Course 049064: Variational Methods in Image Processing

Notes 4: BV and Euler-Lagrange

Guy Gilboa

1 The Space BV

Reminder (i) – integration by parts: To derive it we can begin with the identity: $(uv)' \equiv u'v + uv'$, where $'$ is the spatial differentiation. For 1D, integration of both sides gives $\int_A^B \frac{d}{dx}(uv)dx = \int_A^B u'vdx + \int_A^B uv'dx$, and therefore

$$\int_A^B u'vdx = (uv)|_A^B - \int_A^B uv'dx.$$

In any dimension, for a function $u(x)$ and a vector function $\mathbf{v} = (v_1(x), .., v_N(x))$ we have:

$$\int_\Omega \nabla u \cdot \mathbf{v} dx = \int_{\partial\Omega} u(\mathbf{v}\cdot\mathbf{n})ds - \int_\Omega u \,\mathrm{div}\, \mathbf{v} dx,$$

where $\mathbf{n}$ is the outward unit normal vector to $\partial\Omega$. For $(\mathbf{v}\cdot\mathbf{n})|_{\partial\Omega} = 0$ we get

$$\int_\Omega \nabla u \cdot \mathbf{v} dx = -\int_\Omega u \,\mathrm{div}\, \mathbf{v} dx.$$

Reminder (ii) – distributions: This is a generalization of integrals of functions based on derivatives, where the standard derivatives are not everywhere defined.

Let φ be a test function with the following properties:

1. φ is smooth (infinitely differentiable in principle).
2. φ has a compact support (is identically zero outside some bounded interval).

In cases where we have an integral of an expression with derivatives, we can write it using a test function and use integration by parts to “move” the derivative to the test function. We thus have an alternative way to define derivatives in a weak-sense which holds also in cases where standard derivative notions are undefined.

Distributional derivatives: The distributional (or weak) derivative of order k of f is g if for all $\varphi \in \mathcal{C}_c^\infty(\Omega)$ we have

$$\int_\Omega g(x)\varphi(x)dx = (-1)^k \int_\Omega f(x)\partial_x^k \varphi(x)dx.$$

Example: Consider the function $f(x) \in \mathbb{R}$ defined as $f = 0$ if $x \leq 0$ and $f = x$ if $x > 0$. Its distributional derivative can be computed using:

$$-\int_{\mathbb{R}} f(x)\varphi'(x)dx = -\int_0^\infty x\varphi'(x)dx = \int_0^\infty \varphi(x)dx = \int_{\mathbb{R}} H(x)\varphi(x)dx,$$

where $H(x)$ is the Heaviside function $H(x) = 0$ if $x \leq 0$ and $H(x) = 1$ if $x > 0$. Thus $H(x)$ is the weak derivative of $f(x)$.

1.1 Definition of BV

Let Ω be a bounded open subset of $\mathbb{R}^N$ and let $u \in L^1(\Omega)$.

Let $\varphi = (\varphi_1, ..\varphi_N)$ be a continuously differntiable function with compact support in Ω (belongs to the space $\mathcal{C}_0^1(\Omega)^N$). We denote

$$|\varphi|_{L^\infty} = \sup_x \sqrt{\sum_{i=1}^N \varphi_i^2(x)}.$$

We define the total variation in the distributional sense as:

$$\int_\Omega |Du| = \sup_\varphi \left\{ \int_\Omega u \operatorname{div} \varphi dx, \; |\varphi|_{L^\infty} \leq 1 \right\} \tag{1}$$

We refer to Du as the distributional gradient.

The space $BV(\Omega)$ is defined as the space of functions of bounded variation:

$$BV(\Omega) = \left\{ u \in L^1(\Omega); \; \int_\Omega |Du| < \infty \right\}. \tag{2}$$

$BV(\Omega)$ is a Banach space endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1} + \int_\Omega |Du|$.

2 Euler-Lagrange

2.1 Gâteaux derivative

Let X be a Banach space , $u, v\ inX$ and $F : X \to \mathbb{R}$. The Gâteaux derivative is defined as:

$$F'(u) = \lim_{\lambda \to 0^+} \frac{F(u + \lambda v) - F(u)}{\lambda}.$$

In the variational context, we refer to v as the variation and assume zero contribution on the boundary, $v|_{\partial\Omega} = 0$. We also assume that F is Gâteaux differentiable, that is for all v we reach the same limit.

The equation

$$F'(u) = 0 \tag{3}$$

is called the Euler-Lagrange (E-L) equation. We usually refer to $F'(u)$ as the E-L of functional $F(u)$.

If we have a minimization problem $\inf_{u \in X} F(u)$ (where inf is the infimum) the solution u admits Eq. (3). On the other hand, if F is convex, then a solution u of Eq. (3) is also a solution of the minimization problem. For more details see [2], Ch. 2, p. 37.

2.2 E-L of some functionals

Let a functional F be defined by

$$F(u) = \int_\Omega f(x, u(x), \nabla u(x))dx.$$

We view f as a function of three terms: $f(x, u, \xi)$ in $\mathbb{R}^N$, where $x = (x_1, .., x_N)$, $\xi = (\xi_1, .., \xi_N)$. The derivative of the functionsl (E-L) is:

$$F'(u) = \partial_u F(u) = \frac{\partial f(x, u, \nabla u)}{\partial u} - \sum_{i=1}^{N} \frac{\partial}{\partial x_i} \left(\frac{\partial f(x, u, \nabla u)}{\partial \xi_i} \right) \tag{4}$$

For precise smoothness and growth conditions see [2] (same place as above).

For regularization we are usually concerned only with derivatives of u, thus we can have a simplified formulation. Let $J(u)$ be a regularization functional of the form:

$$J(u) = \int_\Omega \Phi(|\nabla u(x)|)dx. \tag{5}$$

Then the E-L, or the gradient, of the functional is

$$\partial_u J(u) = -\operatorname{div} \left(\frac{\Phi'(|\nabla u|)}{|\nabla u|} \nabla u \right) \tag{6}$$

To minimize a convex functional, we can perform a gradient descent, initializing with some $f(x)$. As a continuous time process the gradient descent evolution is:

$$u_t = -\partial_u J(u) = \operatorname{div}\left(\frac{\Phi'(|\nabla u|)}{|\nabla u|}\nabla u\right), \; u|_{t=0} = f(x) \tag{7}$$

This is exactly a nonlinear, Perona-Malik style, diffusion equation, where the diffusion coefficient c is

$$c(|\nabla u|) = \frac{\Phi'(|\nabla u|)}{|\nabla u|}.$$

Thus we can view nonlinear diffusion as a process which minimizes a regularizing functional.

Remark: Note that the original P-M diffusion coefficients can be shown as minimizing non-convex energy functionals, thus their descent can theoretically get stuck in a local minimum. However it was shown that discretization or the Catte et al gradient smoothing avoid this problem.

2.3 Some useful examples

Linear diffusion: $\Phi(s) = \frac{1}{2}s^2$.

$$J_{Lin}(u) = \frac{1}{2}\int_\Omega |\nabla u(x)|^2 dx.$$

The gradient descent is the linear diffusion equation $c = 1$:

$$u_t = \operatorname{div}(\nabla u) = \Delta u.$$

TV Flow: $\Phi(s) = s$.

$$J_{TV}(u) = \int_\Omega |\nabla u(x)| dx.$$

The gradient descent is called the TV-flow [1, 3], which has special properties, qualitatively similar to the ROF model [5]. We can view it as nonlinear diffusion with $c = \frac{1}{|\nabla u|}$ (note the unbounded value at for zero gradient):

$$u_t = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

Perona-Malik: For the diffusion coefficient $c(s) = \frac{1}{1+(s/k)^2}$ we get $\Phi(s) = \frac{1}{2}k^2 \log(1 + (s/k)^2)$.

$$J_{PM}(u) = \frac{1}{2}k^2 \int_\Omega \log(1 + (|\nabla u(x)|/k)^2) dx.$$

This is a nonconvex functional. The gradient descent is naturally the P-M equation:

$$u_t = \operatorname{div}\left(\frac{\nabla u}{1+(|\nabla u|/k)^2}\right).$$

Charbonnier: For the diffusion coefficient $c(s) = \frac{1}{\sqrt{1+(s/k)^2}}$ of Charbonnier et al [4] we get $\Phi(s) = k\sqrt{k^2+s^2} - k^2$.

$$J_{Ch}(u) = k\int_\Omega \sqrt{k^2+|\nabla u(x)|^2}dx.$$

This is a convex functional. The gradient descent is:

$$u_t = \operatorname{div}\left(\frac{\nabla u}{\sqrt{1+(|\nabla u|/k)^2}}\right).$$

2.4 E-L of common fidelity terms

L^2 **norm square:** $F_{(L^2)^2}(u) = \|u\|^2_{L^2} = \int_\Omega (u(x))^2 dx$.

$$\partial_u F_{(L^2)^2}(u) = 2u.$$

Convolution:
L^1 **norm:** $F_{L^1}(u) = \|u\|_{L^1} = \int_\Omega |u(x)|dx$.

$$\partial_u F_{L^1}(u) = \operatorname{sign}(u).$$

where sign is the signum function: $\operatorname{sign}(q) = 1$ for $q > 0$, $= -1$ and $q < 0$. For $q = 0$ we often keep the vague definition $\operatorname{sign}(0) \in (-1, 1)$, to keep it consistent with the notion of subdifferential.

2.5 Regularization energies

It is now straightforward to obtain the E-L of some standard regularization models. We will write the negative expression $-\partial_u E$ which is useful for the gradient descent equation.

2.5.1 TV Denoising

$$E_{TV} = |u|_{TV} + \lambda\|f-u\|^2_{L^2}.$$

$$-\partial_u E_{TV} = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) + 2\lambda(f-u).$$

Epsilon approximation: As for $\nabla u = 0$ the $E - L$ expression is undefined, for numerical purposes an epsilon approximation is often used:

$$E_{TV-\varepsilon} = |u|_{TV-\varepsilon} + \lambda \|f - u\|_{L^2}^2,$$

where

$$|u|_{TV-\varepsilon} = \int_\Omega \sqrt{|\nabla u(x)|^2 + \varepsilon^2} dx.$$

The respective E-L is

$$-\partial_u E_{TV-\varepsilon} = \text{div}\left(\frac{\nabla u}{\sqrt{|\nabla u(x)|^2 + \varepsilon^2}} \right) + 2\lambda(f - u).$$

So as $|\nabla u| \gg \varepsilon$ we approximate well the total variation descent. Near zero gradient, for $|\nabla u| \ll \varepsilon$ we approach linear diffusion with $c \to \frac{1}{\varepsilon}$.

Numerical implementation problem: We recall that for explicit scheme the time step is bounded by the CFL condition, which is inversely proportional to the diffusion coefficient. Therefore the time step Δt is proportional to ε. Thus as we model more precisely TV we need to use much more smaller time-steps, or many more numerical iterations to simulate the same evolution time.

Thus for TV – alternative non-explicit numerical schemes can be much more efficient.

2.5.2 TV Deconvolution

As we learned earlier, here a kernel $H(x)$ is assumed to convolve the input image. The fidelity term changes to the following energy:

$$E_{TV-deconv} = |u|_{TV} + \lambda \|f - u * H\|_{L^2}^2.$$

$$-\partial_u E_{TV-deconv} = \text{div}\left(\frac{\nabla u(x)}{|\nabla u(x)|} \right) + 2\lambda H(-x) * (f(x) - H(x) * u(x)),$$

Here also, the TV term is often replaced by its epsilon approximation.

2.5.3 TV-L1 outlier removal

$$E_{TV-L1} = |u|_{TV} + \lambda \|f - u\|_{L^1}.$$

$$-\partial_u E_{TV-L1} = \text{div}\left(\frac{\nabla u}{|\nabla u|} \right) + \lambda\, \text{sign}(f - u).$$

In this case both the TV term and the sign term are epsilon approximated.

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] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[3] G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in R^N. *Journal of Differential Equations*, 184(2):475–525, 2002.

[4] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. IEEE ICIP '94*, volume 2, pages 168–172, 1994.

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