Convex Optimization

Part 4: Proximal point method

Namhoon Lee

POSTECH

16 Nov 2022

Admin

Assignment 3

- grading still in progress

Assignment 4

- will be posted on PLMS this week

Proximal point method

an algorithm for minimizing a closed convex function f:

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{t_k f}(x_k) \\ &= \operatorname*{arg\,min}_u \left(f(u) + \frac{1}{2t_k}\|u - x_k\|_2^2 \right) \end{aligned}$$

- can be viewed as proximal gradient method with $g(x) = 0$
- of interest if prox evaluations are much easier than minimizing f directly
- in practice, inexact prox evaluations may be sufficient
- step size $t_k > 0$ affects number of iterations, cost of prox evaluations
- basis of the augmented Lagrangian method

Convergence

Assumptions

- f is closed and convex (hence, $\mathrm{prox}_{tf}(x)$ is uniqely defined for all x)
- optimal value $f^\star$ is finite and attained at $x^\star$

Result

$$f(x_k) - f^\star \leq \frac{\|x_0 - x^\star\|_2^2}{2\sum_{i=0}^{k-1} t_i} \qquad \text{for } k \geq 1$$

- implies convergence if $\sum_i t_i \to \infty$
- rate is $1/k$ if t_i is fixed, or variable but bounded away from zero
- t_i is arbitrary; however cost of prox evaluations will depend on t_i

Proof.
apply analysis of proximal gradient method with $g(x) = 0$; find the lemma for the bound on proximal gradient update □

Accelerated proximal point algorithms

- we take $g(x) = 0$ in FISTA:

$$
\begin{aligned}
x_1 &= \mathrm{prox}_{t_0 f}(x_0) \\
x_{k+1} &= \mathrm{prox}_{t_k f}\left(x_k + \theta_k(\frac{1}{\theta_{k-1}} - 1)(x_k - x_{k-1})\right) \quad \text{for } k \geq 1
\end{aligned}
$$

- choose any $t_k > 0$, determine θ_k from equation

$$
\frac{\theta_k^2}{t_k} = (1 - \theta_k)\frac{\theta_{k-1}^2}{t_{k-1}}
$$

- converges if $\sum_i \sqrt{t_i} \to \infty$
- rate is $1/k^2$ if t_i is fixed or variable but bounded away from zero

Standard problem format

Primal and dual problem

$$
\begin{array}{llc}
\text{primal:} & \text{minimize} & f(x) + g(Ax) \\
\text{dual:} & \text{maximize} & -g^*(z) - f^*(-A^\top z)
\end{array}
$$

Examples

- set constraints ($g(y) = \delta_C(y)$):

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & Ax \in C
\end{array}
$$

- regularized norm approximation ($g(y) = \|y - b\|$):

$$
\text{minimize} \quad f(x) + \|Ax - b\|
$$

Augmented Lagrangian method: proximal point method applied to the dual

Proximal mapping of dual function

Definition: proximal mapping of $h(z) = g^*(z) + f^*(-A^\top z)$ is defined as

$$\mathrm{prox}_{th}(z) = \underset{u}{\arg\min} \left(g^*(u) + f^*(-A^\top u) + \frac{1}{2t}\|u - z\|_2^2 \right)$$

Dual expression: $\mathrm{prox}_{th}(z) = z + t(A\hat{x} - \hat{y})$ where

$$(\hat{x}, \hat{y}) = \underset{x,y}{\arg\min} \left(f(x) + g(y) + z^\top (Ax - y) + \frac{t}{2}\|Ax - y\|_2^2 \right)$$

- $\hat{x}, \hat{y}$ minimize the augmented Lagrangian (Lagrangian + quadratic penalty)
- $f(x) + g(y) + z^\top (Ax - y)$ is Lagrangian of primal problem reformulated as

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax - y = 0 \end{array}$$

Proof.

- write augmented Lagrangian minimization as

$$\begin{array}{ll} \text{minimize (over } x, y, w) & f(x) + g(y) + \frac{t}{2}\|w\|_2^2 \\ \text{subject to} & Ax - y + z/t = w \end{array}$$

- optimality conditions (u is the multiplier for the equality constraint):

$$Ax - y + \frac{1}{t}z = w, \qquad -A^\top u \in \partial f(x), \qquad u \in \partial g(y), \qquad tw = u$$

- eliminating w gives

$$u = z + t(Ax - y), \qquad -A^\top u \in \partial f(x), \qquad u \in \partial g(y)$$

- eliminating x, y gives

$$0 \in \partial g^*(u) - A\partial f^*(-A^\top u) + \frac{1}{t}(u - z)$$

this is the optimality condition for the problem in the definition of $u = \operatorname{prox}_{th}(z)$

Augmented Lagrangian method

choose initial z_0 and repeat:

1. minimize augmented Lagrangian

$$(\hat{x}, \hat{y}) = \operatorname*{arg\,min}_{x,y} \left(f(x) + g(y) + \frac{t_k}{2} \|Ax - y + z_k/t_k\|_2^2 \right)$$

2. dual update

$$z_{k+1} = z_k + t_k(A\hat{x} - \hat{y})$$

- also known as method of multipliers
- this is the proximal point method applied to the dual problm
- as variants, can apply the accelerated proximal point methods to the dual
- usually implemented with inexact minimization step 1

Examples

$$\text{minimize } f(x) + g(Ax)$$

Equality constraints: g is indicator of $\{b\}$

$$\hat{x} = \arg\min_x \left(f(x) + \frac{t}{2}\|Ax - b + z/t\|_2^2 \right)$$
$$z := z + t(A\hat{x} - b)$$

Set constraint: g indicator of convex set C

$$\hat{x} = \arg\min_x \left(f(x) + \frac{t}{2} d(Ax + z/t)^2 \right)$$
$$z := z + t(A\hat{x} - \mathrm{P}_C(A\hat{x} + z/t))$$

- in step 1 on previous page, $\hat{y} = \mathrm{P}_C(A\hat{x} + z/t)$ where P_C is projection on C
- $d(u) = \|u - \mathrm{P}_C(u)\|_2$ is Euclidean distance of u to C

Moreau-Yosida smoothing

Definition: the Moreau-Yosida regularization of a closed convex function f is

$$
\begin{aligned}
f_{(t)}(x) &= \inf_u \left(f(u) + \frac{1}{2t}\|u-x\|_2^2 \right) \qquad \text{(with } t>0\text{)} \\
&= f\Big(\operatorname{prox}_{tf}(x)\Big) + \frac{1}{2t}\Big\|\operatorname{prox}_{tf}(x) - x\Big\|_2^2
\end{aligned}
$$

this is also known as the Moreau envelope of f

Immediate properties

- $f_{(t)}$ is convex (infimum over u of a convex function of x, u)
- domain of $f_{(t)}$ is $\mathbb{R}^n$ (recall that $\operatorname{prox}_{tf}(x)$ is defined for all x)

Examples

Indicator function: smoothed f is squared Euclidean distance

$$f(x) = \delta_C(x), \qquad f_{(t)}(x) = \frac{1}{2t} d(x)^2$$

1-norm: smoothed function is Huber penalty

$$f(x) = \|x\|_1, \qquad f_{(t)}(x) = \sum_{k=1}^{n} \phi_t(x_k)$$

$$\phi_t(z) = \begin{cases} z^2/(2t) & |z| \leq t \\ |z| - t/2 & |z| \geq t \end{cases}$$

Conjugate of Moreau envelope

$$f_{(t)}(x) = \inf_u \left(f(u) + \frac{1}{2t}\|u - x\|_2^2 \right)$$

- $f_{(t)}$ is infimal convolution of $f(u)$ and $\|v\|_2^2/(2t)$:

$$f_{(t)}(x) = \inf_{u+v=x} \left(f(u) + \frac{1}{2t}\|v\|_2^2 \right)$$

- conjugate is sum of conjugates of $f(u)$ and $\|v\|_2^2/(2t)$:

$$\left(f_{(t)}\right)^*(y) = f^*(y) + \frac{t}{2}\|y\|_2^2$$

- hence, conjugate is strongly convex with parameter t

Gradient of Moreau envelope

$$f_{(t)}(x) = \sup_y \left(x^\top y - f^*(y) - \frac{t}{2}\|y\|_2^2 \right)$$

- maximizer y in definition is unique and satisfies

$$\begin{aligned} x - ty \in \partial f^*(y) \quad &\Longleftrightarrow \quad y \in \partial f(x - ty) \\ &\Longleftrightarrow \quad y = \frac{1}{t}(x - \mathrm{prox}_{tf}(x)) \end{aligned}$$

- maximizer y is the gradient of $f_{(t)}$:

$$\nabla f_{(t)}(x) = \frac{1}{t}(x - \mathrm{prox}_{tf}(x)) = \mathrm{prox}_{t^{-1}f^*(x/t)}$$

  we applied the Moreau decomposition
- gradient $\nabla f_{(t)}$ is Lipschitz continuous with constant $1/t$

Interpretation of proximal point algorithm

apply gradient method to minimize Moreau envelope

$$\text{minimize} \quad f_{(t)}(x) = \inf_u \left(f(u) + \frac{1}{2t}\|u - x\|_2^2 \right)$$

this is an exact smooth reformulation of problem of minimizing $f(x)$:

- solution x is minimizer of f
- $f_{(t)}$ is differentiable with Lipschitz continuous gradient $(L = 1/t)$

Gradient update: with fixed $t_k = 1/L = t$

$$x_{k+1} = x_k - t\nabla f_{(t)}(x_k) = \mathrm{prox}_{tf}(x_k)$$

... the proximal point update with constant step size $t_k = t$

Interpretation of augmented Lagrangian algorithm

$$\text{minimize} \quad f(x) + g(Ax)$$

- augmented Lagrangian iteration is

$$(\hat{x}, \hat{y}) = \operatorname*{arg\,min}_{x,y} \left(f(x) + g(y) + \frac{t}{2}\|Ax - y + (1/t)z\|_2^2 \right)$$
$$z := z + t(A\hat{x} - \hat{y})$$

- with fixed t, dual update is gradient step applied to a smoothed dual
- after eliminating y, primal step can be written as

$$\hat{x} = \operatorname*{arg\,min}_{x} \left(f(x) + g_{(1/t)}(Ax + (1/t)z) \right)$$

- second term $g_{(1/t)}(Ax + (1/t)z)$ is a smooth approximation of $g(Ax)$
- adding the offset z/t allows us to use a fixed t

Example

$$\text{minimize} \quad f(x) + \|Ax - b\|_1$$

- augmented Lagrangian iteration is

$$(\hat{x}, \hat{y}) = \underset{x,y}{\operatorname{arg\,min}} \left(f(x) + \|y - b\|_1 + \frac{t}{2}\|Ax - y + (1/t)z\|_2^2 \right)$$
$$z := z + t(A\hat{x} - \hat{y})$$

- primal step after eliminating y: $\hat{x}$ is the solution of

$$\text{minimize} \quad f(x) + \phi_{1/t}(Ax - b + (1/t)z)$$

with $\phi_{1/t}$ the Huber penalty applied componentwise

Any questions?