Course 049064: Variational Methods in Image Processing

Notes 10: Convex Analysis and Spectral Total-Variation

Guy Gilboa

1 Some Basic Notions in Convex Analysis

We will start by defining several basic notions in convex analysis. They serve as very good mathematical tools to better understand minimization of convex energies. Some numerical algorithms rely on this theory, like Chambolle's projection algorithm [4]. This is a sketchy outline on the topic. More can be found in the following books [10, 5, 3].

Convex set: A set $C \subseteq \Omega$ is convex iff (if and only if) for any $x, y \in C$ and $\alpha \in [0, 1]$

$$\alpha x + (1 - \alpha)y \in C.$$

Convex function and functional: A function $f(x)$ is convex iff for any $x, y \in \Omega$ and $\alpha \in [0, 1]$

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

We can define the *epigraph* of the function as the set of points above or equal to f:

$$\text{epi}(f) = \{(x, w) : x \in \Omega,\, w \in \mathbb{R},\, w \geq f(x)\}.$$

Then an equivalent definition for convexity: f is convex iff epi(f) is a convex set.

Functionals can be defined in the same way. A functional $J(u(x))$ is convex iff for any $u_1(x), u_2(x)$, $x \in \Omega$ and $\alpha \in [0, 1]$

$$J(\alpha u_1(x) + (1 - \alpha)u_2(x)) \leq \alpha J(u_1(x)) + (1 - \alpha)J(u_2(x)).$$

It is ease to show that functionals of the form $J = \int_\Omega f(u(x))dx$ are convex if $f(\cdot)$ is convex.

As q^2, $|q|$ and $\sqrt{q^2+\varepsilon^2}$ are all convex in q, we can assign $q = |\nabla u(x)|$ and see that

- $J_{H^1} = \int_\Omega |\nabla u|^2 dx$
- $J_{TV} = \int_\Omega |\nabla u| dx$
- $J_{TV-\varepsilon} = \int_\Omega \sqrt{|\nabla u|^2 + \varepsilon^2} dx$

are all convex functionals.

Remark 1: We say that a function is strictly convex if for $\forall x \neq y$, $\alpha \in (0,1)$ the inequality is strict:

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y).$$

The same is for functionals. Therefore J_{H^1} is strictly convex, whereas J_{TV} is not, which yields some unique properties.

Remark 2: If f is twice differentiable in 1D then iff $f'' \geq 0$ then f is convex. If $f'' > 0$ then f is strictly convex (but not vice-versa, see e.g. $f(x) = x^4$ at $x = 0$). For N dimensions, if the $N \times N$ Hessian matrix $H = (\partial_{x_i x_j} f)$, $(i, j = 1, .., N)$ is positive semi-definite then the function is convex.

Why convex functions are good?

1. All local minima are global minima.
2. Very well understood functions.
3. There are many efficient numerical methods to minimize convex functionals.
4. Strong duality tools are available both for efficient numerics and for developing theory.

Subdifferential: Extension of the derivative to non-differentiable cases. For functions we define the following set:

$$\partial f(x) := \{m : f(y) \geq f(x) + m(y-x), \forall y \in \Omega\}. \tag{1}$$

- Every element m in the set is a called a *subgradient* and the entire set is called the *subdifferential.*
- If f is differentiable (at least once) then $\partial f(x)$ has exactly one element which is the gradient at the point $\partial f(x) = \nabla f(x)$.

- A point x_0 is a global minimum of f iff zero is contained in the subdifferential at that point $0 \in \partial f(x_0)$.

For functionals we assume u, v are in some space X and p is in the dual space X^* and define

$$\partial J(u) := \{p \in X^* : J(v) \geq J(u) + \langle p, v-u \rangle, \forall v \in X\} . \tag{2}$$

Duality - Legender-Fenchel transform:

$$f^*(m) := \sup_{x \in \Omega} \{m : mx - f(x)\} . \tag{3}$$

For functionals:

$$J^*(p) := \sup_{u \in X} \{p : \langle p, u \rangle - J(u)\} . \tag{4}$$

Some properties:

1. J^* is convex.
2. $J^{**} = J$.
3. For differentiable J we have $\partial_u J(u) = p$, $\partial_p J^*(p) = x$.
4. if J is not convex, J^* and J^{**} are still convex, J^{**} is the largest convex function satisfying $J^{**}(u) \leq J(u)$ and is called the convex envelope of $J(u)$.

2 Spectral TV Decomposition

In [6, 7] a TV-spectral framework was formulated with a TV transform and inverse transform. It affords a different representation of the image and allows a continuous separation of scales with full edge and contrast preservation.

2.1 Preliminaries

2.1.1 Reminder: TV and TV-flow

The total variation for smooth u:

$$J_{\tilde{TV}} = \int_{\Omega} |\nabla u| dx. \tag{5}$$

For nonsmooth and possibly discontinuous u the distributional definition is used:

$$J_{TV} = \int_\Omega |Du| = \sup_z \{\langle u, \mathrm{div}(z)\rangle, \|z\|_\infty \leq 1\}. \tag{6}$$

The gradient descent which minimizes the total variation energy (with the input image f as initial condition), referred to also as *TV-flow*, is

$$u_t(t,x) = -\partial_u J_{TV}(u(t,x)) = \mathrm{div}\left(\frac{Du}{Du}\right), \; u(t=0,x) = f(x). \tag{7}$$

Existence and uniqueness of the TV-flow were proved in [1].

2.1.2 Convex one-homogeneous functionals

Convex one-homogeneous functionals have the following property for any $0 < \alpha \in \mathbb{R}$:

$$J(\alpha u) = \alpha J(u). \tag{8}$$

We denote a subgradient p which admits (1) $p(u) \in \partial J(u)$. For one-homogeneous functionals we have

$$J(u) = \langle u, p(u)\rangle, \tag{9}$$

and thus

$$J(v) \geq \langle p(u), v\rangle, \; \forall v. \tag{10}$$

We also have for all $\alpha > 0$

$$p(\alpha u) = p(u). \tag{11}$$

2.2 Nonlinear Eigenfunctions

Eigenfunctions of linear operators admit the following linear eigenvalue problem

$$Lu = \lambda u, \tag{12}$$

where L is a linear operator. Eigenfunctions of the graph-Laplacian have been used extensively to solve many signal processing, computer vision and machine-learning problems such as segmentation [11], clustering [9] and subspace clustering [8], dimensionality reduction [2], and more.

We view the subdifferential as a nonlinear operator and examine the following non-linear eigenvalue problem:

$$p(u) = \lambda u, \; p(u) \in \partial J(u). \tag{13}$$

Functions u admitting (13) are called eigenfunctions (of $J(u)$) with λ their corresponding eigenvalue.

When f is an eigenfunction, the TV-flow, Eq. (7), has a simple analytic solution:

$$u(t,x) = (1-\lambda t)^{+} f(x), \tag{14}$$

where $(X)^{+} := \max(0, X)$.

2.3 TV Transform

The TV transform is defined by:

$$\phi(t;x) = u_{tt}(t;x)t, \tag{15}$$

where u_{tt} is the second time derivative of the solution $u(t;x)$ of the TV-flow (7). The inverse transform is:

$$f(x) = \int_0^\infty \phi(t;x)dt + \bar{f}, \tag{16}$$

where $\bar{f} = \frac{1}{\Omega}\int_\Omega f(x)dx$ is the mean value of the initial condition. Filtering is defined by a non-negative amplification function $H(t)$:

$$\phi_H(t;x) = \phi(t;x)H(t),$$

where the filtered image f_H is computed by reconstructing with ϕ_H (instead of ϕ) in Eq. (16). The spectrum $S(t)$ corresponds to the amplitude of each scale:

$$S(t) = \|\phi(t;x)\|_{L^1} = \int_\Omega |\phi(t;x)|dx. \tag{17}$$

Two significant results were shown in [7] for this transform:

- **Atoms as eigenfunctions:** Let $f(x)$ be an eigenfunction (admitting (13)) with eigenvalue λ, then the transform yields a single impulse, multiplied by $u(x)$, at $t = 1/\lambda$ and is zero for all other t:

$$\phi(t,x) = \delta(t - \frac{1}{\lambda})f(x). \tag{18}$$

- **Relations to TV-flow:** The TV flow solution $u(t_1)$ is a specific low-pass filter in this framework with the following scale-attenuation:

$$H_{TVF,t_1}(t) = \frac{t-t_1}{t},\ t_1 \leq t < \infty \text{ and zero otherwise.} \tag{19}$$

Ideal low-pass-filter (iLPF):

$$iLPF(t) := \int_t^\infty \phi(\tau)d\tau = u(t) + tp(u(t)). \tag{20}$$

Orthogonality of ϕ and u for any $t > 0$,

$$\langle \phi(t,x), u(t,x) \rangle = 0. \tag{21}$$

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] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[3] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar. *Convex analysis and optimization.* Athena Scientific Belmont, 2003.

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

[5] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* Elsevier, 1976.

[6] G. Gilboa. A spectral approach to total variation. In *A. Kuijper et al. (Eds.): SSVM 2013*, volume 7893 of *Lecture Notes in Computer Science*, pages 36–47. Springer, 2013.

[7] G. Gilboa. A total variation spectral framework for scale and texture analysis. *SIAM J. Imaging Sciences*, 7(4):1937–1961, 2014.

[8] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):171–184, 2013.

[9] A.Y. Ng, M.I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[10] R. T. Rockafellar. *Convex analysis.* Princeton university press, 1997.

[11] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.