A Short Note for Nonlocal TV Minimization

Xavier Bresson

June 25, 2009

Abstract

This is a short note to explain how to efficiently minimize the nonlocal Total Variation (NLTV) energy. The method is based on the Split-Bregman (SB), introduced by Goldstein-Osher in [8], and extended to a nonlocal/graph version by Zhang-Burger-Bresson-Osher in [13]. For the 256×256 Barbara picture, the computation of weights takes around 1 second for a patch size 5×5 and a search window 11×11 and the NLTV minimization takes less 2 seconds. So, the total time for an image 256×256 for the NLTV minimization is less than 3 seconds. For the 512×512 Barbara picture, patch 5×5, window search 13×13, the weight computation takes 7 seconds and the NLTV minimization 8 seconds.
We also compare the SB version of NLTV with the dual version of NLTV, the NLH^1, and the NL-Means. Experiments show that the SB-NLTV provides the best denoising result.

Keywords: Image denoising, Nonlocal Total Variation (NLTV), Split-Bregman method, Comparisons with NLH^1 and NL-Means.

1 Definitions and Notations

In this section, we introduce some definitions and notations regarding nonlocal/graph (NL/G) quantities that will be useful in the rest of the paper. Many definitions are borrowed from Zhou-Scholkopf in [14] and Gilboa-Osher in [7].

The NL/G gradient of a function $u : \Omega \to \mathbb{R}$ is defined for the pair of points $(x, y) \in \Omega \times \Omega$ as:

$$\nabla_{\mathrm{NL}} u(x,y) \quad = \quad (u(y) - u(x))\sqrt{w(x,y)} : \Omega \times \Omega \to \mathbb{R}.$$

where $w : \Omega \times \Omega \to \mathbb{R}_+$ is the edge weight between the points x, y (w defines a graph). In this note, we assume that w is symmetric, i.e. $w(x,y) = w(y,x)$. The NL gradient of u does not define a vector in the standard sense since it is a mapping from $\Omega \times \Omega$ to $\mathbb{R}$. However, we will speak of NL vectors for mappings $p : \Omega \times \Omega \to \mathbb{R}$ such as $\nabla_{\mathrm{NL}} u$. The inner product between two NL vectors $p_1, p_2 : \Omega \times \Omega \to \mathbb{R}$ at point $x \in \Omega$ is defined as:

$$< p_1, p_2 > (x) \quad = \quad \int_\Omega p_1(x,y) p_2(x,y) dy : \Omega \to \mathbb{R},$$

which gives the norm of a NL vector $p : \Omega \times \Omega \to \mathbb{R}$ at point $x \in \Omega$ as follows:

$$|p|(x) = \sqrt{\int_\Omega p(x,y)^2 dy} : \Omega \to \mathbb{R}_+.$$

Hence, the norm of the NL gradient of a function $u : \Omega \to \mathbb{R}$ at $x \in \Omega$ is defined as:

$$|\nabla_{\rm NL} u|(x) = \sqrt{\int_\Omega (u(y) - u(x))^2 w(x,y) dy} : \Omega \to \mathbb{R}_+.$$

The NL divergence operator can be defined by the standard adjoint relation with the NL gradient (divergence theorem extended to graph):

$$< \nabla_{\rm NL} u, p >= - < u, \mathrm{div}_{\rm NL}\, p >, \ \forall u : \Omega \to \mathbb{R}, \ \forall p : \Omega \times \Omega \to \mathbb{R},$$

which defines the NL divergence of the NL vector $p : \Omega \times \Omega \to \mathbb{R}$ at $x \in \Omega$:

$$\mathrm{div}_{\rm NL}\, p(x) = \int_\Omega (p(x,y) - p(y,x)) \sqrt{w(x,y)} dy : \Omega \to \mathbb{R},$$

2 Nonlocal Means

In this section, we recall the recent image denoising model of Buades-Coll-Morel introduced in [2] and called Nonlocal Means (NLM). The NLM algorithm can be seen as the generalization of the Yaroslavsky filter [12] and bilateral filters [11] to intensity patch feature instead of single pixel feature. See also the paper on Texture Synthesis [5]. The objective of NLM is to denoise the noisy image using patches around pixels and looking for other similar patches in the image. Then, a weighted intensity average of pixels with similar patches produces the new denoised intensity value. The model of NLM is a linear filter-based model, whose filter is non-linearly computed with the distance between patches. The continuous version of the NLM is as follows:

$$u(x) \quad = \quad \frac{\int_\Omega w(x,y) u_0(y) dy}{\int_\Omega w(x,y) dy}.$$

where $u_0 : \Omega \to \mathbb{R}$ is the given noisy image, $u : \Omega \to \mathbb{R}$ is the outcome of the NLM algorithm, function w is defined as

$$w(x,y) \quad = \quad e^{-\int_\Omega G_a(z)|u_0(x+z) - u_0(y+z)|^2 dz / h^2},$$

where $\int_\Omega G_a(z)|u_0(x+z) - u_0(y+z)|^2 dz$ is the distance between patches located at x and y, G_a is a Gaussian function with standard deviation a and h is a positive constant which acts as a scale parameter.

If we denote i, j for discretized points $x, y \in \Omega \times \Omega$, the discrete NLM is as follows:

$$u_i \quad = \quad \frac{\Sigma_j w_{ij} u_j}{\Sigma_j w_{ij}}.$$

3 Nonlocal H^1

In this section, we recall the nonlocal image denoising model of Gilboa-Osher in [6]. Based on [9], Gilboa-Osher embed the NLM model into a variational formulation. They also propose denoising images using the NLH^1 operator (defined as $\int_\Omega |\nabla_{\rm NL} u|^2$):

$$\begin{aligned} \inf_u F(u) \quad &= \quad \int_\Omega |\nabla_{\rm NL} u|^2 + \lambda (u - u_0)^2 dx, \qquad (3.1) \\ &= \quad \int_{\Omega \times \Omega} w(x,y)(u(y) - u(x))^2 + \lambda (u - u_0)^2 dy dx \end{aligned}$$

where $u, u_0 : \Omega \to \mathbb{R}$ and $\nabla_{\mathrm{NL}} u : \Omega \times \Omega \to \mathbb{R}$, u_0 is the given noisy image and λ is a positive constant that controls the trade-off between the regularization process and the fidelity with respect to the original image.

3.1 Steepest Descent Minimization Scheme

The steepest gradient descent scheme based on the Euler-Lagrange equation of (3.1) is given by:

$$\begin{aligned} \partial_t u(x,t) &= \Delta_{\mathrm{NL}} u - 2\lambda(u - u_0) \\ &= 2\int_\Omega (u(y) - u(x))w(x,y)dy - 2\lambda(u(x) - u_0(x)), \end{aligned} \tag{3.2}$$

where $\Delta_{\mathrm{NL}} := \mathrm{div}_{\mathrm{NL}} \nabla_{\mathrm{NL}}$ denotes the NL/graph Laplacian. (3.2) is a linear heat diffusion equation carried out on a graph defined by the weights w. The discrete iterative minimization scheme at pixel i and at iteration k is as follows:

$$u_i^{k+1} = u_i^k + \delta t\Big(\Sigma_j w_{ij}(u_j^k - u_i^k) + \lambda(u_i^k - {u_0}_i^k)\Big),\ k \geq 0,\ u^0 = u_0, \tag{3.3}$$

where $\delta t > 0$ is the temporal step. The minimization scheme (3.3) is slow. A faster numerical scheme is introduced in the next section.

3.2 Fixed Point Minimization Scheme

In this section, a fixed point approach is used to minimize (3.1). The Euler-Lagrange equation of (3.1) in a discrete formulation is as follows:

$$\Sigma_j w_{ij}(u_j - u_i) + \lambda(u_i - u_{0i}) = 0,$$

which solution is given by

$$u_i = \frac{\lambda u_{0i} + \Sigma_j w_{ij} u_j}{\lambda + \Sigma_j w_{ij}}.$$

The following iterative minimization scheme:

$$u_i^{k+1} = \frac{\lambda u_{0i} + \Sigma_j w_{ij} u_j^k}{\lambda + \Sigma_j w_{ij}},$$

converges much faster to the minimizer and does not depend on a time step like in the previous section.

4 Nonlocal TV

In this section, we recall the nonlocal image denoising model of Gilboa-Osher in [7]. Gilboa-Osher proposed using the NLTV operator (defined as $\int_\Omega |\nabla_{\mathrm{NL}} u|$) as follows:

$$\begin{aligned} \inf_u F(u) &= \int_\Omega |\nabla_{\mathrm{NL}} u| + \frac{\lambda}{2}(u - u_0)^2 dx, \\ &= \int_\Omega \sqrt{\int_\Omega (u(y) - u(x))^2 w(x,y)dy} + \frac{\lambda}{2}(u - u_0)^2 dx \end{aligned} \tag{4.1}$$

where $u, u_0 : \Omega \to \mathbb{R}$, $\nabla_{\mathrm{NL}} u : \Omega \times \Omega \to \mathbb{R}$ and u_0 is the given noisy image.

4.1 Steepest Descent Minimization Scheme

The steepest gradient descent scheme based on the Euler-Lagrange equation of (4.1) is:

$$\begin{aligned}\partial_t u(x,t) &= \kappa_{\mathrm{NL}}(u) - \lambda(u-u_0)\\ &= \int_\Omega w(x,y)(u(y)-u(x))(|\nabla_{\mathrm{NL}} u|^{-1}(x) + |\nabla_{\mathrm{NL}} u|^{-1}(y))dy\\ &\quad -\lambda(u(x)-u_0(x)),\end{aligned}$$

where $\kappa_{\mathrm{NL}}(.) := \mathrm{div}_{\mathrm{NL}} \frac{\nabla_{\mathrm{NL}}\cdot}{|\nabla_{\mathrm{NL}}\cdot|}$ denotes the NL/graph curvature. The previous minimizing flow is slow. Gilboa-Osher in [7] extend the TV minimization algorithm of Chambolle [3] to the NLTV.

4.2 Gilboa-Osher's Model: Dual Version of NLTV Minimization

Gilboa-Osher extend the Chambolle's projection algorithm introduced in [3] to minimize the NLTV. As shown by Chan-Golub-Mulet in [4], the minimization of (4.1) is equivalent to this min-max problem:

$$\inf_u \sup_{|p|\leq 1} \int_\Omega < \nabla_{\mathrm{NL}} u, p > + \frac{\lambda}{2}(u-u_0)^2 dx,$$

where $u, u_0 : \Omega \to \mathbb{R}$ and $\nabla_{\mathrm{NL}} u, p : \Omega \times \Omega \to \mathbb{R}$ are NL vectors. The inf and sup can be swapped according to the minimax theorem, which gives the explicit minimizing solution for u:

$$u = u_0 - \mathrm{div}_{\mathrm{NL}}\, p_\star/\lambda,$$

and $p_\star$ is given by solving the following max problem for p:

$$\sup_{|p|\leq 1} \int_\Omega < \nabla_{\mathrm{NL}} u, p > + \frac{1}{2\lambda}|\,\mathrm{div}_{\mathrm{NL}}\, p|^2 dx$$

which is solved with the following semi-implicit fixed point iterated method:

$$p^{k+1}(x,y) = \frac{p^k + \delta t \nabla_{\mathrm{NL}}(\mathrm{div}_{\mathrm{NL}}\, p^k - \lambda u_0)}{1 + \delta t |\nabla_{\mathrm{NL}}(\mathrm{div}_{\mathrm{NL}}\, p^k - \lambda u_0)|},\ k > 0,\ p^0 = 0, \tag{4.2}$$

where $\delta t < 1/\|\,\mathrm{div}_{\mathrm{NL}}\,\|$ is the time step that guaranties the convergence of the iterative scheme (4.2).

The discrete formulation of (4.2) can be written as follows at discrete points i, j at iteration k:

$$p_{ij}^{k+1} = \frac{p_{ij}^k + \delta t\big(\sum_i w_{ij}(p_{ji}^k - p_{ij}^k) - \sum_j w_{ij}(p_{ij}^k - p_{ji}^k) - \lambda(u_{0j} - u_{0i})\sqrt{w_{ij}}\big)}{1 + \delta t \sqrt{\sum_j \big(\sum_i (p_{ji}^k - p_{ij}^k) w_{ij} - \sum_j (p_{ij}^k - p_{ji}^k) w_{ij} - \lambda(u_{0j} - u_{0i})\sqrt{w_{ij}}\big)^2}}.$$

The previous NLTV minimizing scheme is much faster than the direct steepest descent scheme. However, we will introduce in the next section a NLTV minimization scheme which is faster than the dual version of the NLTV algorithm and presents better denoising results.

4.3 Zhang-Burger-Bresson-Osher's Model: Split-Bregman version of NLTV Minimization

In this section, we develop the Split-Bregman version [8] of the NLTV minimization. The minimization problem (4.1) is as follows:

$$\inf_u F(u) = \int_\Omega |\nabla_{\mathrm{NL}} u| + \frac{\lambda}{2}(u-u_0)^2 dx,$$

where $u, u_0 : \Omega \to \mathbb{R}$ and $\nabla_{\text{NL}} u : \Omega \times \Omega \to \mathbb{R}$. A new variable $d : \Omega \times \Omega \to \mathbb{R}$ is introduced as follows:

$$\inf_{u,d} \int_\Omega |d| + \frac{\lambda}{2}(u - u_0)^2, \quad \text{such that } d = \nabla_{\text{NL}} u$$

The constraint $d = \nabla u$ is enforced using the efficient Bregman iteration approach [8, 10, 1] defined as:

$$\left\{ \begin{array}{rcl} (u^{k+1}, d^{k+1}) & = & \arg\min\limits_{u,d} \int_\Omega |d| + \frac{\lambda}{2}(u - u_0)^2 + \frac{\beta}{2}|d - \nabla_{\text{NL}} u - b^k|^2 dx \\ b^{k+1} & = & b^k + d^{k+1} - \nabla_{\text{NL}} u^{k+1}. \end{array} \right. , \; k \geq 0$$

The minimizing solution $u^{k+1} : \Omega \to \mathbb{R}$ is characterized by the optimality condition:

$$\lambda(u - u_0) - \beta \, \text{div}_{\text{NL}}(d^k - \nabla_{\text{NL}} u - b^k) = 0,$$

A fast approximated solution is provided by a Gauss-Seidel iterative scheme, i.e. for $n \geq 0$:

$$\begin{aligned} u_i^{k+1,n+1} = \quad & \frac{1}{\lambda + \beta \sum_j w_{ij}} \Big(\beta \sum_j w_{ij} u_j^{k+1,n} + \\ & \lambda u_{0i} - \beta \sum_j \sqrt{w_{ij}} (d_{ij}^{k+1,n} - d_{ji}^{k+1,n} - b_{ij}^{k+1,n} + b_{ji}^{k+1,n}) \Big), \; u^{k+1,n=0} = u^k, \end{aligned} \tag{4.3}$$

In experiments, we observed that $n = 2$ iterations of (4.3) are enough to determine a good approximation of the minimizer.

The minimizing solution $d^{k+1} : \Omega \times \Omega \to \mathbb{R}$ is given by extension of the soft-thresholding formula to the nonlocal/graph case as follows:

$$d^{k+1} = \frac{\nabla_{\text{NL}} u^{k+1} + b^k}{|\nabla_{\text{NL}} u^{k+1} + b^k|} \max(|\nabla_{\text{NL}} u^{k+1} + b^k| - \beta, 0).$$

which discrete solution at pixel i at iteration $k+1$ is as follows:

$$\begin{aligned} d_{ij}^{k+1} = \quad & \frac{\sqrt{w_{ij}}(u_j^{k+1} - u_i^{k+1}) + b_{ij}^k}{\sqrt{\sum_j w_{ij}(u_j^{k+1} - u_i^{k+1})^2 + {b_{ij}^k}^2}} \\ & \max\Big(\sqrt{\sum_j w_{ij}(u_j^{k+1} - u_i^{k+1})^2 + {b_{ij}^k}^2} - \beta, 0 \Big). \end{aligned}$$

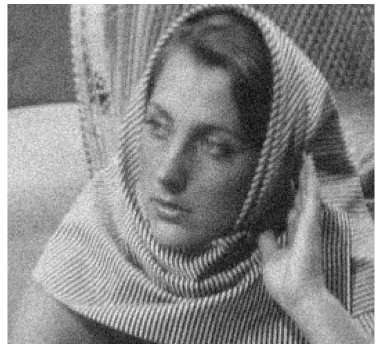
Finally, the Bregman variable is updated as follows:

$$b_{ij}^{k+1} = b_{ij}^k + \sqrt{w_{ij}}(u_j^{k+1} - u_i^{k+1}) - d_{ij}^{k+1}.$$

5 Comparisons between nonlocal Denoising Models

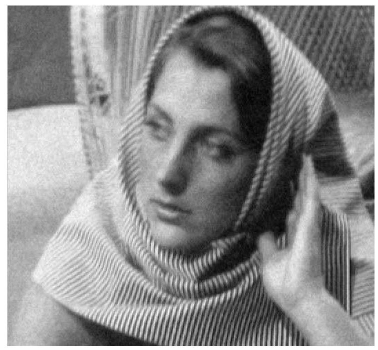
Given the same weight function w for all nonlocal image denoising models, we will compare the quality and speed of these models. We use the Barbara picture, which contains smooth and textured regions. Standard schemes s.a. standard TV minimization scheme do not produce as good denoising results for texture denoising as the nonlocal versions.

(a) Original.



(b) Noisy Image.

(c) nonlocal Means. SNR=17.3. Time=0.03sec.



(d) nonlocal H^1. SNR=17.07. Time=0.16sec.

(e) nonlocal TV (Projection). SNR=17.62. Time=1.91sec.

(f) nonlocal TV (Split-Bregman). SNR=17.87. Time=1.51sec.

Figure 1: nonlocal denoising results.

References

[1] L. Bregman. The Relaxation Method of Finding the Common Points of Convex Sets and Its Application to the Solution of Problems in Convex Programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[2] A. Buades, B. Coll, and J.M Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling and Simulation (MMS)*, 4(2):490–530, 2005.

[3] A. Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[4] T.F. Chan, G.H. Golub, and P. Mulet. A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration. *SIAM journal on scientific computing*, 20(6):1964–1977, 1999.

[5] A. Efros and T. Leung. Texture Synthesis by Non-Parametric Sampling. In *International Conference on Computer Vision*, pages 1033–1038, 1999.

[6] G. Gilboa and S. Osher. Nonlocal Linear Image Regularization and Supervised Segmentation. *SIAM Multiscale Modeling and Simulation (MMS)*, 6(2):595–630, 2007.

[7] G. Gilboa and S. Osher. Nonlocal Operators with Applications to Image Processing. *SIAM Multiscale Modeling and Simulation (MMS)*, 7(3):1005–1028, 2007.

[8] T. Goldstein and S. Osher. The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[9] S. Kindermann, S. Osher, and P. Jones. Deblurring and Denoising of Images by Nonlocal Functionals. *SIAM Multiscale Modeling and Simulation (MMS)*, 4(4):1091–1115, 2005.

[10] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An Iterative Regularization Method for Total Variation-based Image Restoration. *SIAM Multiscale Modeling and Simulation*, 4:460–489, 2005.

[11] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. In *International Conference on Computer Vision (ICCV)*, pages 839–846, 1998.

[12] L.P. Yaroslavsky. *Digital Picture Processing, an Introduction.* Springer-Verlag, 1985.

[13] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction. *CAM Report 09-03*, 2009.

[14] D. Zhou and B. Scholkopf. A Regularization Framework for Learning from Graph Data. In *Workshop on Statistical Relational Learning and Its Connections to Other Fields*, 2004.