Dual methods for the minimization of the total variation

Rémy Abergel
supervisor Lionel Moisan

MAP5 - CNRS UMR 8145

– Different Learning Seminar, LTCI –

Thursday 21st April 2016

Plan

Plan

Mathematical framework

A gray level image is represented as a function

$$u : \Omega \to \mathbb{R}$$

where Ω denotes

- **Continuous framework**: a bounded open set of $\mathbb{R}^2$.
- **Discrete framework**: a rectangular subset of $\mathbb{Z}^2$.

In both cases, we will note $u \in \mathbb{R}^\Omega$.

Total variation (continuous framework)

We will focus on image restoration process involving the total variation functional, which is defined by

$$\forall u \in W^{1,1}(\Omega), \quad \mathrm{TV}(u) = \int_\Omega \|\nabla u(x)\|_2 dx \,,$$

or, more generally,

$$\forall u \in \mathrm{BV}(\Omega), \quad \mathrm{TV}(u) = \sup_{\substack{\phi \in \mathscr{C}_c^\infty(\Omega;\mathbb{R}^2) \\ \forall x \in \Omega,\ \|\phi(x)\|_2 \leq 1}} -\int_\Omega u(x) \mathrm{div}\phi(x) dx \,,$$

where $\mathrm{BV}(\Omega) = \left\{u \in L^1_{\mathrm{loc}}(\Omega);\, \mathrm{TV}(u) < +\infty\right\}$.

Total variation (discrete framework)

Let $\Omega = \{0, \dots, M-1\} \times \{0, \dots, N-1\}$ denote a discrete rectangular domain, and $u \in \mathbb{R}^\Omega$ a discrete image. We generally adapt the continuous definition of $\mathrm{TV}(u)$ as follows,

$$\mathrm{TV}(u) = \|\nabla u\|_{1,2} := \sum_{(x,y)\in\Omega} \|\nabla u(x,y)\|_2 \,,$$

where ∇ denotes a finite difference operator.

Plan

Optimization problem

We are interested in the computation of $\widehat{u} \in E$, a minimizer of a given cost function J over a subset $\mathscr{C} \subset E$ (constraint set). Such a problem is usually written

$$\widehat{u} \in \operatorname*{argmin}_{u \in \mathscr{C}} J(u)$$

- J denotes a function from E to $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$,
- E denotes (for shake of simplicity) a Hilbert space,
- in general, $\mathscr{C} = \{u \in \mathbb{E};\ g(u) \leq 0\ ,\ h(u) = 0\}$
  - where g is called the inequality constraint,
  - and h is called the equality constraint.

Differentiable and unconstrained framework

Theorem (first order necessary condition for optimality)

If $\widehat{u}$ achieves a minimum of J over E, then $\nabla J(\widehat{u}) = 0$.

This condition becomes **sufficient** when the cost function J is **convex**.

Theorem (sufficient condition for the existence of a minimizer)

*If $J : E \to \overline{\mathbb{R}}$ is a **proper**, **continuous** and **coercive** function, then the **unconstrained** problem admits at least one solution.*

If moreover J is **strictly convex**, the problem admits **exactly** one solution.

Example of resolvant algorithm ($\mathscr{C} = E = \mathbb{R}^n$)

Algorithm (gradient descent)

1. ***Initialization:***
   - *Choose* $u_0 \in \mathbb{R}^n$, $\alpha_0 > 0$ *and* $\varepsilon > 0$.
2. ***Iteration: k***
   - *compute* $\nabla J(u_k)$
   - *compute* α_k
   - $u_{k+1} = u_k - \alpha_k \nabla J(u_k)$
3. ***Example of stopping criterion:***
   - *if* $\|J(u_{k+1}) - J(u_k)\| < \varepsilon$, *STOP*
   - *otherwise, set* $k = k + 1$ *and go back to 2.*

Remark: a first order Taylor expansion of $J(u_k + \alpha_k \nabla J(u_k))$ at point u_k helps to understand that $J(u_{k+1}) \leq J(u_k)$ as soon as α_k is small enough.

Differentiable and constrained framework

- Theorems can be adapted (in the convex setting), leading to the so-called **Karush-Kuhn-Tucker** conditions.
- A numerical solution of the constrained problem can be numerically computed using the **projected gradient** algorithm, which simply consists in replacing

$$u_{k+1} = u_k - \alpha_k \nabla J(u_k)$$

by

$$u_{k+1} = \text{Proj}_{\mathscr{C}}\left(u_k - \alpha_k \nabla J(u_k)\right)$$

into the gradient descent algorithm.

Legendre-Fenchel transform

Let E denote a finite dimensional Hilbert space, $E^\star$ its dual space, and $\langle \cdot, \cdot \rangle$ the bilinear mapping over $E^\star \times E$ defined by

$$\forall \varphi \in E^\star, \quad \forall u \in E, \quad \langle \varphi, u \rangle = \varphi(u)\,.$$

Definition (affine continuous applications)

An affine continuous application is a funtion of the type

$$\mathcal{A} : u \;\mapsto\; \langle \varphi, u \rangle + \alpha$$

- *where $\varphi \in E^\star$ is called **the slope** of $\mathcal{A}$,*
- *and α is a real number, called the **constant term** of $\mathcal{A}$.*

Legendre-Fenchel transform

Q. At which condition(s) does the affine continuous application $\mathcal{A}$, with slope $\varphi \in E^\star$ and constant term $\alpha \in \mathbb{R}$, lower bound J everywhere on E?

$$\forall u \in E, \quad \mathcal{A}(u) \leq J(u)$$

$$\begin{aligned}
&\Leftrightarrow \quad \forall u \in E, \quad \langle \varphi, u \rangle + \alpha \leq J(u) \\
&\Leftrightarrow \quad \forall u \in E, \quad \langle \varphi, u \rangle - J(u) \leq -\alpha \\
&\Leftrightarrow \quad \sup_{u \in E} \{ \langle \varphi, u \rangle - J(u) \} \leq -\alpha \\
&\Leftrightarrow \quad J^\star(\varphi) \leq -\alpha \\
&\Leftrightarrow \quad -J^\star(\varphi) \geq \alpha
\end{aligned}$$

Legendre-Fenchel transform

Definition (Legendre-Fenchel transform)

Let $J : E \to \overline{\mathbb{R}}$, the Legendre-Fenchel transform of J is the application $J^\star : E^\star \to \overline{\mathbb{R}}$ defined by:

$$\forall \varphi \in E^\star, \quad J^\star(\varphi) = \sup_{u \in E} \{ \langle \varphi, u \rangle \; - \; J(u) \}$$

Geometrical intuition:

$-J^\star(\varphi)$ represents the largest constant term α that can assume any affine continuous function with slope φ, to remain under J everywhere on E.

Transformée de Legendre-Fenchel

By definition of $J^\star$, we have

$$\forall \varphi \in E^\star, \quad J^\star(\varphi) = \sup_{u \in E} \{ \, \langle \varphi, u \rangle \; - \; J(u) \, \} \, .$$

We remark that

- $J^\star(0_{E^\star}) = - \inf_{u \in E} J(u)$
- we retrieve here a link between "**null slope**" and "**infimum of J**"

Subdifferentiability

Definition (exact applications)

Let $u \in E$, $\varphi \in E^\star$, then, the affine continuous application

$$\mathcal{A} : v \mapsto \langle \varphi, v - u \rangle + J(u)$$

*satisfies $\mathcal{A}(u) = J(u)$. We say that $\mathcal{A}$ is **exact** at u.*

Definition (subdifferentiability & subgradient)

*A $J : E \to \overline{\mathbb{R}}$ is said **subdifferentiable at the point** $\mathbf{u} \in \mathbf{E}$ if it admits at least one **lower bounding affine continuous function** which is **exact at** $\mathbf{u}$.*

- *The slope φ of such an affine function is then called a subgradient of J at the point u.*
- *The set of all subgradients of J at u is noted $\partial J(u)$.*

Subdifferentiability

Basic properties:

- $\varphi \in \partial J(u) \quad \Leftrightarrow \quad \forall v \in E, \ \langle \varphi, v-u \rangle + J(u) \leq J(v)$
- $0 \in \partial J(\widehat{u}) \quad \Leftrightarrow \quad \widehat{u} \in \operatorname*{argmin}_{u \in E} J(u)$

Remark: transformation of a constrained problem into an unconstrained problem

$$\operatorname*{argmin}_{u \in \mathscr{C}} J(u) \quad = \quad \operatorname*{argmin}_{u \in E} J(u) + \imath_{\mathscr{C}}(u)$$

$$\text{where } \imath_{\mathscr{C}}(u) = \begin{cases} 0 & \text{si } u \in \mathscr{C} \\ +\infty & \text{si } u \notin \mathscr{C} \end{cases}$$

Properties & subdifferential calculus

- Any convex and lower semi-continous (l.s.c.) function is subdifferentiable over the interior of its domain.
- If J is convex and differentiable at u, then $\partial J(u) = \{\nabla J(u)\}$.
- $\forall u \in E,\ \partial(J_1 + J_2)(u) \supset \partial J_1(u) + \partial J_2(u)\,.$
- The converse inclusion is satisfied under some additional (but weak) hypotheses on J_1 and J_2.
- If J is convex, lower semi-continuous, then

$$\varphi \in \partial J(u) \quad \Leftrightarrow \quad u \in \partial J^\star(\varphi)\,.$$

- If J is convex, and lower semi-continuous, then $J^{\star\star}(u) = J(u)\,.$

Plan

Legendre-Fenchel transform of the discrete TV

Theorem (Legendre-Fenchel transform of the discrete TV)

The Legendre-Fenchel transform of TV *is the indicator function of the convex set* $\mathscr{C} = \mathrm{div}\mathscr{B}$, *where*

$$\mathscr{B} = \left\{p \in \mathbb{R}^\Omega \times \mathbb{R}^\Omega,\ \|p\|_{\infty,2} \leq 1\right\},$$

and $\|\cdot\|_{\infty,2} := p \mapsto \max_{(x,y)\in\Omega} \|p(x,y)\|_2$ *is the dual norm of the* $\|\cdot\|_{1,2}$ *norm.*

In other words:

$$\mathrm{TV}^\star(\varphi) \;=\; \imath_{\mathscr{C}}(\varphi) = \begin{cases} 0 & \text{if } \exists p \in \mathscr{B},\ \varphi = \mathrm{div} p \\ +\infty & \text{otherwise.} \end{cases}$$

Proof: this result is easy to prove using the convex analysis tools presented before (see the proof in appendix).

The ROF (Rudin, Osher, Fatemi) model

We are interested in the computation of

$$\widehat{u}_{\mathrm{MAP}} = \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} J(u) := \frac{1}{2}\|u - u_0\|_2^2 + \lambda \mathrm{TV}(u) .$$

Thanks to the previous properties, we have

$$\begin{aligned}
\widehat{u}_{\mathrm{MAP}} &= \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \frac{1}{2}\|u - u_0\|_2^2 + \lambda \mathrm{TV}(u) \\
\Leftrightarrow\ & 0 \in \widehat{u}_{\mathrm{MAP}} - u_0 + \lambda \partial \mathrm{TV}(\widehat{u}_{\mathrm{MAP}}) \\
\Leftrightarrow\ & \widehat{u}_{\mathrm{MAP}} \in \partial \mathrm{TV}^\star \left(\frac{u_0 - \widehat{u}_{\mathrm{MAP}}}{\lambda} \right) \\
\Leftrightarrow\ & \frac{u_0}{\lambda} \in \frac{u_0 - \widehat{u}_{\mathrm{MAP}}}{\lambda} + \frac{1}{\lambda} \partial \mathrm{TV}^\star \left(\frac{u_0 - \widehat{u}_{\mathrm{MAP}}}{\lambda} \right)
\end{aligned}$$

The ROF (Rudin, Osher, Fatemi) model

Dual formulation of the ROF problem: Let $\widehat{w} = \frac{u_0 - \widehat{u}_{\text{MAP}}}{\lambda}$, we have

$$0 \in \widehat{w} - u_0/\lambda + \frac{1}{\lambda}\partial \mathrm{TV}^\star(\widehat{w}),$$

Thus,

$$\widehat{w} = \operatorname*{argmin}_{w \in \mathbb{R}^\Omega} \frac{1}{2}\|w - u_0/\lambda\|_2^2 + \frac{1}{\lambda}\mathrm{TV}^\star(w).$$

Last, since $\mathrm{TV}^\star(w) = \imath_{\mathscr{C}}(w)$, we have

$$\widehat{w} = \operatorname*{argmin}_{w \in \mathscr{C}} \|w - u_0/\lambda\|_2^2 = \mathrm{Proj}_{\mathscr{C}}(u_0/\lambda),$$

an thus, $\widehat{u}_{\text{map}} = u_0 - \lambda\,\mathrm{Proj}_{\mathscr{C}}(u_0/\lambda)$.

Inverse problems (primal-dual approach)

$$A : \mathbb{R}^\Omega \to \mathbb{R}^\omega, \qquad \widehat{u}_{\text{MAP}} = \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \frac{1}{2}\|Au - u_0\|^2 \; + \; \lambda \text{TV}(u) \,.$$

Primal-dual formulation: Let us use $F^{\star\star} = F$ (valid as soon as F is convex, and lower semi-continuous).

- $\text{TV}(u) = \text{TV}^{\star\star}(u)$ yields a dual formulation (also called weak formulation) of the discrete TV,

$$\text{TV}(u) = \max_{p \in \mathscr{B}} \langle \nabla u, p \rangle \,.$$

- $\frac{1}{2}\|Au - u_0\|_2^2 = f(Au) = f^{\star\star}(Au) = \max_{q \in \mathbb{R}^\omega} \langle q, Au \rangle - f^\star(q) \,,$
  and we can easily show that $f^\star(q) = \frac{1}{2}\|q + u_0\|_2^2 - \frac{1}{2}\|u_0\|_2^2$.

Inverse problems (primal-dual approach)

By replacing these two terms into the initial problem, we get a **primal-dual reformulation**:

$$\widehat{u}_{\text{MAP}} = \operatorname*{argmin}_{u \in \mathbb{R}^\Omega} \max_{\substack{p \in \mathscr{B} \\ q \in \mathbb{R}^\omega}} \ \langle (\lambda \nabla u, Au), (p, q) \rangle \ - \ \frac{1}{2} \|q + u_0\|_2^2$$

Such a problem can be handled using the Chambolle-Pock algorithm (2011), which boils down to the numerical scheme

$$\begin{cases} p^{n+1} & = \operatorname{Proj}_{\mathscr{B}} \left(p^n + \sigma \lambda \nabla \overline{u}^n \right) \\ q^{n+1} & = \left(q^n + \sigma \left(A \overline{u}^n - u_0 \right) \right) / (1 + \sigma) \\ u^{n+1} & = u^n + \tau \lambda \operatorname{div} p^{n+1} - \tau A^* q^{n+1} \\ \overline{u}^{n+1} & = u^{n+1} + \theta \left(u^{n+1} - u^n \right) \end{cases}$$

The convergence of the iterates (u^n, p^n, q^n) toward a solution of the primal-dual problem is ensured for $\theta = 1$ and $\tau\sigma < |||K|||^2$, noting $K = u \mapsto (\lambda \nabla u, Au)$.

Plan

Conclusion

- The tools presented here are based on very **simple notions**.
- They are useful to reformulate a (convex) problem into a dual (or primal-dual) one, which can be sometimes **much more** simple than the initial problem.
- What is the **good framework** for using these tools?
  - The cost function must be **convex** and **lower semi-continuous** (Γ space). When it is not the case, it may be replaced by a convex approximation (Γ-regularization, Moreau-Yoshida envelope, surrogate functions, etc.).
  - A dual reformulation often starts with the computation of the **Legendre-Fenchel transform** of a part of the cost function (which is particularly easy in the case of ℓ^p norms).
  - The dual variables are easy to manipulate when E is a **Hilbert space**.

Bibliography

- I. Ekeland, and R. Témam: "Convex analysis and variational problems", 1999.
- A. Chambolle, and T. Pock: "A first-order primal-dual algorithm for convex problems with applications to imaging", 2011.
- A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock: "An introduction to total variation for image analysis", 2010.
- R. T. Rockafellar: "Convex analysis", 1997.
- S. Boyd, and L. Vandenberghe: "Convex optimization", 2009.
- L. I. Rudin, S. Osher, and E. Fatemi: "Nonlinear total variation based noise removal algorithms", 1992.

Appendix (Computation of TV*)

Lemma (Legendre-Fenchel transform of a norm)

Let E denote a Hilbert space, endowed with a norm $\|\cdot\|$, and a scalar product $\langle\cdot,\cdot\rangle$. We have

$$\forall v \in E, \quad \|v\|^* = \imath_{\mathscr{B}_*}(v) := \begin{cases} 0 & \text{if } \|v\|_* \leq 1 \\ +\infty & \text{otherwise,} \end{cases}$$

where $\|\cdot\|_* = v \mapsto \sup_{u\in E,\ \|u\|\leq 1}\langle v, u\rangle$ *denotes the dual norm.*

In other words, $\|\cdot\|^\star$ is the indicator function of the closed unit ball for the dual norm $\|\cdot\|_*$.

Proof. We have $\imath^\star_{\mathscr{B}_*}(u) = \sup_{v\in E,\ \|v\|_*\leq 1}\langle v, u\rangle = \|u\|_{**} = \|u\|$, for any $u \in E$. Thus, $\|\cdot\|^\star = \imath^{\star\star}_{\mathscr{B}_*} = \imath_{\mathscr{B}_*}$, since $\imath_{\mathscr{B}_*} \in \Gamma(E)$.

Appendix (Computation of TV*)

Lemma (dual norm of the $\|\cdot\|_{1,2}$ norm)

The two norms $\|\cdot\|_{1,2}$ and $\|\cdot\|_{\infty,2}$ over the Hilbert space $E := \mathbb{R}^\Omega \times \mathbb{R}^\Omega$ are dual to each other.

Proof. Since E is reflexive, we just need to show that one norm is the dual of the other. Let us show that $\|\cdot\|_{1,2}$ is the dual norm of $\|\cdot\|_{\infty,2}$. For any $p \in E$, we have

$$\begin{aligned}
\sup_{q\in E,\ \|q\|_{\infty,2}\leq 1} \langle p, q\rangle_E &= \sup_{\substack{q\in E\\ \forall x\in\Omega,\ \|q(x)\|_2\leq 1}} \textstyle\sum_{x\in\Omega} \langle p(x), q(x)\rangle_{\mathbb{R}^2} \\
&= \textstyle\sum_{x\in\Omega} \displaystyle\sup_{q(x)\in\mathbb{R}^2,\ \|q(x)\|_2\leq 1} \langle p(x), q(x)\rangle_{\mathbb{R}^2} \\
&= \textstyle\sum_{x\in\Omega} \|p(x)\|_2 \\
&= \|p\|_{1,2}\,.
\end{aligned}$$

Appendix (Computation of $\mathrm{TV}^\star$)

Theorem (Legendre-Fenchel transform of TV)

$\mathrm{TV}^\star = \imath_{\mathscr{C}}$, *where* $\mathscr{C} = \mathrm{div}\mathscr{B}$ *and* $\mathscr{B} = \{p \in E,\ \|p\|_{\infty,2} \leq 1\}$.

Proof.

- Since the two norms $\|\cdot\|_{1,2}$ and $\|\cdot\|_{\infty,2}$ are dual to each other, we have $\|\cdot\|_{1,2}^\star = \imath_{\mathscr{B}}$, and thus $\|\cdot\|_{1,2} = \|\cdot\|_{1,2}^{\star\star} = \imath_{\mathscr{B}}^\star$.
- Besides, for all $u \in \mathbb{R}^\Omega$, we have

$$\imath_{\mathscr{C}}^\star(u) = \sup_{v \in \mathscr{C}} \langle u, v\rangle = \sup_{p \in \mathscr{B}} \langle u, \mathrm{div} p\rangle = \sup_{p \in \mathscr{B}} \langle \nabla u, p\rangle = \imath_{\mathscr{B}}^\star(\nabla u)\,.$$

- Therefore, $\imath_{\mathscr{C}}^\star(u) = \imath_{\mathscr{B}}^\star(\nabla u) = \|\nabla u\|_{1,2} = \mathrm{TV}(u)$, for any u.
- Thus $\mathrm{TV} = \imath_{\mathscr{C}}^\star$, and finally $\mathrm{TV}^\star = \imath_{\mathscr{C}}^{\star\star} = \imath_{\mathscr{C}}$.