Introduction to proximal mapping

3:06 PM

Should go to Convex Optimization Self-study Proximal Alg

Boyd Proximal Algorithms Introduction

[Boyd]_Introduction_...

*Introduction:

Page 1

Newton's method: base operations: low level: (linear algebra operations, compute $\nabla f(x)$, $\nabla^2 f(x)$)

Proximal algorithms: analogous tool to Newton's method (but has a higher level of abstraction)

solves non-smooth, constrained, large scale, distributed convex optimization problems

well suited for large data set

base operation: evaluate proximal operator of a function = solve a small optimization problem,

∵ often admit closed form solution ∨

simple specialized methods can be used to solve very quickly

- has interesting interpretations
- connected to many different topics in optimization + applied math

Page 2

$f: (\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\})$, CPC, extended value (closed, proper, convex)

f: CPC $\Leftrightarrow$ $\text{epi} f = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \le t\}$: (非ϕ, [ ], □) # Closed convex proper CPC function

$\text{dom} f = \{x \in \mathbb{R}^n \mid f(x) < \infty\}$ # set of points for which f takes on finite values)

{proximal operator of f at □} = $\text{prox}_f(□): \mathbb{R}^n \to \mathbb{R}^n = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2}\|x - □\|_2^2 \right)$

□ $\in \mathbb{R}^n$

i.e., □ doesn't have to belong to dom f

$f(x) + \frac{1}{2}\|x\|_2^2 - 2 \cdot \frac{1}{2} \cdot □^T x + \frac{1}{2}\|□\|_2^2$

$= \left(f(x) - □^T x + \frac{1}{2}\|□\|_2^2 \right) + \frac{1}{2}\|x\|_2^2$ → this is strongly convex # f: □$_{\text{strongly}} \Leftrightarrow \exists_m \left(f(x) - \frac{m}{2}\|x\|_2^2 \right)$: □

($f(x)$: □, $-□^T x$: affine, $\frac{1}{2}\|□\|_2^2$: constant, together: □; $\frac{1}{2}\|x\|_2^2$: eucledian norm squared)

$//\left(□ - \frac{1}{2}\|x\|_2^2\right)$: □ ($m=1$)

the argmin of a □$_{\text{strongly}}$ function will be unique $\forall_{□ \in \mathbb{R}^n}$ (even when $\text{dom} f \subsetneq \mathbb{R}^n$)

λ>0 [illegible] argmin will be the same

{proximal operator of scaled function λf: $\lambda > 0$} = $\text{prox}_{\lambda f}(□) = \underset{x}{\text{argmin}} \left(\lambda f(x) + \frac{1}{2}\|x - □\|_2^2 \right) = \underset{x}{\text{argmin}} \left(\lambda \left(f(x) + \frac{1}{2\lambda}\|x - □\|_2^2 \right) \right) = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2\lambda}\|x - □\|_2^2 \right)$

Code to generate function:

sublevel-sets

sublevel-sets Mathematica code

Page 3: Brief Interpretations:

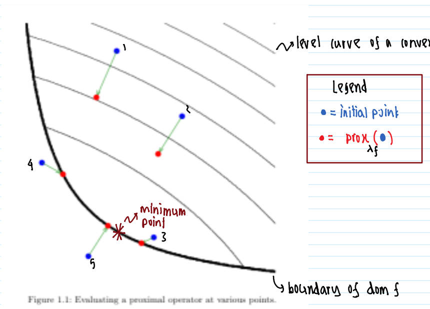
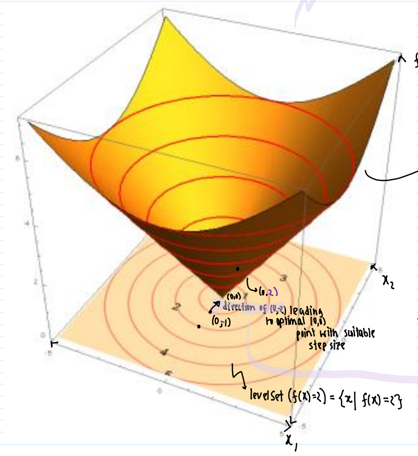


Figure 1.1: Evaluating a proximal operator at various points.



$f(x) = \|x\|_2^2$

$\text{dom} f = \{x \mid \|x\|_\infty \le 5\}$

* Some charecteristics of level curves of convex function:

* Out of curosity: How gradient descent will work at (0,-1)?

Ans: Because $f(x) = \|x\|^2 = x^T I x$ // $\nabla_x □ = 2x$

So, we go opposite to $2x|_{x=(0,-1)} = (0,-2)$ i.e. direction is (0, 2)

Clearly with proper step size we will head towards the optimal point (0,0)

if stepsize is not determined properly we may go beyond the optimal point

*Some charecteristics:

- points in the domain will stay in the domain, and move towards the optimal point (e.g. point 1,2,3)
- points outside the domain (4,5) will move to the boundary of the domain and towards the minimum of the function.
- λ controls the extent to which proximal operator maps to the optimal point
  - large → mapped points near the minimum
  - smaller → smaller movement near the mininum

∴ Analogous to stepsize in Newton's method

1.2

Basic interpretations of prox operator

* $\text{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2\lambda}\|x - v\|_2^2 \right)$

∵ comprises of minimizing $f(x)$ and (x being near to v)

$\lambda \sim 0 \to \text{prox}_{\lambda f}(v) \sim \underset{x}{\text{argmin}} \left(\frac{1}{2}\|x - v\|_2^2 \right)$

$\lambda \sim +\infty \to \text{prox}_{\lambda f}(v) \sim \underset{x}{\text{argmin}} \left(f(x) \right)$

*When

Page 4 f = Indicator function $= I_C(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases}$

(∵ [ ], □ set)

$\text{prox}_{\lambda I_C}(v) = \text{prox}_{\lambda f}(v) = \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2}\|x - v\|_2^2 \right) = \underset{x}{\text{argmin}} \left(I_C(x) + \frac{1}{2}\|x - v\|_2^2 \right)$

$= \underset{x}{\text{argmin}} \left(\left[\begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases} \right] + \frac{1}{2}\|x - v\|_2^2 \right) = \underset{x \in C}{\text{argmin}} \left(\frac{1}{2}\|x - v\|_2^2 \right) = \Pi_C(v)$ ∴ $\text{prox}_{\lambda I_C}(□) = \Pi_C(□)$ [$\text{prox}_{\lambda I_C}(□) = \Pi_C(□)$]

∴ So, proximal operator is sort of generalized projections

We can hope that various properties of projection that will be obeyed by proximal operator

* Under some assumptions,

$\text{prox}_{\lambda f}(v) \approx v - \lambda \nabla f(v)$ (λ: small; f: differentiable) # remember in steepest descent, $x^{(k+1)} = x^{(k)} - s_k \nabla f(x^{(k)})$ (s_k: step size)

this suggests: • close connections between proximal operators and gradient methods

Proximal map of \lambda I_C is the projection on C

- proximal operator may be useful in optimization
- λ : similar to stepsize

* $x^* = \text{prox}_{\lambda f}(x^*) \leftrightarrow x^* = \underset{x}{\text{argmin}}(f(x))$

this suggests • close connection between fixed point theory and proximal algorithm

- proximal algorithm can be interpreted as solving optimization problems by finding fixed points of appropriate operators

1.3. ★ Proximal Algorithms:
- Proximal gradient algorithm
- ADMM
- Parallel, distributed algorithm
- Evaluating it easily. → most useful when this happens

★ Why study proximal algorithm:
- extremely general (non-smooth, extended real values)
- fast
- amenable to distributed optimization
  very large scale "
- conceptually, mathematically simple

Proximal mapping = Resolvent of the subdifferential operator:

$$\forall_{x \in \text{dom} f} \quad \underset{\text{a point}}{x} \xrightarrow{\partial f} \underset{\text{a set}}{\partial f(x)}$$

So, ∂f is a mapping or relation

$$\partial f = \{ \bigcup_{x \in \text{dom} f} (x, \partial f(x)) \}$$

now: $\text{prox}_{\lambda f} = \underbrace{(I + \lambda \partial f)^{-1}}_{\text{resolvent of the operator } \partial f \text{ with operator}}$

$$\forall_{x \in \text{dom} f} \quad \text{prox}_{\lambda f}(x) = (I + \lambda \partial f)^{-1} x$$

thm: proximal operator is the resolvent of subdifferential operator

relation, sum, inverse } Operations on relations

however $(I + \lambda f)^{-1}$ {dom[·] = $\mathbb{R}^n$, single-valued}

Proof: f {subdifferentiable on its domain} 2017-04-06 10:38 PM

$\forall_{x \in \text{dom} f}$

$$(I + \lambda \partial f)^{-1}(x) = \{\boxdot_i\}$$

let

$$z \in (I + \lambda \partial f)^{-1}(x) \qquad \{\boxdot_i\}$$

$$\leftrightarrow \quad x \in (I + \lambda \partial f) z = z + \lambda \partial f(z) = \{ z + \lambda \boxdot_i \}_i$$

overloaded plus

$$\leftrightarrow \exists_{\boxdot_{i^*} \in \partial f(z)} \quad x = z + \lambda \boxdot_{i^*}$$

$$\cancel{\lambda} \boxdot_{i^*} = -\frac{1}{\lambda}(z - x) \quad [\because \lambda > 0]$$

$$\boxdot_{i^*} + \frac{1}{\lambda}(z - x) = 0$$

$$\leftrightarrow \exists_{\boxdot_{i^*} \in \partial f(z)} \quad \boxdot_{i^*} + \frac{1}{\lambda}(z - x) = 0$$

$$\leftrightarrow \quad 0 \in \underbrace{\partial f(z) + \frac{1}{\lambda}(z - x)}_{\partial_z \left(f(z) + \frac{1}{\lambda} \frac{1}{2} \|z - x\|_2^2 \right)} = \partial_z \left(f(z) + \frac{1}{\lambda}\frac{1}{2} \|z - x\|_2^2 \right) = \left[\partial_{ta} \left(f(ta) + \frac{1}{\lambda} \frac{1}{2} \| ta - x \|_2^2 \right) \right]_{ta = z}$$

\# remember: 1) $\underbrace{f(z)}_{\text{convex}} + \underbrace{\frac{1}{\lambda} \|z - x\|_2^2}_{\text{strongly convex}}$ is a strongly convex as $\triangledown + \triangledown_{\text{strongly}} = \triangledown_{\text{strongly}} \Rightarrow \triangledown_{\text{strict}}$

\#

\# so (·) has unique minimizer at z, as (·) is strictly convex

$$\leftrightarrow \quad z = \underset{ta}{\text{argmin}} \left[f(ta) + \frac{1}{\lambda} \frac{1}{2} \| ta - x \|_2^2 \right]$$

$$= \underset{ta}{\text{argmin}} \ \underset{\text{constant so can be dropped}}{\frac{1}{\lambda}} \left[\lambda f(ta) + \frac{1}{2} \| ta - x \|_2^2 \right]$$

$$= \underset{ta}{\text{argmin}} \left[\lambda f(ta) + \frac{1}{2} \| ta - x \|_2^2 \right]$$

$$= \text{prox}_{\lambda f}(x)$$

$\therefore$ $z \in (I + \lambda \partial f)^{-1}(x) \leftrightarrow z = \text{prox}_{\lambda f}(x)$ which is a singleton (so, $(I + \lambda \partial f)^{-1}$ is a function)

$\therefore \ \mathrm{prox}_{\lambda f}(\cdot) = (I + \lambda \partial f)^{-1}(\cdot)$ {single valued}

(proved)