0.
\end{aligned}$$

Property (d) is shown by

$$\begin{aligned}
\int_\Omega p(u)(x)dx = \tfrac{1}{2} \int_{\Omega\times\Omega} [&(u(x) - u(y)) \tfrac{\phi'(|u(x)-u(y)|)}{|u(x)-u(y)|} w(x, y) + \\
&(u(y) - u(x)) \tfrac{\phi'(|u(y)-u(x)|)}{|u(y)-u(x)|} w(y, x)]dydx = 0.
\end{aligned}$$

□

2.4 Basic properties of the flow

We would like to establish some results regarding the flow (6). We examine the strictly convex case $\phi(s) = \sqrt{s^2 + \epsilon^2} - \epsilon$. Thus

$$p(u)(x) = \int_\Omega \frac{u(x) - u(y)}{\sqrt{(u(x) - u(y))^2 + \epsilon^2}} w(x, y)dy. \tag{13}$$

We assume condition (12) is satisfied, thus we can use the properties of Proposition 1.

Proposition 2 *The flow* (6) *with* $p(u)$ *satisfying the conditions of Proposition 1 admits the following properties:*

(i) *The mean value is preserved,* $\frac{1}{|\Omega|}\int_\Omega u(x,t)dx = \frac{1}{|\Omega|}\int_\Omega f(x)dx,\ \forall t \geq 0$.
(ii) *The extremum principle holds,* $\min_x(f(x)) \leq u(x,t) \leq \max_x(f(x)),\ \forall x \in \Omega,\ \forall t \geq 0$.
(iii) *The solution converges to a constant,* $u(x, t \to \infty) \equiv const = \int_\Omega f(x)dx$.
(iv) *The following estimate holds:* $\frac{1}{2}\frac{d}{dt}\|u(x,t)\|^2_{L^2} \leq 0$.

Proof. (i) can be shown by computing the time derivative of the mean value and using Property (d):

$$\frac{d}{dt}\Big(\frac{1}{|\Omega|}\int_\Omega u(x)dx\Big) = -\frac{1}{|\Omega|}\int_\Omega p(u)(x)dx = 0.$$

(ii) is validated by Property (b) as any point x where $u(x)$ is maximal is non-increasing with time, and, similarly, any point x where $u(x)$ is minimal is non-decreasing with time.

Let us first validate (iv). Using Property (c) we obtain:

$$\frac{1}{2}\frac{d}{dt}\|u(x,t)\|^2_{L^2} = \langle u, u_t\rangle = -\langle u, p(u)\rangle \leq 0.$$

To prove (iii) we can use the estimate of (iv). It can be shown that the estimate is strictly negative unless $p(u) \equiv 0$. Then we use property (a) to show that the only steady state solution $u_t = -p(u) = 0$ is a constant. □

2.5 Basic properties of the variational problem

Similar results can be obtained for the variational formulation, Eq. (7).

Proposition 3 *The minimizer* u^λ *of* (7) *with* $J(u)$ *satisfying the conditions of Proposition 1 admits the following properties:*

(i) *The mean value is preserved,* $\frac{1}{|\Omega|}\int_\Omega u^\lambda(x)dx = \frac{1}{|\Omega|}\int_\Omega f(x)dx,\ \forall \lambda \geq 0$.
(ii) *The extremum principle holds,* $\min_x(f(x)) \leq u^\lambda(x) \leq \max_x(f(x)),\ \forall x \in \Omega,\ \forall \lambda \geq 0$.
(iii) *The solution converges to a constant as* $\lambda \to 0$, $\lim_{\lambda\to 0} u^\lambda(x) \equiv const = \int_\Omega f(x)dx$.
(iv) *The following estimate holds:* $\frac{1}{2}\frac{d}{d\lambda}\|f - u^\lambda\|^2_{L^2} \leq 0$.

Proof. (i) can be shown by integrating the E-L equation (8), using Property (d). One can prove (ii) by contradiction: Let us assume $\max_x(u(x)) > \max_x(f(x))$. Denoting $x_0 \in \{x : u(x) = \max_x(u(x))\}$, using Property (b) we have that $p(u)(x_0) \geq 0$, also in this case $\lambda(u(x_0) - f(x_0)) > 0$, thus the E-L equation is not satisfied. To validate (iv), we can compute the derivative of the E-L equation

(8) with respect to λ to obtain: $f - u^\lambda = (p' + \lambda)\frac{du^\lambda}{d\lambda}$, where $p'(u) = J''(u)$ is the Hessian of J, which is positive semidefinite for convex functions. Hence:

$$\frac{1}{2}\frac{d}{d\lambda}\|f - u^\lambda\|_{L^2}^2 = \langle f - u^\lambda, -\frac{du^\lambda}{d\lambda}\rangle = -\lambda\|\frac{du^\lambda}{d\lambda}\|_2^2 - \langle p'\frac{du^\lambda}{d\lambda}, \frac{du^\lambda}{d\lambda}\rangle \leq 0.$$

To prove (iii) we can use (iv) and (i) where for $\lambda = 0$ we have shown in the previous proposition that the solution of the E-L equation with no fidelity term is a constant. □

3 Fast minimization based on graph cuts

In this section we briefly describe the algorithms we have to optimize this new functional. They are defined in a discrete setting and rely on graph-cuts techniques [3, 16].

First, assume that images are defined on a discrete lattice S which is to be seen as a discretization of the continuous domain Ω. The value of the image u at a site $s \in S$ is denoted by u_s, and we assume that u_s takes value in the linearly ordered discrete set $\mathcal{L} = [\![0, \delta, \dots L - 1 - \delta, L - 1]\!]$, where $\delta > 0$ is a quantization step. The discrete version of the L^2 fidelity with any image g writes as: $\sum_s (u_s - g_s)^2$. The grid is endowed with a neighboring relationship. Let us denote by $s \sim t$ the neighborhood relationship between two adjacent sites s and t. In this paper we only need to consider pairwise interactions, and such a clique is referred to as (s, t) where $s \sim t$. Now the discrete version J_d of J given by Eq. (3) is:

$$J_d(u) = \frac{1}{2}\sum_{(s,t)} w_{st}\phi(|u_s - u_t|) \ .$$

In this paper we focus on the TV case, where $\phi(s) = s$. We show in the sequel that both variational denoising and flow computation can be obtained by minimizing a sequence of energies of the form:

$$E_d(u, g) = \frac{\lambda}{2}\sum_s (u_s - g_s)^2 + \frac{1}{2}\sum_{(s,t)} w_{st}|u_s - u_t| \ .$$

This formulation fits exactly in the framework proposed by Darbon and Sigelle in [12] to optimize TV-based energies with convex data fidelity. The algorithm is based on maximum-flow techniques. Its main characteristic is that it computes an exact minimizer (wrt. the quantization step) and is very fast. We now describe how to solve the variational denoising problem and how to compute the flow.

Minimization for Variational Denoising

We assume that the variance of the noise σ is known. For any fixed λ, let us denote by $u^\star$ a global minimizer of $E_d(\cdot, f)$. Our goal is to find the optimal value of the Lagrange multiplier $\lambda^\star$ so that we have $\sum_s (u_s^\star - f_s)^2 = |S|\sigma^2$, where $|S|$ denote the cardinal of S. Assume we have an upper bound M on λ, i.e., $0 \leq \lambda^\star \leq M$. Such a bound can be obtained using the technique described in [11]. Then the value $\lambda^\star$ can be find with the precision ϵ using the following dichotomic approach:

1. Set $\lambda_m \leftarrow 0$ and $\lambda_M = M$
2. Set $\lambda = (\lambda_m + \lambda_M)/2$ and compute $u^\star$
   - if $\sum_s (u_s^\star - f_s)^2 \leq |S|\sigma^2$ set $\lambda_M = \lambda$
   - else set $\lambda_m = \lambda$
3. if $\lambda_M - \lambda_m > \epsilon$ then go to 2, else return λ.

This algorithm performs $O(-\log_2(\epsilon))$ energy minimizations.

Flow computation

To compute the flow, we generate a series of images $\{u^n\}$ that are obtained by minimizing the following series of energies:

$$u^{n+1} = arg\min_u \frac{1}{2}\sum_s (u_s - u_s^n)^2 + \triangle t J_d(u) \ ,$$

where we start with $u^0 = f$ and $\triangle t$ is a time step. It is easily seen that the Euler-Lagrange equations of the above functionals correspond to the discrete version of the flow given by equation (6). For denoising purposes we stop the flow as soon as we have $\sum_s (u_s^n - f_s)^2 \geq |S|\sigma^2$.

4 Experimental results

Below we present some experiments for image denoising. We first describe how the weights were computed.

Computing the weights: There are many ways to choose the appropriate functional and weights. In our experiment we focused on a general denoiser for white noise based on a nonlocal TV-type energy. In this case, one may want to have the following basic guidelines for the weights: uniform high weights for smooth regions (strong denosing); low weights near random texture (reduced denosing); for edges and lines - higher weights for regions which on average are closer to the center pixel value (to reduce contrast loss).

To construct weights which approximate these guidelines we can use L^2 distances between patches, as done in [4]:

$$w(x,y) = e^{-d_a(f(x),f(y))/2h^2}, \tag{14}$$

where $d_a(\cdot,\cdot)$ is defined in (2). A semilocal neighborhood of 11×11 pixels was used as the search window around each pixel (which apparently gives better results than using larger windows or a completely nonlocal domain, see [13]). The patches that were compared were of size 5×5 pixels. The parameter h was fixed according to the noise level to $h = \sqrt{2}\sigma$ so constant regions would have approximately uniform weights with a value close to 1. A small set of the maximal weights is used, as in [13], with an average of about 10 connections per pixel.

Comments on the experiments: Fig. 1 depicts the original cameraman image and its noisy version corrupted by a white Gaussian additive noise ($\sigma = 20$), along with results for different methods: namely, original ROF model [20], NL-ROF, Eq. (7) with $\phi(s) = s$ and NL-means [4]. For each method the residual image $f - u$ is shown. In terms of SNR, NL-ROF performs better than the two other methods. For NL-ROF the result is very sharp. Both NL-ROF and NL-means preserve very well the contrast and lose very little details (of the camera), as seen clearly in the residual part. However, NL-ROF tends to produce some staircase effects (most noticeable in the face).

Fig. 2 depicts an original MRI image and its noisy version corrupted by a white Gaussian additive noise ($\sigma = 10$). Results using ROF, NL-ROF and the NL-TV-flow, Eq. (6) with $\phi(s) = s$, are shown. Both nonlocal variations are better than the original ROF model. The NL-TV-flow turns out to perform slightly better than its variational minimization with L^2 square fidelity. This behavior was also noticed for the quadratic regularizer, $\phi(s) = \frac{1}{2}s^2$, in [13]. Note that one can compute the flow using standard regularized Euler-Lagrange equation, Eq. (13). In this case we observed less staircasing than the pure TV-flow, as might be expected. The SNR in that case is 17.42 ($\epsilon = 1$).

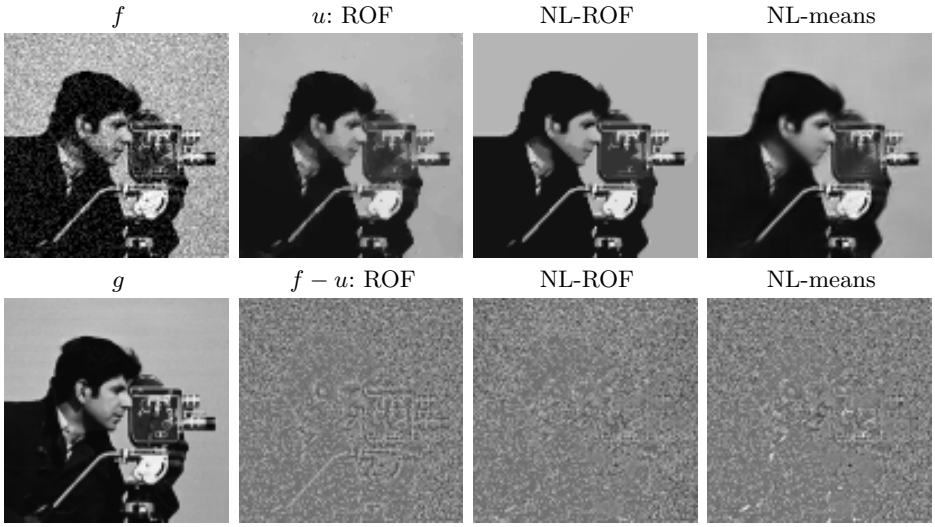

Fig. 1. Cameraman (head part), $\sigma = 20$. Top (from left): noisy image (SNR=12.53), ROF (SNR=17.12), Nonlocal ROF (SNR=18.03), NL-means (SNR=17.23). Bottom: clean image and residual parts. For all filtered images the variance of the residual equals the variance of the noise.

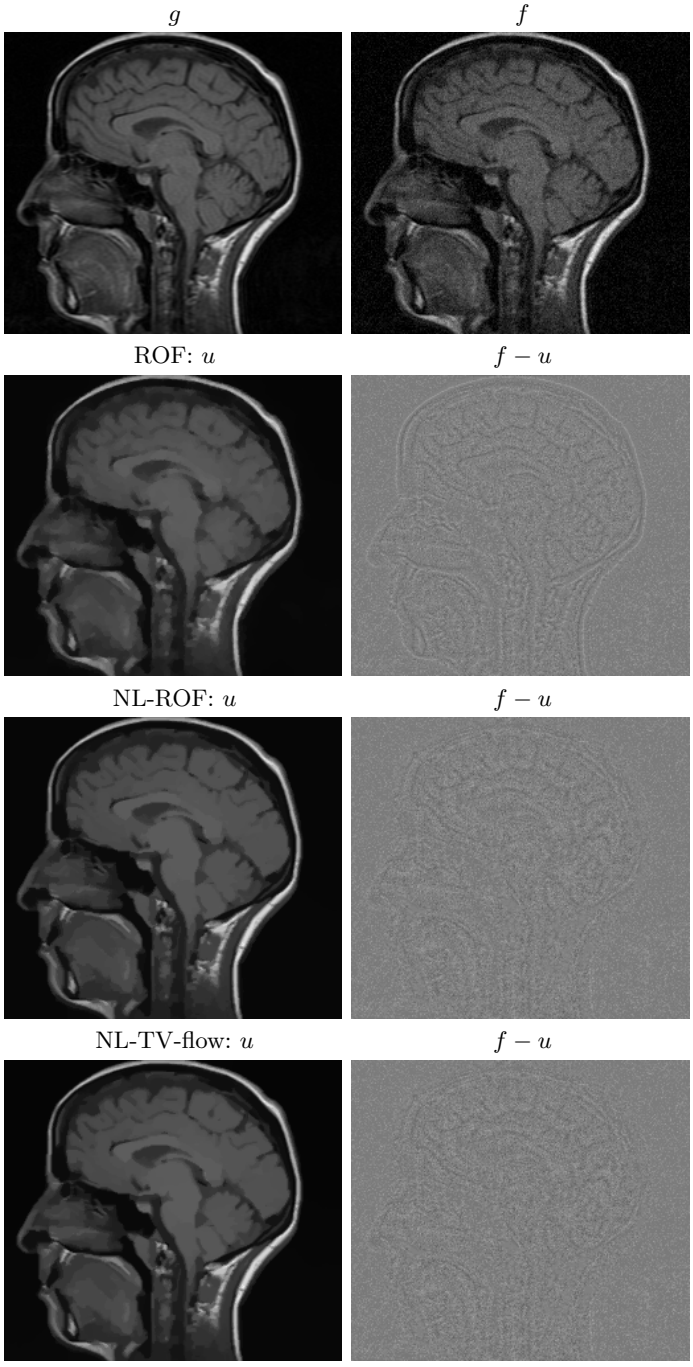

Fig. 2. MRI image, $\sigma = 10$. Top (from left): clean and noisy image (SNR=12.70). From second to fourth rows the filtered image u (left) and the residual part $f - u$ of ROF (SNR=16.16), Nonlocal ROF (SNR=16.90), Nonlocal-TV-flow (SNR=17.16), respectively. For all filtered images the variance of the residual equals the variance of the noise.

5 Conclusion

We have proposed a non local convex regularization framework along with an efficient graph-cut based algorithm for its numerical solution. In essence, we suggest to separate the problem into two parts: the first is feature extraction (computed once and represented by the weights $w(x, y)$), and the second is a more standard regularization procedure. Thus we are able to embed complex feature information into a classical functional. Very mild assumptions on the weights are required. This approach allows us to use classical convex theory and its numerical algorithms.

In this paper L^2 distances between patches have been used to compute the weights as in [4, 13]. We proposed fast numerical solutions for the nonlocal TV case (L^1-type regularization $\phi(s) = s$). In a forthcoming paper we will study alternative methods for computing the weights which take into account both: the type of images and the convex regularizer. Moreover we will present a graph-cut based numerical scheme for the general convex case.

References

1. F. Andreu, C. Ballester, V. Caselles, and J. M. Mazn. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, 2001.
2. B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf. Contour extraction based on anisotropic image classification, 2006. UCLA CAM Report 06-42.
3. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
4. A. Buades, B. Coll, and J-M. Morel. On image denoising methods. *SIAM Multiscale Modeling and Simulation*, 4(2):490–530, 2005.
5. M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Comm. in Math. Sci.*, 4(1):179–212, 2006.
6. M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.
7. M. Burger, S. Osher, J. Xu, and G. Gilboa. Nonlinear inverse scale space methods for image restoration. In *VLSM '05*, volume 3752 of *Lecture Notes in Computer Science*, pages 25–36, 2005.
8. T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.
9. F. Chung. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, 1997.
10. R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker. Geometric diffusion as a tool for harmonic analysis and structure definition of data, part i: Diffusion maps. *Proceedings of the National Academy of Sciences*, 102(21):7426–7431, 2005.
11. J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part ii: Levelable functions, convex priors and non-convex cases. *Accepted to the Journal of Mathematical Imaging and Vision*, 2005.
12. J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part i: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 2006.

13. G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation, 2006. UCLA CAM Report 06-47.
14. L. Grady and G. Funka-Lea. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. In *ECCV, Workshop on Computer Vision Approaches to Medical Image Analysis*, pages 230–245, 2004.
15. S. Kindermann, S. Osher, and P. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Multiscale Modeling and Simulation*, 4(4):1091 – 1115, 2005.
16. V. Kolmogorov and R. Zabih. What energy can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.
17. A. Marquina. Inverse scale space methods for blind dconvolution, 2006. UCLA CAM Report 06-36.
18. B. Mohar. The Laplacian spectrum of graphs. In *Y. Alavi, G. Chartrand, O. R. Oellermann, A. J. Schwenk (Eds.), Graph Theory, Combinatorics, and Applications, Wiley*, volume 2, pages 871–898, 1991.
19. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation based image restoration. *SIAM Journal on Multiscale Modeling and Simulation*, 4:460–489, 2005.
20. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
21. J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
22. A.D. Szlam. Non-stationary analysis of datasets and applications, 2006. PhD. Thesis, Yale.
23. Y. Weiss. Segmentation using eigenvectors: A unifying view. In *International Conference on Computer Vision*, pages 975–982, 1999.
24. J. Xu and S. Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising, 2006. UCLA CAM Report 06-11.
25. S.X. Yu and J. Shi. Segmentation given partial grouping constraints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):173– 183, 2004.