Course 049064: Variational Methods in Image Processing

Notes 12: Numerical Methods II

Guy Gilboa

Motivation

Here we will just give a very preliminary intro to the topic of efficient convex optimization of nonlinear problems, which today is undergoing extensive research. These notes mainly follow the review paper of Parikh and Boyd [4].

1 Linear equations

We generally want to solve problems of the form:

$$\min_x f(x) + g(x)$$

or

$$\min_x f(x) \text{ s.t. } Ax - b = 0,$$

where A is $n \times n$ matrix and x and b are column vectors. Here we will show several popular and general methods, where a key ingredient is the use of proximal operators.

2 The Proximal Operator

The proximal operator $\text{prox}_{f,\lambda} : \mathbb{R}^n \to \mathbb{R}^n$ is

$$\text{prox}_{f,\lambda}(v) = \text{argmin}_x \left\{ f(x) + \frac{1}{2\lambda}\|x - v\|^2 \right\}, \tag{1}$$

where $\|\cdot\|$ is the standard Euclidean (ℓ^2) norm and $\lambda > 0$. This operator yields a compromise between minimizing f and being close to v (the larger

λ is, the closer it is to the minimizer of f). The Euler-Lagrange yields for $\lambda \to 0$ $\text{prox}_{f,\lambda}(v) \approx v - \lambda \nabla f$, therefore the operator can be interpreted as a step in a gradient descent.

2.1 Properties

Fixed points iterations of the proximal operator are minimizers of f. $\text{prox}_{f,\lambda}(x^*) = x^*$ if and only if x^* minimizes f.

If f is seperable across two vaiables, $f(x) = \varphi(x) + \psi(y)$ then

$$\text{prox}_{f,\lambda}(v, w) = (\text{prox}_{\varphi,\lambda}(v), \text{prox}_{\psi,\lambda}(w)).$$

Similarly, in the general case if we can write f as $f(x) = \sum_{i=1}^{n} f_i(x_i)$, then

$$(\text{prox}_{f,\lambda}(v))_i = \text{prox}_{f_i,\lambda}(v_i).$$

2.2 Examples of proximal functions

In general if f is a norm and B is the unit ball of the dual norm we have

$$\text{prox}_{f,\lambda}(v) = v - \lambda \Pi_B(v/\lambda),$$

where Π_B is a projection onto the ball B.

Let $f = \|\cdot\|_2$ be the Euclidean norm in $\mathbb{R}^n$ (reminder, the dual of ℓ^2 is ℓ^2). A projection Π_B onto the unit ball B is

$$\Pi_B(v) = \begin{cases} v/\|v\|_2 & , \|v\|_2 > 1 \\ v & , \|v\|_2 \leq 1. \end{cases} \tag{2}$$

The proximal operator is thus

$$\text{prox}_{f,\lambda}(v) = (1 - \lambda/\|v\|_2)^+ v = \begin{cases} (1 - \lambda/\|v\|_2)v & , \|v\|_2 > \lambda \\ 0 & , \|v\|_2 \leq \lambda. \end{cases} \tag{3}$$

For the ℓ^1 norm, $f = \|\cdot\|_1$, we get elementwise soft thresholding. The proximal operator is

$$\text{prox}_{f,\lambda}(v)_i = \begin{cases} v_i - \lambda & , v_i > \lambda \\ 0 & , |v_i| \leq \lambda \\ v_i + \lambda & , v_i < -\lambda. \end{cases} \tag{4}$$

3 Proximal Point Algorithms

3.1 ADMM

Consider the problem of finding x which minimizes

$$f(x) + g(x) \tag{5}$$

where f, g are (closed, proper) convex functionals from $\mathbb{R}^n \to \mathbb{R}$, which are not necessarily smooth.

The Alternating Direction Method of Multipliers (ADMM) to solve this problem is as follows. Initialize with $x^0 = z^0 = u^0 = 0$. Set $k = 0$, $\lambda = \lambda_0$. Do the following iterations until convergence:

1. $x^{k+1} = \text{prox}_{f,\lambda}(z^k - u^k)$.
2. $z^{k+1} = \text{prox}_{g,\lambda}(x^{k+1} + u^k)$.
3. $u^{k+1} = u^k + x^{k+1} - z^{k+1}$.

Here x^k and z^k converge to each other and to optimality.

A variant of the ADMM is very popular in denoising total-variation and similar functionals, referred to as the *Split-Bregman* algorithm by Goldstein and Osher [3], see a review on this and the relation to ADMM in [2].

3.2 FISTA

Initialize with $x^0 = y^0$, $t^1 = 1$. Set $k = 0$, L is the Lipschitz constant of ∇f. Do the following iterations until convergence:

1. $x^k = \text{argmin}_x \left\{ g(x) + \frac{L}{2}\|x - (y - \frac{1}{L}\nabla f(y))\|^2 \right\}$.
2. $t^{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$.
3. $y^{k+1} = x^k + \frac{t_k - 1}{t_{k+1}}(x_k - x^{k-1})$.

3.3 Chambolle-Pock

Given K a linear operater (such as a convolution blur kernel) the algorithm aims at minimizing

$$f(Kx) + g(x) \tag{6}$$

This is done by alternating between the primal and dual spaces. In the dual space the problem is to find y which *maximizes* the dual expression

$$-f^*(y) - g^*(-K^T y). \tag{7}$$

Initialize with $x^0 = y^0 = \bar{x}^0 = 0$, $\tau = \sigma = \theta = 1$. Set $k = 0$. Do the following iterations until convergence:

1. $y^{k+1} = \text{prox}_{f^*,\sigma}(y^k + \sigma K \bar{x}^k)$.
2. $x^{k+1} = \text{prox}_{g,\tau}(x^k - \tau K^T y^{k+1})$.
3. $\bar{x}^{k+1} = x^{k+1} + \theta(x^{k+1} - x^k)$.

Faster convergence are attained in accelerated versions (which are valid in a subset of the cases) where *theta* and the other parameters change every iteration:

$$.\theta^k = 1/\sqrt{1 + 2\gamma\tau^k},\ \tau^{k+1} = \theta^k \tau^k,\ \sigma^{k+1} = \sigma^k/\theta^k,$$

(with some $\gamma > 0$.).

The algorithm initiated by Pock-Cremers-Bischof-Chambolle for the Mumford Shah functional [5]. Later it was analyzed and generalized in [1].

References

[1] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[2] Ernie Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM report*, 9:31, 2009.

[3] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[4] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231, 2013.

[5] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the mumford-shah functional. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1133–1140. IEEE, 2009.