Course 049064: Variational Methods in Image Processing

Notes 10: Nonlocal Differential Framework

Guy Gilboa

1 Background

One can often better process image pixels by exploiting non-local correlations, in addition to the classical local correlations of adjacent pixels. Most notably this has been shown by Efros and Leung for texture synthesis [8] and by Buades-Coll-Morel for denoising [4].

A first suggestion of a nonlocal functional was proposed by Kindermann et al in [12]. However, no nonlocal operators were suggested and it was very hard to generalized these formulations and to relate them to local functionals.

In [9] and [10] Gilboa and Osher proposed a complete nonlocal framework in the continuous setting, derived the basic nonlocal differential operators and related them to the local setting and to spectral graph theory [7, 3].

More advances on this subject can be seen in [16, 14, 17, 11]. In these notes we will focus on the paper of [10].

1.1 Nonlocal Means

In [4] Buades-Coll-Morel suggested the following nonlocal filter for image denoising:

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

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:

$$c(x) = \int_\Omega e^{-d_a(u(x),u(y))/h^2} dy. \tag{3}$$

The corresponding discrete formulation is:

$$NL(u)(i) = \sum_j \alpha(i,j)u(j)$$

where

$$\alpha(i,j) = \frac{1}{c(i)} e^{-\|u(B_i)-v(B_j)\|^2_{2,a}/h^2}$$

$u(B_i) = (u(k) : \ k \in B_i)$, B_i is a small ball (patch, in general) around pixel i.

1.2 Graph Laplacian

We will see below that the operators can be related to similar ones on graphs, and specifically to the well know graph Laplacian.

Let $G = (V, E)$ be a connected undirected weighted graph with (a finite set of) vertices (nodes) V and edges E. To each edge $e_{kl} \in E$ between nodes k and l a corresponding weight $w_{kl} \in E$ is defined. The weights are non-negative and symmetric: $w_{kl} \geq 0$, $w_{kl} = w_{lk}$. We assume that a discrete function u is defined on the nodes of the graph and denote by $u(k) \in V$ the value of u at node k. The (weighted) *graph Laplacian* is

$$\Delta_G(u(k)) := \sum_{l \in \mathcal{N}_k} w_{kl}(u(l) - u(k)), \quad k, l \in V, \tag{4}$$

where $l \in \mathcal{N}_k$ is the set of nodes with edges connected to k. Note that we define here the Laplacian with an *opposite sign* to the usual graph theoretic definition so it will coincide with the continuous definition.

2 A nonlocal mathematical framework

2.1 Basic operators

Let $\Omega \subset \mathbb{R}^n$, $x \in \Omega$, $u(x)$ a real function $u : \Omega \to \mathbb{R}$. We extend the notion of derivatives to a nonlocal framework by the following definition:

$$\partial_y u(x) := \frac{u(y) - u(x)}{\tilde{d}(x,y)}, \quad y, x \in \Omega,$$

where $0 < \tilde{d}(x,y) \leq \infty$ is a positive measure defined between points x and y. To keep with standard notations related to graphs we define the weights as

$$w(x,y) = \tilde{d}^{-2}(x,y).$$

Thus $0 \leq w(x,y) < \infty$. The weights are symmetric, that is $w(x,y) = w(y,x)$. The nonlocal derivative can be written as

$$\partial_y u(x) := (u(y) - u(x))\sqrt{w(x,y)}. \tag{5}$$

The nonlocal gradient $\nabla_w u(x) : \Omega \to \Omega \times \Omega$ is defined as the vector of all partial derivatives:

$$(\nabla_w u)(x,y) := (u(y) - u(x))\sqrt{w(x,y)}, \quad x, y \in \Omega. \tag{6}$$

Vectors are denoted as $\vec{v} = v(x,y) \in \Omega \times \Omega$. The standard L^2 inner product is used for functions

$$\langle u_1, u_2 \rangle := \int_\Omega u_1(x) u_2(x) dx.$$

For vectors we define a dot product

$$(\vec{v_1} \cdot \vec{v_2})(x) := \int_\Omega v_1(x,y) v_2(x,y) dy,$$

and an inner product

$$\langle \vec{v}_1, \vec{v}_2 \rangle := \langle \vec{v}_1 \cdot \vec{v}_2, 1 \rangle = \int_{\Omega \times \Omega} v_1(x,y) v_2(x,y) dx dy.$$

The magnitude of a vector is

$$|\vec{v}|(x) := \sqrt{\vec{v_1} \cdot \vec{v_1}} = \sqrt{\int_\Omega v(x,y)^2 dy}.$$

With the above inner products the nonlocal divergence $\text{div}_w \, \vec{v}(x) : \Omega \times \Omega \to \Omega$ is defined as the adjoint of the nonlocal gradient:

$$(\text{div}_w \, \vec{v})(x) := \int_\Omega (v(x,y) - v(y,x))\sqrt{w(x,y)} dy. \tag{7}$$

The Laplacian can now be defined by:

$$\Delta_w u(x) := \frac{1}{2} \text{div}_w (\nabla_w u(x)) = \int_\Omega (u(y) - u(x)) w(x,y) dy. \tag{8}$$

Note that in order to get the standard Laplacian definition which relates to the graph Laplacian we need a factor of $1/2$.

2.2 Some properties

Most of the properties involving a double integral can be shown by expanding an integral of the form $\int_{\Omega\times\Omega} f(x,y)dxdy$ to $\frac{1}{2}\int_{\Omega\times\Omega}(f(x,y)+f(y,x))dxdy$, changing the order of integration and using the fact that $w(x,y)=w(y,x)$. We give an example showing the adjoint relation

$$\langle \nabla_w u, \vec{v}\rangle = \langle u, -\operatorname{div}_w \vec{v}\rangle, \tag{9}$$

$$\begin{aligned}
\langle \nabla_w u, \vec{v}\rangle \quad &= \int_{\Omega\times\Omega}(u(y)-u(x))\sqrt{w(x,y)}v(x,y)dxdy \\
&= \tfrac{1}{2}\int_{\Omega\times\Omega}\left[(u(y)-u(x))\sqrt{w(x,y)}v(x,y)+(u(x)-u(y))\sqrt{w(y,x)}v(y,x)\right]dxdy \\
&= \tfrac{1}{2}\int_{\Omega\times\Omega}\left[u(y)(v(x,y)-v(y,x))-u(x)(v(x,y)-v(y,x))\right]\sqrt{w(x,y)}dxdy \\
&= \tfrac{1}{2}\int_{\Omega\times\Omega}\left[u(x)(v(y,x)-v(x,y))-u(x)(v(x,y)-v(y,x))\right]\sqrt{w(x,y)}dxdy \\
&= \int_\Omega u(x)\left(-\int_\Omega(v(x,y)-v(y,x))\sqrt{w(x,y)}dy\right)dx.
\end{aligned}$$

"Divergence theorem":

$$\int_\Omega \operatorname{div}_w \vec{v}dx = 0. \tag{10}$$

The Laplacian is self adjoint

$$\langle \Delta_w u, u\rangle = \langle u, \Delta_w u\rangle \tag{11}$$

and negative semidefinite

$$\langle \Delta_w u, u\rangle = -\langle \nabla_w u, \nabla_w u\rangle \le 0. \tag{12}$$

We can also formulate a nonlocal (mean) curvature:

$$\begin{aligned}
\kappa_w &:= \operatorname{div}_w\left(\frac{\nabla_w u}{|\nabla_w u|}\right) \\
&= \int_\Omega (u(y)-u(x))w(x,y)\left(\frac{1}{|\nabla_w u|(x)}+\frac{1}{|\nabla_w u|(y)}\right)dy,
\end{aligned} \tag{13}$$

where

$$|\nabla_w u|(q) := \sqrt{\int_\Omega (u(z)-u(q))^2 w(q,z)dz}.$$

2.3 The Regularizing Functionals

Two types of regularizing nonlocal functionals are proposed. The first type is based on the nonlocal gradient. It is set within the mathematical framework described above. The second type is based on differences, it appears to be easier to implement, where the minimization can be accomplished using graph cut techniques,

The gradient-based functional is

$$\begin{aligned} J(u) &= \int_\Omega \phi(|\nabla_w u|^2)dx, \\ &= \int_\Omega \phi(\int_\Omega (u(y)-u(x))^2 w(x,y)dy)dx, \end{aligned} \tag{14}$$

where $\phi(s)$ is a positive function, convex in $\sqrt{s}$ with $\phi(0) = 0$.

The difference-based functional is

$$J_a(u) = \int_{\Omega\times\Omega} \phi((u(y)-u(x))^2 w(x,y))dydx. \tag{15}$$

The variation with respect to u (Euler-Lagrange) of (14) is

$$\partial_u J(u) = -2\int_\Omega (u(y)-u(x))w(x,y)(\phi'(|\nabla_w u|^2(x))+\phi'(|\nabla_w u|^2(y)))dy, \tag{16}$$

where $\phi'(s)$ is the derivative of ϕ with respect to s. This can be written more concisely as

$$\partial_u J(u) = -2\,\mathrm{div}_w\left(\nabla_w u \phi'(|\nabla_w u|^2(x))\right).$$

The variation with respect to u of (15) is

$$\partial_u J_a(u) = -4\int_\Omega (u(y)-u(x))w(x,y)\phi'((u(y)-u(x))^2 w(x,y))dy. \tag{17}$$

Note that for the quadratic case $\phi(s) = s$ the functionals (14) and (15) coincide (and naturally so do Eqs. (16) and (17)).

2.3.1 Relation to isotropic and anisotropic local functionals

The functionals which can be written in the form of Eq. (14) correspond in the local case to *isotropic* functionals (which have no preferred directionality). The second category, Eq. (15), can be related to *anisotropic* functionals in the local case. We suggest later two different methods for efficiently computing each category.

As an example, for total-variation, $\phi(s) = \sqrt{s}$, Eq. (14) becomes:

$$J_{NL-TV}(u) = \int_\Omega |\nabla_w u|dx = \int_\Omega \sqrt{\int_\Omega (u(y)-u(x))^2 w(x,y)dy}dx \tag{18}$$

whereas Eq. (15) becomes

$$J_{NL-TVa}(u) = \int_{\Omega\times\Omega} |u(x)-u(y)|\sqrt{w(x,y)}dydx \tag{19}$$

The above functionals correspond in the local two dimensional case to the isotropic TV

$$J_{TV}(u) = \int_{\Omega} |\nabla u| dx = \int_{\Omega} \sqrt{u_{x_1}^2 + u_{x_2}^2} dx$$

and to the anisotropic TV

$$J_{TVa}(u) = \int_{\Omega} (|u_{x_1}| + |u_{x_2}|) dx.$$

3 Basic Models

Classical PDE's can have their nonlocal generalization:

Nonlocal diffusion:

$$u_t(x) = \Delta_w u(x), \quad u_{t=0} = f(x), \tag{20}$$

where $\Delta_w u(x)$ is the nonlocal Laplacian defined in 8.

Nonlocal TV-flow:

$$u_t(x) = -\partial J_{NL-TV}(u(x)), \quad u_{t=0} = f(x), \tag{21}$$

where $J_{NL-TV}(u)$ is defined in (18).

One can also define regularization models, based for instance on the NL-TV.

Nonlocal ROF:

$$J_{NL-TV}(u) + \lambda \|f-u\|_{L^2}^2. \tag{22}$$

Nonlocal TV-L1: Another very important model following [13] is the extension of $TV - L^1$ to a nonlocal version:

$$J_{NL-TV}(u) + \lambda \|f-u\|_{L^1}. \tag{23}$$

An interesting application of texture regularization can be shown using Tv-L1. It can both detect and remove anomalies or irregularities from images, and specifically textures.

Nonlocal TV-G:

$$J_{NL-TV}(u) + \lambda \|f-u\|_{NL-G}. \tag{24}$$

where the G-norm is defined below.

Inpainting:

Follows the local TV-inpainting model of [6]:

$$J_{NL-TV}(u) + \int_\Omega \lambda(x)(f-u)^2 dx, \tag{25}$$

with $\lambda(x) = 0$ in the inpainting region and $\lambda(x) = c$ in the rest of the image.

4 Discrete Computations

4.1 Basic Discretization

Let u_i denote the value of a pixel i in the image ($1 \leq i \leq N$), $w_{i,j}$ is the sparsely discrete version of $w(x,y)$. We use the neighbors set notation $j \in \mathcal{N}_i$ defined as $j \in \mathcal{N}_i := \{j : w_{i,j} > 0\}$.

Let ∇_{wd} be the discretization of ∇_w:

$$\nabla_{wd}(u_i) := (u_j - u_i)\sqrt{w_{i,j}}, \quad j \in \mathcal{N}_i \tag{26}$$

Let div_{wd} be the discretization of div_w:

$$\text{div}_{wd}(p_{i,j}) := \sum_{j \in \mathcal{N}_i} (p_{i,j} - p_{j,i})\sqrt{w_{i,j}}. \tag{27}$$

The discrete inner product for functions is $< u, v >:= \sum_i (u_i v_i)$ and for vectors we have the discretized dot product $(p \cdot q)_i := \sum_j (p_{i,j} q_{i,j})$ and inner product $< p, q >:= \sum_i \sum_j (p_{i,j} q_{i,j})$. The vector magnitude is therefore $|p|_i := \sqrt{\sum_j (p_{i,j})^2}$.

weights discretization: The weights are descretized as follows: we take a patch around a pixel i, compute the distances $(d_a)_{i,j}$, a discretization of $d_a(x,y)$, Eq. (2), to all the patches in the search window and select the k closest (with the lowest distance value). The number of neighbors k is an integer proportional to the area γ. For each selected neighbor j we assign the value 1 to $w_{i,j}$ and to $w_{j,i}$. A maximum of up to $m = 2k$ neighbors for each pixel is allowed in our implementation. A reasonable setting is to take 5×5 pixel patches, a search window of size 21×21 and $m = 10$.

4.2 Steepest Descent

In this convex framework, one can resort as usual to a steepest descent method for computing the solutions. One initializes u at $t = 0$, e.g. with the input image: $u|_{t=0} = f$, and evolves numerically the flow:

$$u_t = -\partial_u J_d - \partial_u H_d(f, u),$$

where $\partial_u J_d$ is the discretized version of Eq. (16) or Eq. (17) and $H_d(f, u)$ is the discretized fidelity term functional. As in the local case, here also one should introduce a regularized version of the total variation: $\phi(s) = \sqrt{s + \epsilon^2}$ (where s is the square gradient magnitude). Thus the E-L equations are well defined, also for a zero gradient.

4.2.1 Nonlocal ROF

Chambolle's projection algorithm [5] for solving ROF [15] can be extended to solve nonlocal ROF.

A minimizer for the discrete version of Eq. (22) can be computed by the following iterations (fixed point method):

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla_{wd}(\mathrm{div}_{wd}(p^n) - 2\lambda f))_{i,j}}{1 + \tau|(\nabla_{wd}(\mathrm{div}_{wd}(p^n) - 2\lambda f))_{i,j}|} \tag{28}$$

where $p^0 = 0$, and the operators ∇_{wd} and div_{wd} are defined in (26) and (27), repectively. The solution is $u = f - \frac{1}{2\lambda}\,\mathrm{div}_{wd}(p)$.

Theorem: The algorithm converges to the global minimizer as $n \to \infty$ for any $0 < \tau \leq \frac{1}{\|\,\mathrm{div}_{wd}\,\|_{L^2}^2}$.

For the proof see [10].

A bound on τ The bound on τ depends on the operator norm $\|\,\mathrm{div}_{wd}\,\|^2$ which is a function of the weights $w_{i,j}$. As the weights are image dependent, so is $\|\,\mathrm{div}_{wd}\,\|^2$.

Let m be the maximal number of neighbors of a pixel, $m := \max_i\{\sum_j(\mathrm{sign}(w_{i,j}))\}$. If the weights are in the range $0 \leq w_{i,j} \leq 1\ \forall i, j$, then for $0 < \tau \leq \frac{1}{4m}$ the algorithm converges.

We need to show that $\|\,\mathrm{div}_{wd}\,\|^2 \leq 4m$:

$$\begin{aligned}
\|\,\mathrm{div}_{wd}(p)\|^2 &= \textstyle\sum_i \left(\sum_j (p_{i,j} - p_{j,i})\sqrt{w_{i,j}}\right)^2 \\
&\leq 2\textstyle\sum_i \left(\sum_j (p_{i,j}^2 + p_{j,i}^2)\right)\left(\sum_j w_{i,j}\right) \\
&\leq 4\max_i \left(\textstyle\sum_j w_{i,j}\right) \sum_i \sum_j p_{i,j}^2 \\
&\leq 4m\|p\|^2.
\end{aligned}$$

4.2.2 Nonlocal TV-L1

To solve (23) we generalize the algorithm of [2]. We consider the problem:

$$\inf_{u,v} \left\{ J_{NL-TV}(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + \lambda\|v\|_{L^1} \right\} \tag{29}$$

The parameter α is small so that we almost have $f = u + v$, thus (29) is a very good approximation of (23). We can solve the discretized version of (29) by iterating:

- v being fixed (we have a nonlocal ROF problem), find u using the nonlocal Chambolle's projection algorithm:

$$\inf_{u} \left(J_{NL-TV}(u) + \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 \right)$$

- u being fixed, find v which satisfies:

$$\inf_{v} \frac{1}{2\alpha}\|f-u-v\|_{L^2}^2 + \lambda\|v\|_{L^1}.$$

The solution for v is given by soft-thresholding $f-u$ with $\alpha\lambda$ as the threshold, denoted by $ST_{\alpha\lambda}(f-u)$, where

$$ST_\beta(q) := \begin{cases} q-\beta, & q > \beta \\ 0, & |q| \leq \beta \\ q+\beta, & q < -\beta. \end{cases} \tag{30}$$

The algorithm converges to the global minimizer as $n \to \infty$ for any $0 < \tau \leq \frac{1}{\|\operatorname{div}_{wd}\|_{L^2}^2}$.

A similar extension to the iterative projection algorithm of [1] can be made also to the nonlocal setting.

References

[1] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *JMIV*, 22(1), January 2005.

[2] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition – modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

[3] S. Bougleux, A. Elmoataz, and M. Melkemi. Discrete regularization on weighted graphs for image and mesh filtering. In *1st International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, volume 4485 of *Lecture Notes in Computer Science*, pages 128–139, 2007.

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

[5] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20:89–97, 2004.

[6] T.F. Chan and J. Shen. Mathematical models of local non-texture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2001.

[7] F. Chung. *Spectral Graph Theory.* Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, 1997.

[8] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *ICCV (2)*, pages 1033–1038, 1999.

[9] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Multiscale Modeling and Simulation*, 6(2):595–630, 2007.

[10] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *SIAM Multiscale Modeling and Simulation*, 7(3):1005–1028, 2008.

[11] Varun P Gopi, P Palanisamy, Khan A Wahid, and Paul Babyn. Mr image reconstruction based on framelets and nonlocal total variation using split bregman method. *International journal of computer assisted radiology and surgery*, 9(3):459–472, 2014.

[12] S. Kindermann, S. Osher, and P. Jones. Deblurring and denoising of images by nonlocal functionals. *SIAM Multiscale Modeling & Simul.*, 4(4):1091 – 1115, 2005.

[13] M. Nikolova. A variational approach to remove outliers and impulse noise. *JMIV*, 20(1-2):99–120, 2004.

[14] Gabriel Peyré, Sébastien Bougleux, and Laurent D Cohen. Non-local regularization of inverse problems. *Inverse Problems and Imaging*, 5(2):511–530, 2011.

[15] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[16] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.

[17] D. Zosso, G. Tran, and S. Osher. Non-local retinex—a unifying framework and beyond. *SIAM Journal on Imaging Sciences*, 8(2):787–826, 2015.