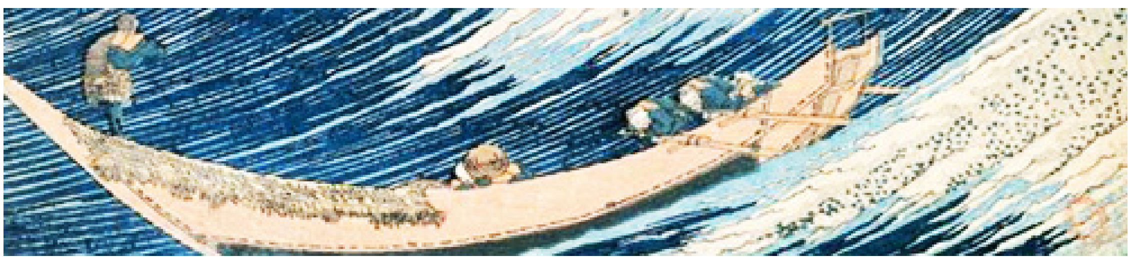
Mathematical Foundations of Data Sciences



Gabriel Peyré
CNRS & DMA
École Normale Supérieure
gabriel.peyre@ens.fr
www.gpeyre.com
www.numerical-tours.com

October 7, 2017

Chapter 2

Convex Duality

Main ref [5, 6, 3], [8, 2, 1]
TODO.

2.1 Forward-backward on the Dual

Chambolle's algorithm. Chambolle in [4] detail an algorithm to minimize exactly the TV denoising problem

$$f_\lambda^\star = \underset{g\in\mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2}\|f-g\|^2 + \lambda\|g\|_{\mathrm{TV}}. \tag{2.1}$$

It uses a relationship between the vectorial ℓ^1 and ℓ^∞ norms

$$\|v\|_1 = \sum_{m=0}^{N-1} \|v_m\| \quad \text{and} \quad \|v\|_\infty = \max_{0\leqslant m<N} \|v_m\|$$

where each $v_m \in \mathbb{R}^2$ and $v \in \mathbb{R}^{N\times 2}$. One has

$$\|v\|_1 = \max_{\|w\|_\infty \leqslant 1} \langle w,\, u\rangle$$

which allows one to re-write the optimization (2.1) as

$$\min_{g\in\mathbb{R}^N} \max_{\|w\|_\infty\leqslant 1} \frac{1}{2}\|f-g\|^2 + \lambda\langle w,\, \nabla g\rangle.$$

Exchanging the roles of the min and the max, one proves that the solution of (2.1) is re-written as

$$f_\lambda^\star = f + \lambda \operatorname{div}(w^\star) \tag{2.2}$$

where

$$w^\star \in \underset{\|w\|_\infty\leqslant 1}{\operatorname{argmin}} \|f + \lambda \operatorname{div}(w^\star)\|^2. \tag{2.3}$$

The convex optimization problem (2.3) computes a dual vector field $w^\star \in \mathbb{R}^{N\times 2}$, from which the denoised image is recovered using (2.2).

The dual problem (2.3) is the minimization of a quadratic functional subject to a convex ℓ^∞ constraint. It can thus be solved using for instance a projected gradient descent

$$w_m^{(k+1)} = \frac{\tilde{w}_m^{(k)}}{\max(|\tilde{w}_m^{(k)}|, 1)} \quad \text{where} \quad \tilde{w}^{(k)} = w^{(k)} + \tau\nabla(f/\lambda + \operatorname{div}(w^{(k)})).$$

If the gradient step size satisfy $0 < \tau < 1/4$, one can prove that

$$f + \lambda\,\mathrm{div}(w^{(k)}) \longrightarrow f^\star_\lambda \quad \text{when} \quad k \to +\infty.$$

Bibliography

[1] Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, 2014.

[2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

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

[5] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[6] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[7] Philippe G Ciarlet. Introduction à l'analyse numérique matricielle et à l'optimisation. 1982.

[8] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.