Structured Sparsity in Numerical Optimisation

Université Côte D'Azur, Inria, France

by

Matteo Frigo

Nice - December 4, 2017

- PhD Student
- Supervisor: Prof. Rachid Deriche
- Inria - Athena Project Team
- ERC Computational Brain Connectivity Mapping
- EDSTIC

Contents

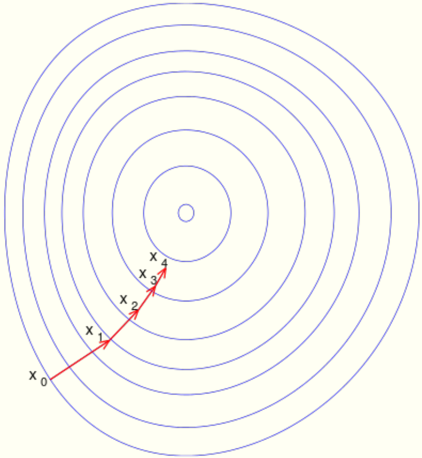
Mathematics

Numerical Optimisation

- $\mathcal{H}$ is a set
- $\Phi : \mathcal{H} \to \mathbb{R}$
- Find

$$x^* = \operatorname*{argmin}_{x \in \mathcal{H}} \Phi(x)$$

Smooth case



Smooth case: $g(x) = 0$

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^d} f(x).$$

Iteration:

$$x^+ = x - \gamma \nabla f(x)$$

with $\gamma = \frac{1}{L}$

Non-smooth case

Let $\{C_j\}_j$ sequence of cvx subsets of $\mathbb{R}^d$ with non-empty intersection,

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^d} \sum_j \iota_{C_j}(c)$$

Non-smooth case

Let $\{C_j\}_j$ sequence of cvx subsets of $\mathbb{R}^d$ with non-empty intersection,

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^d} \sum_j \iota_{C_j}(c)$$

POCS: Projection Onto Convex Sets

$$x_{k+1} = \Pi_{C_1} \Pi_{C_2} \cdots \Pi_{C_N} x_k.$$

Non-smooth case

Let $\{C_j\}_j$ sequence of cvx subsets of $\mathbb{R}^d$ with non-empty intersection,

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^d} \sum_j \iota_{C_j}(c)$$

POCS: Projection Onto Convex Sets

$$x_{k+1} = \Pi_{C_1} \Pi_{C_2} \cdots \Pi_{C_N} x_k.$$

Each projection is the solution of

$$\operatorname*{argmin}_{y \in \mathbb{R}^d} \frac{1}{2} \|x - y\|_2^2 + \iota_{C_j}(y).$$

Proximal operator

$\Gamma_0(\mathcal{H}) = \{f : \mathcal{H} \to \mathbb{R} \text{ l.s.c. and cvx with } \operatorname{dom}(f) \neq \emptyset\}$

Definition (Proximal Operator)

Let $f \in \Gamma_0(\mathbb{R}^N)$. For every $x \in \mathbb{R}^N$, the minimisation problem

$$\operatorname*{argmin}_{y \in \mathbb{R}^N} f(y) + \frac{1}{2} \|x - y\|_2^2$$

admits a unique solution which is called as $\operatorname{prox}_f(x)$.

Properties of prox

Separability:

$$f(x, y) = \varphi(x) + \psi(y) = \mathrm{prox}_{\varphi}(x) + \mathrm{prox}_{\psi}(y)$$

Nonexpansiveness:

$$\|\mathrm{prox}_f(x) - \mathrm{prox}_f(y)\|_2^2 \leq (x - y)^T (\mathrm{prox}_f(x) - \mathrm{prox}_f(y))$$

Resolvent operator:

$$\mathrm{prox}_f(\cdot) = (I + \partial f)^{-1}(\cdot)$$

Properties of prox: part 2

The convex conjugate of $f : X \to \mathbb{R}$ is $f^* : X^* \to \mathbb{R}$

$$f^*(\xi) = \sup_{x \in X} \langle \xi, x \rangle - f(x)$$

Properties of prox: part 2

The convex conjugate of $f : X \to \mathbb{R}$ is $f^* : X^* \to \mathbb{R}$

$$f^*(\xi) = \sup_{x \in X} \langle \xi, x \rangle - f(x)$$

Moreau decomposition:

$$\begin{aligned} v &= \mathrm{prox}_f(v) + \mathrm{prox}_{f^*}(v) \\ v &= \Pi_L(v) + \Pi_{L^\perp}(v) \\ v &= \Pi_K(v) + \Pi_{K^\circ}(v) \end{aligned}$$

Proof.

$2 + 2 = 4 - 1 = 3$ quick mafhs. □

Prox of norms

Consider $f = \|\cdot\|$ and $\mathcal{B} = \{x : \|x\|_* \leq 1\}$, then

$$f^*(\xi) = \iota_{C_j}\mathcal{B}(\xi)$$

By Moreau decomposition:

$$v = \mathrm{prox}_f(v) + \Pi_{\mathcal{B}}(v)$$
$$\mathrm{prox}_{\|\cdot\|}(v) = v - \Pi_{\mathcal{B}}(v)$$

Prox of norms

Consider $f = \|\cdot\|$ and $\mathcal{B} = \{x : \|x\|_* \leq 1\}$, then

$$f^*(\xi) = \iota_{C_j}\mathcal{B}(\xi)$$

By Moreau decomposition:

$$v = \mathrm{prox}_f(v) + \Pi_{\mathcal{B}}(v)$$
$$\mathrm{prox}_{\|\cdot\|}(v) = v - \Pi_{\mathcal{B}}(v)$$

We have the proximal operator of norms!

Key property

A point $x^* \in \mathbb{R}^n$ is a minimiser of f if and only if

$$\operatorname{prox}_f(x^*) = x^*. \tag{1}$$

Greedy algorithm

Objective: minimise f, for which we can compute prox_f

Greedy algorithm

Objective: minimise f, for which we can compute prox_f

- prox is firmly non-expansive

Greedy algorithm

Objective: minimise f, for which we can compute prox_f

- prox is firmly non-expansive
- The minimiser x^* is a fixed point of prox

Greedy algorithm

Objective: minimise f, for which we can compute prox_f

- prox is firmly non-expansive
- The minimiser x^* is a fixed point of prox

Iteration:

$$x_{k+1} = \mathrm{prox}_f\left(x_k\right)$$

Greedy algorithm

Objective: minimise f, for which we can compute prox_f

- prox is firmly non-expansive
- The minimiser x^* is a fixed point of prox

Iteration:

$$x_{k+1} = \operatorname{prox}_f(x_k)$$

If we don't know anything about the Lipschitz constant of prox_f

$$x_{k+1} = [(1-\alpha)I + \alpha \operatorname{prox}_f](x_k)$$

Cominetti et al., On the rate of convergence of Krasnoselskii-Mann iterations and their connection with sums of Bernoullis, 2014

Interpretation

Gradient descent of the Moreau envelope

$$M_f = \left(f^* + (1/2)\left\|\cdot\right\|_2^2\right)^*$$

Interpretation

Gradient descent of the Moreau envelope

$$M_f = \left(f^* + (1/2)\left\|\cdot\right\|_2^2\right)^*$$

NB: the minimisers of f and M_f coincide