The proximal mapping

http://bicmr.pku.edu.cn/~wenzw/opt-2018-fall.html

Acknowledgement: this slides is based on Prof. Lieven Vandenberghes lecture notes

Outline

Closed set

a set C is closed if it contains its boundary:

$$x^k \in C, \quad x^k \to \bar{x} \qquad \Longrightarrow \qquad \bar{x} \in C$$

operations that preserve closedness

- the intersection of (finitely or infinitely many) closed sets is closed
- the union of a finite number of closed sets is closed
- inverse under linear mapping: $\{x \mid Ax \in C\}$ is closed if C is closed

Image under linear mapping

the image of a closed set under a linear mapping is not necessarily closed

example(C is closed, $AC = \{Ax \mid x \in C\}$ is open):

$$C = \{(x_1, x_2) \in \mathbf{R}^2_+ \mid x_1 x_2 \geq 1\}, \quad A = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad AC = \mathbf{R}_{++}$$

sufficient condition: AC is closed if

- C is closed and convex
- and C does not have a recession direction in the nullspace of A. *i.e.*,

$$Ay = 0, \quad \hat{x} \in C, \quad \hat{x} + \alpha y \in C \quad \forall \alpha \geq 0 \quad \Longrightarrow \quad y = 0$$

in particular, this holds for any A if C is bounded

Closed function

definition: a function is closed if its epigraph is a closed set or if for each $\alpha \in \mathbb{R}$ the sublevel set $\{x \in \mathbf{dom} f \mid f(x) \leq \alpha\}$ is a closed set.

- If f is a continuous function and $\mathbf{dom} f$ is closed, then f is closed.
- If f is a continuous function and $\mathbf{dom} f$ is open, then f is closed iff it converges to ∞ along every sequence converging to a boundary point of $\mathbf{dom} f$

examples

- $f(x) = -\log(1-x^2)$ with $\mathbf{dom}\, f = \{x \mid |x| < 1\}$
- $f(x) = x \log x$ with $\mathbf{dom}\, f = \mathbf{R}_+$ and $f(0) = 0$
- indicator function of a closed set C: $f(x) = 0$ if $x \in C = \mathbf{dom}\, f$

not closed

- $f(x) = x \log x$ with $\mathbf{dom}\, f = \mathbf{R}_{++}$ or $\mathbf{dom}\, f = \mathbf{R}_+$ and $f(0) = 1$
- indicator function of a set C if C is not closed

Properties

sublevel sets: f is closed if and only if all its sublevel sets are closed

minimum: if f is closed with bounded sublevel sets then it has a minimizer

Weierstrass

Suppose that the set $D \subset E$ (a finite dimensional vector space over R^n) is nonempty and closed, and that all sublevel sets of the continuous function $f : D \to R$ are bounded. Then f has a global minimizer.

common operations on convex functions that preserve closedness

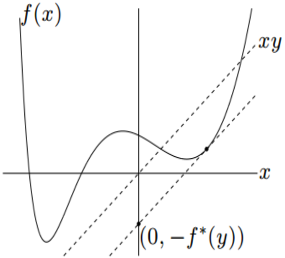
- sum: $f + g$ is closed if f and g are closed (and $\textbf{dom}\, f \cap \textbf{dom}\, g \neq \emptyset$)
- composition with affine mapping: $f(Ax + b)$ is closed if f is closed
- supremum: $\sup_\alpha f_\alpha(x)$ is closed if each function f_α is closed

Outline

Conjugate function

the **conjugate** of a function f is

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x))$$



f^* is closed and convex even if f is not

Fenchel's s inequality

$$f(x) + f^*(y) \geq x^T y \quad \forall x, y$$

(extends inequality $x^T x/2 + y^T y/2 \geq x^T y$ to non-quadratic convex f)

Quadratic function

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c$$

strictly convex case ($A \succ 0$)

$$f^*(y) = \frac{1}{2}(y-b)^T A^{-1}(y-b) - c$$

general convex case ($A \succeq 0$)

$$f^*(y) = \frac{1}{2}(y-b)^T A^{\dagger}(y-b) - c, \quad \mathbf{dom}\, f^* = range(A) + b$$

Negative entropy and negative logarithm

negative entropy

$$f(x) = \sum_{i=1}^{n} x_i \log x_i \qquad f^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$$

negative logarithm

$$f(x) = -\sum_{i=1}^{n} \log x_i \qquad f^*(y) = -\sum_{i=1}^{n} \log(-y_i) - n$$

matrix logarithm

$$f(x) = -\log\det X \quad (\mathbf{dom}\, f = \mathbf{S}^n_{++}) \quad f^*(Y) = -\log\det(-Y) - n$$

Indicator function and norm

indicator of convex set C: **conjugate** is support function of C

$$f(x) = \begin{cases} 0, & x \in C \\ +\infty, & x \notin C \end{cases} \qquad f^*(y) = \sup_{x \in C} y^T x$$

norm: conjugate is indicator of unit dual norm ball

$$f(x) = ||x|| \qquad f^*(y) = \begin{cases} 0, & ||y||_* \leq 1 \\ +\infty, & ||y||_* > 1 \end{cases}$$

(see next page)

proof: recall the definition of dual norm:

$$||y||_* = \sup_{||x|| \leq 1} x^T y$$

to evaluate $f^*(y) = \sup_x (y^T x - ||x||)$ we distinguish two cases

- if $||y||_* \leq 1$, then (by definition of dual norm)

  $$y^T x \leq ||x|| \quad \forall x$$

  and equality holds if $x = 0$; therefore $\sup_x (y^T x - ||x||) = 0$
- if $||y||_* > 1$, there exists an x with $||x|| \leq 1, x^T y > 1$; then

  $$f^*(y) \geq y^T (tx) - ||tx|| = t(y^T x - ||x||)$$

  and $r.h.s.$ goes to infinity if $t \to \infty$

The second conjugate

$$f^{**}(x) = \sup_{y \in \mathbf{dom}\, f^*} (x^T y - f^*(y))$$

- $f^{**}(x)$ is closed and convex
- from Fenchel's inequality $x^T y - f^*(y) \le f(x)$ for all y and x):

$$f^{**} \le f(x) \quad \forall x$$

  equivalently, epi $f \subseteq$ epi f^{**} (for any f)
- if f is closed and convex, then

$$f^{**}(x) = f(x) \quad \forall x$$

  equivalently, epi $f =$ epi f^{**} (if f is closed convex); proof on next page

proof ($f^{**} = f$ if f is closed and convex): by contradiction
suppose $(x, f^{**}(x)) \notin$ epi f; then there is a strict separating hyperplane:

$$\begin{bmatrix} \mathrm{a} \\ b \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \leq c \leq 0 \qquad \forall (z, s) \in \text{ epi } f$$

for some a, b, c with $b \leq 0$ ($b > 0$ gives a contradiction as $s \to \infty$)

- if $b < 0$, define $y = a/(-b)$ and maximize l.h.s. over $(z, s) \in$ epi f:

  $$f^*(y) - y^T x + f^{**}(x) \leq c/(-b) < 0$$

  this contradicts **Fenchel's** inequality

- if $b = 0$, choose $\hat{y} \in \mathbf{dom}\, f^*$ and add small multiple of $(\hat{y}, -1)$ to (a, b):

  $$\begin{bmatrix} a + \epsilon \hat{y} \\ -\epsilon \end{bmatrix}^T \begin{bmatrix} z - x \\ s - f^{**}(x) \end{bmatrix} \leq c + \epsilon(f^*(\hat{y}) - x^T \hat{y} + f^{**}(x)) < 0$$

  now apply the argument for $b < 0$

Conjugates and subgradients

if f is closed and convex, then

$$y \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(y) \quad \Leftrightarrow \quad x^T y = f(x) + f^*(y)$$

proof: if $y \in \partial f(x)$, then $f^*(y) = \sup_u (y^T u - f(u)) = y^T x - f(x)$

$$\begin{aligned} f^*(v) &= \sup_u (v^T u - f(u)) \\ &\geq v^T x - f(x) \\ &= x^T(v-y) - f(x) + y^T x \\ &= f^*(y) + x^T(v-y) \end{aligned} \tag{1}$$

for all v; therefore, x is a subgradient of f^* at y ($x \in \partial f^*(y)$)
reverse implication $x \in \partial f^*(y) \Rightarrow y \in \partial f(x)$ follows from $f^{**} = f$

Some calculus rules

separable sum

$$f(x_1, x_2) = g(x_1) + h(x_2) \qquad f^*(y_1, y_2) = g^*(y_1) + h^*(y_2)$$

scalar multiplication: (for $\alpha > 0$)

$$f(x) = \alpha g(x) \qquad f^*(y) = \alpha g^*(y/\alpha)$$

addition to affine function

$$f(x) = g(x) + a^T x + b \qquad f^*(y) = g^*(y - a) - b$$

infimal convolution

$$f(x) = \inf_{u+v=x} (g(u) + h(v)) \qquad f^*(y) = g^*(y) + h^*(y)$$

Outline

Proximal mapping

$$prox_f(x) = \operatorname*{argmin}_u \left(f(u) + \frac{1}{2}\|u - x\|_2^2 \right)$$

if f is closed and convex then $prox_f(x)$ **exists** and is **unique** for all x

- existence: $f(u) + (1/2)\|u - x\|_2^2$ is closed with bounded sublevel sets
- uniqueness: $f(u) + (1/2)\|u - x\|_2^2$ is strictly (in fact, strongly) convex

subgradient characterization

$$u = prox_f(x) \quad \Longleftrightarrow \quad x - u \in \partial f(u)$$

we are interested in functions f for which prox_{tf} is inexpensive

Examples

quadratic function ($A \succeq 0$)

$$f(x) = \frac{1}{2}x^T Ax + b^T x + c, \qquad \mathrm{prox}_{tf}(x) = (I + tA)^{-1}(x - tb)$$

Euclidean norm: $f(x) = \|x\|_2$

$$\mathrm{prox}_{tf}(x) = \begin{cases} (1 - t/\|x\|_2)x & \|x\|_2 \geq t \\ 0 & otherwise \end{cases}$$

logarithmic barrier

$$f(x) = -\sum_{i=1}^{n} \log x_i, \qquad \mathrm{prox}_{tf}(x)_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}, \quad i = 1, \ldots, n$$

Some simple calculus rules

separable sum

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = g(x) + h(y), \qquad \mathrm{prox}_f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \mathrm{prox}_g(x) \\ \mathrm{prox}_h(y) \end{bmatrix}$$

scaling and translation of argument : with $\lambda \neq 0$

$$f(x) = g(\lambda x + a), \qquad \mathrm{prox}_f(x) = \frac{1}{\lambda}(\mathrm{prox}_{\lambda^2 g}(\lambda x + a) - a)$$

scaling and translation of argument : with $\lambda > 0$

$$f(x) = \lambda g(x/\lambda), \qquad \mathrm{prox}_f = \lambda \mathrm{prox}_{\lambda^{-1} g}(x/\lambda)$$

Addition to linear or quadratic function

linear function

$$f(x) = g(x) + a^T x, \qquad \mathrm{prox}_f = \mathrm{prox}_g(x - a)$$

quadratic function: with $u > 0$

$$f(x) = g(x) + \frac{u}{2}\|x - a\|_2^2, \qquad \mathrm{prox}_f(x) = \mathrm{prox}_{\theta g}(\theta x + (1 - \theta)a),$$

where $\theta = 1/(1 + u)$

Moreau decomposition

$$x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) \qquad \forall x$$

- follows from properties of conjugates and subgradients:

$$\begin{aligned} u = prox_f(x) &\iff x - u \in \partial f(u) \\ &\iff u \in \partial f^*(x-u) \\ &\iff x - u = prox_{f^*}(x) \end{aligned}$$

- generalizes decomposition by orthogonal projection on subspaces:

$$x = P_L(x) + P_{L^\perp}(x)$$

  if L is a subspace, $L^\perp$ its orthogonal complement (this is Moreau decomposition with $f = I_L$, $f^* = I_{L^\perp}$)

Extended Moreau decomposition

for $\lambda > 0$

$$x = prox_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f*}(x/\lambda) \qquad \forall x$$

proof: apply Moreau decomposition to λf

$$\begin{aligned} x &= \text{prox}_{\lambda f}(x) + \text{prox}_{(\lambda f)*}(x) \\ &= \text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f*}(x/\lambda) \end{aligned}$$

second line uses $(\lambda f)^*(y) = \lambda f^*(y/\lambda)$

Composition with affine mapping

- for general A, the prox-operator of

$$f(x) = g(Ax + b)$$

does not follow easily from the prox-operator of g

- however if $AA^T = (1/\alpha)I$, we have

$$prox_f(x) = (I - \alpha A^T A)x + \alpha A^T(\mathrm{prox}_{\alpha^{-1}g}(Ax + b) - b)$$

example: $f(x_1, \ldots, x_m) = g(x_1 + x_2 + \cdots + x_m)$

$$prox_f(x_1, \ldots, x_m)_i = x_i - \frac{1}{m}\left(\sum_{j=1}^{m} x_j - \mathrm{prox}_{mg}(\sum_{j=1}^{m} x_j)\right)$$

proof: $u = prox_f(x)$ is the solution of the optimization problem

$$\begin{aligned} \min_{u,y} \ & g(y) + \frac{1}{2}\|u - x\|_2^2 \\ \text{s.t.} \ & Au + b = y \end{aligned}$$

with variables u, y

- eliminate u using the expression

$$\begin{aligned} u &= x + A^T(AA^T)^{-1}(y - b - Ax) \\ &= (I - \alpha A^T A)x + \alpha A^T(y - b) \end{aligned}$$

- optimal y is minimizer of

$$g(y) + \frac{\alpha^2}{2}\|A^T(y - b - Ax)\|_2^2 = g(y) + \frac{\alpha}{2}\|y - b - Ax\|_2^2$$

  solution is $y = prox_{\alpha^{-1}g}(Ax + b)$

Projection on affine sets

hyperplane: $C = \{x | a^T x = b\}$ (with $a \neq 0$)

$$P_C(x) = x + \frac{b - a^T x}{\|a\|_2^2} a$$

affine set: $C = \{x | Ax = b\}$ (with $A \in \mathbb{R}^{p \times n}$ and $\mathrm{rank}(A) = p$)

$$P_C(x) = x + A^T (AA^T)^{-1} (b - Ax)$$

inexpensive if $p \ll n, or \quad AA^T = I, \ldots$

Projection on simple polyhedral sets

halfspace: $C = \{x | a^T x \leq b\}$(with $a \neq 0$)

$$P_C(x) = \begin{cases} x + \frac{b - a^T x}{\|a\|_2^2} a & \textit{if } a^T x > b \\ x & \textit{if } a^T x \leq b \end{cases}$$

rectangle: $C = [l, u] = \{l \preceq x \preceq u\}$

$$P_C(x)_i = \begin{cases} l_i & x_i \leq l_i \\ x_i & l_i \leq x_i \leq u_i \\ u_i & x_i \geq u_i \end{cases}$$

nonnegative orthant: $C = \mathbf{R}^n_+$

$$P_C(x) = x_+ \qquad (x_+ \text{ is componentwise maximum of 0 and } x)$$

probability simplex: $C = \{x | 1^T x = 1, x \geq 0\}$

$$P_C(x) = (x - \lambda 1)_+$$

where λ is the solution of the equation

$$1^T (x - \lambda 1)_+ = \sum_{i=1}^{n} \max\{0, x_k - \lambda\} = 1$$

probability simplex: $C = \{x | a^T x = b, l \leq x \leq u\}$

$$P_c(x) = P_{[l,u]}(x - \lambda a)$$

where λ is the solution of

$$a^T P_{[l,u]}(x - \lambda a) = b$$

Projection on norm balls

Euclidean ball: $C = \{x | \|x\|_2 \leq 1\}$

$$P_C(x) = \begin{cases} \frac{1}{\|x\|_2} x & \textit{if } \|x\|_2 > 1 \\ x & \textit{if } \|x\|_2 \leq 1 \end{cases}$$

1-norm ball: $C = \{x | \|x\|_1 \leq 1\}$

$$P_c(x)_k = \begin{cases} x_k - \lambda, & x_k > \lambda \\ 0, & -\lambda \leq \ x_k \leq \lambda \\ x_k + \lambda, & x_k < -\lambda \end{cases}$$

$\lambda = 0$ if $\|x\|_1 \leq 1$; otherwise λ is the solution of the equation

$$\sum_{k=1}^{n} \max\{|x_k| - \lambda, 0\} = 1$$

Projection on simple cones

second order cone $C = \{(x,t) \in \mathbf{R}^{n\times 1} | \|x\|_2 \le t\}$

$$P_C(x,t) = (x,t) \quad \textit{if } \|x\|_2 \le t, \qquad P_C(x,t) = (0,0) \quad \textit{if } \|x\|_2 \le -t$$

and

$$P_C(x,t) = \frac{t + \|x\|_2}{2\|x\|_2} \begin{bmatrix} x \\ \|x\|_2 \end{bmatrix} \quad \textit{if } \|x\|_2 > t,\ x \neq 0$$

positive semidefinite cone $C = \mathbf{S}^n_+$

$$P_C(X) = \sum_{i=1}^{n} \max\{0, \lambda_i\} q_i q_i^T$$

if $X = \sum_{i=1}^n \lambda_i q_i q_i^T$ is the eigenvalue decomposition of X

Support function

conjugate of support function of closed convex set is indicator function

$$f(x) = S_C(x) = sup_{y \in C} x^T y, \qquad f^*(y) = I_C(y)$$

prox-operator of support function: apply Moreau decomposition

$$\begin{aligned} \mathrm{prox}_{tf} &= x - t\,\mathrm{prox}_{t^{-1}f^*}(x/t) \\ &= x - tP_C(x/t) \end{aligned}$$

example: $f(x)$ is sum of largest r components of x

$$f(x) = x_{[1]} + \cdots + x_{[r]} = S_C(x), \qquad C = \{y | 0 \le y \le 1, 1^T y = r\}$$

prox-operator of f is easily evaluated via projection on C

Norms

conjugate of norm is indicator function of dual norm ball:

$$f(x) = \|x\|, \qquad f^*(x) = I_B(y) \qquad (B = \{y \mid \|y\|_* \leq 1\})$$

prox-operator of norm: apply Moreau decomposition

$$\begin{aligned} \mathrm{prox}_{tf} &= x - t\mathrm{prox}_{t^{-1}f^*}(x/t) \\ &= x - tP_B(x/t) \\ &= x - P_{tB}(x) \end{aligned}$$

useful formula for $\mathrm{prox}_{t\|\cdot\|}$ when projection on $tB = \{x \mid \|x\| \leq t\}$ is cheap

examples: $\|\cdot\|_1$, $\|\cdot\|_2$

Distance to a point

distance (in general norm)

$$f(x) = \|x - a\|$$

prox-operator: from page 20, with $g(x) = \|x\|$

$$\begin{aligned}
\mathrm{prox}_{tf} &= a + \mathrm{prox}_{tg}(x - a) \\
&= a + x - a - tP_B(\frac{x - a}{t}) \\
&= x - P_{tB}(x - a)
\end{aligned}$$

B is the unit ball for the dual norm $\|\cdot\|_*$

Euclidean distance to a set

Euclidean distance (to a closed convex set C)

$$d(x) = \inf_{y \in C} \|x - y\|_2$$

prox-operator of distance

$$\mathrm{prox}_{td}(x) = \theta P_C(x) + (1-\theta)x, \qquad \theta = \begin{cases} t/d(x) & d(x) \geq t \\ 1 & \textit{otherwise} \end{cases}$$

prox-operator of squared distance: $f(x) = d(x)^2/2$

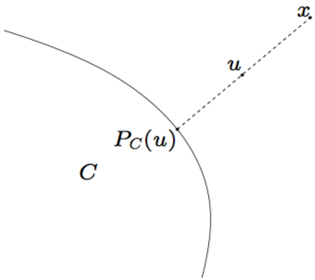
$$\mathrm{prox}_{tf} = \frac{1}{1+t}x + \frac{t}{1+t}P_C(x)$$

proof (expression for $\text{prox}_{td}(x)$)

- if $u = \text{prox}_{td}(x) \notin C$, then from the definition and subgradient for d

$$x - u = \frac{t}{d(u)}(u - P_C(u))$$

  implies $P_C(u) = P_C(x)$, $d(x) \geq t$, u is convex combination of x, $P_C(x)$



- if $u \in C$ minimizes $d(u) + (1/(2t))\|u - x\|_2^2$, then $u = P_C(x)$

proof (expression for $\mathrm{prox}_{tf}(x)$ when $f(x) = d(x)^2/2$)

$$\begin{aligned}\mathrm{prox}_{tf}(x) &= \arg\min_u \left(\frac{1}{2}d(u)^2 + \frac{1}{2t}\|u-x\|_2^2\right) \\ &= \arg\min_u \inf_{v\in C}\left(\frac{1}{2}\|u-v\|_2^2 + \frac{1}{2t}\|u-x\|_2^2\right)\end{aligned}$$

optimal u as a function of v is

$$u = \frac{t}{t+1}v + \frac{1}{t+1}x$$

optimal v minimizes

$$\frac{1}{2}\left\|\frac{t}{t+1}v + \frac{1}{t+1}x - v\right\|_2^2 + \frac{1}{2t}\left\|\frac{t}{t+1}v + \frac{1}{t+1}x - x\right\|_2^2 = \frac{1}{2(1+t)}\|v-x\|_2^2$$

Over C, *i.e.*, $v = P_C(x)$

Refrences

- P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting, Multiscale Modeling and Simulation (2005).*
- *P. L. Combettes and J.-Ch. Pesquet,* Proximal splitting methods in signal processsing, Fixed-Point Algorithms for Inverse Problems in Science and Engineering (2011).
- N. Parikh and S. Boyd, *Proximal algorithms (2013).*