6. Proximal mapping

- introduction
- review of conjugate functions
- proximal mapping

Proximal mapping

the proximal mapping (prox-operator) of a convex function h is

$$\mathbf{prox}_h(x) = \underset{u}{\operatorname{argmin}} \left(h(u) + \frac{1}{2}\|u - x\|_2^2 \right)$$

examples

- $h(x) = 0$: $\mathbf{prox}_h(x) = x$
- $h(x) = I_C(x)$ (indicator function of C): $\mathbf{prox}_h$ is projection on C

$$\mathbf{prox}_h(x) = \underset{u \in C}{\operatorname{argmin}} \|u - x\|_2^2 = P_C(x)$$

- $h(x) = t\|x\|_1$: $\mathbf{prox}_h$ is the 'soft-threshold' (shrinkage) operation

$$\mathbf{prox}_h(x)_i = \begin{cases} x_i - t & x_i \geq t \\ 0 & |x_i| \leq t \\ x_i + t & x_i \leq -t \end{cases}$$

Proximal gradient method

unconstrained problem with cost function split in two components

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

- g convex, differentiable, with $\mathbf{dom}\, g = \mathbf{R}^n$
- h convex, possibly nondifferentiable, with inexpensive prox-operator

proximal gradient algorithm

$$x^{(k)} = \mathbf{prox}_{t_k h}\left(x^{(k-1)} - t_k \nabla g(x^{(k-1)})\right)$$

$t_k > 0$ is step size, constant or determined by line search

Interpretation

$$x^+ = \mathbf{prox}_{th}\left(x - t\nabla g(x)\right)$$

from definition of proximal operator:

$$\begin{aligned} x^+ &= \operatorname*{argmin}_u \left(h(u) + \frac{1}{2t}\|u - x + t\nabla g(x)\|_2^2 \right) \\ &= \operatorname*{argmin}_u \left(h(u) + g(x) + \nabla g(x)^T(u - x) + \frac{1}{2t}\|u - x\|_2^2 \right) \end{aligned}$$

x^+ minimizes $h(u)$ plus a simple quadratic local model of $g(u)$ around x

Examples

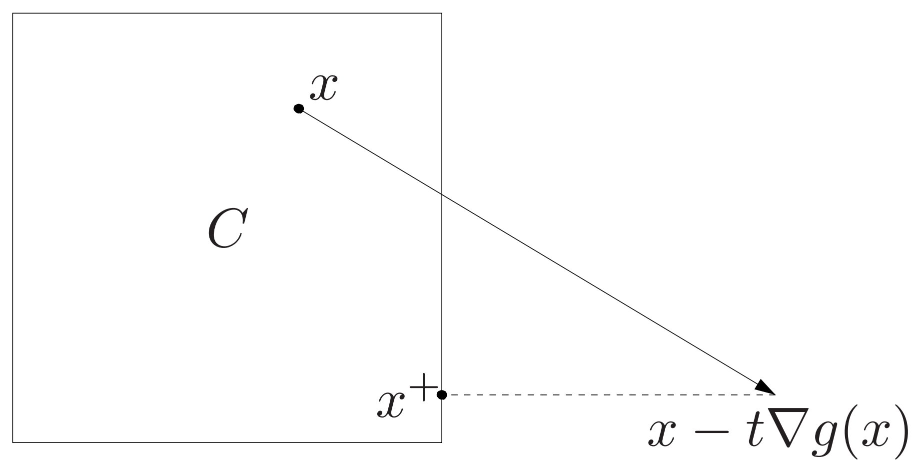
$$\text{minimize} \quad g(x) + h(x)$$

gradient method: $h(x) = 0$, *i.e.*, minimize $g(x)$

$$x^+ = x - t\nabla g(x)$$

gradient projection method: $h(x) = I_C(x)$, *i.e.*, minimize $g(x)$ over C

$$x^+ = P_C\left(x - t\nabla g(x)\right)$$

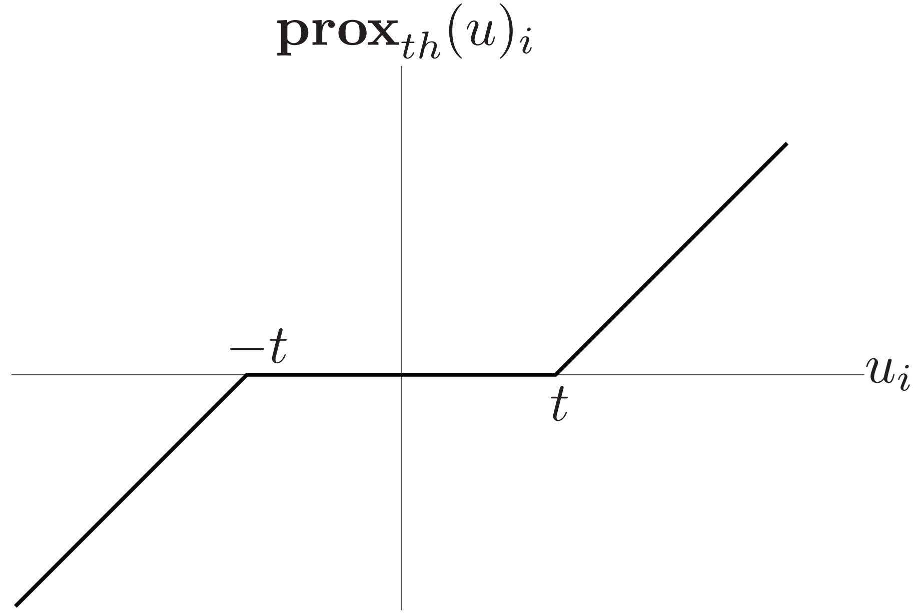


soft-thresholding: $h(x) = \|x\|_1$, *i.e.*, minimize $g(x) + \|x\|_1$

$$x^+ = \mathbf{prox}_{th}\left(x - t\nabla g(x)\right)$$

and

$$\mathbf{prox}_{th}(u)_i = \begin{cases} u_i - t & u_i \geq t \\ 0 & -t \leq u_i \leq t \\ u_i + t & u_i \leq -t \end{cases}$$



more on proximal algorithms in next lecture. . .

Outline

- introduction
- **review of conjugate functions**
- proximal mapping

Closed convex function

a function with a closed convex epigraph

examples

$$f(x) = |x|, \qquad f(x) = -\log(1-x^2), \qquad f(x) = \begin{cases} x\log x & x > 0 \\ 0 & x = 0 \\ +\infty & x < 0 \end{cases}$$

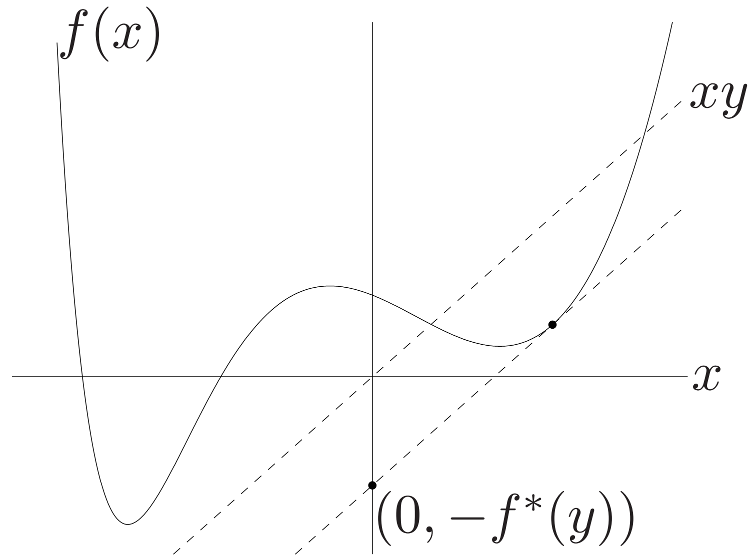
convex functions that are not closed

$$f(x) = \begin{cases} |x| & |x| < 1 \\ +\infty & |x| \geq 1 \end{cases}, \qquad f(x) = \begin{cases} x\log x & x > 0 \\ 1 & x = 0 \\ +\infty & x < 0 \end{cases}$$

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 inequality

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

Examples

negative logarithm $f(x) = -\log x$

$$\begin{aligned} f^*(y) &= \sup_{x>0} (xy + \log x) \\ &= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

quadratic function $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}^n_{++}$

$$\begin{aligned} f^*(y) &= \sup_x (y^T x - \frac{1}{2} x^T Q x) \\ &= \frac{1}{2} y^T Q^{-1} y \end{aligned}$$

Conjugate of closed functions

for f closed and convex

- $f^{**} = f$
- $y \in \partial f(x)$ if and only if $x \in \partial f^*(y)$

  proof sketch: if $\hat{y} \in \partial f(\hat{x})$, then $f(z) \geq f(\hat{x}) + \hat{y}^T(z - \hat{x})$, $\forall z$, or

  $$z^T\hat{y} - f(z) \leq \hat{x}^T\hat{y} - f(\hat{x}), \quad \forall z$$

  so $z^T\hat{y} - f(z)$ reaches its maximum over z at $\hat{x}$; therefore $\hat{x} \in \partial f^*(\hat{y})$; this shows

  $$\hat{y} \in \partial f(\hat{x}) \quad \Longrightarrow \quad \hat{x} \in \partial f^*(\hat{y})$$

  reverse implication follows from $f^{**} = f$

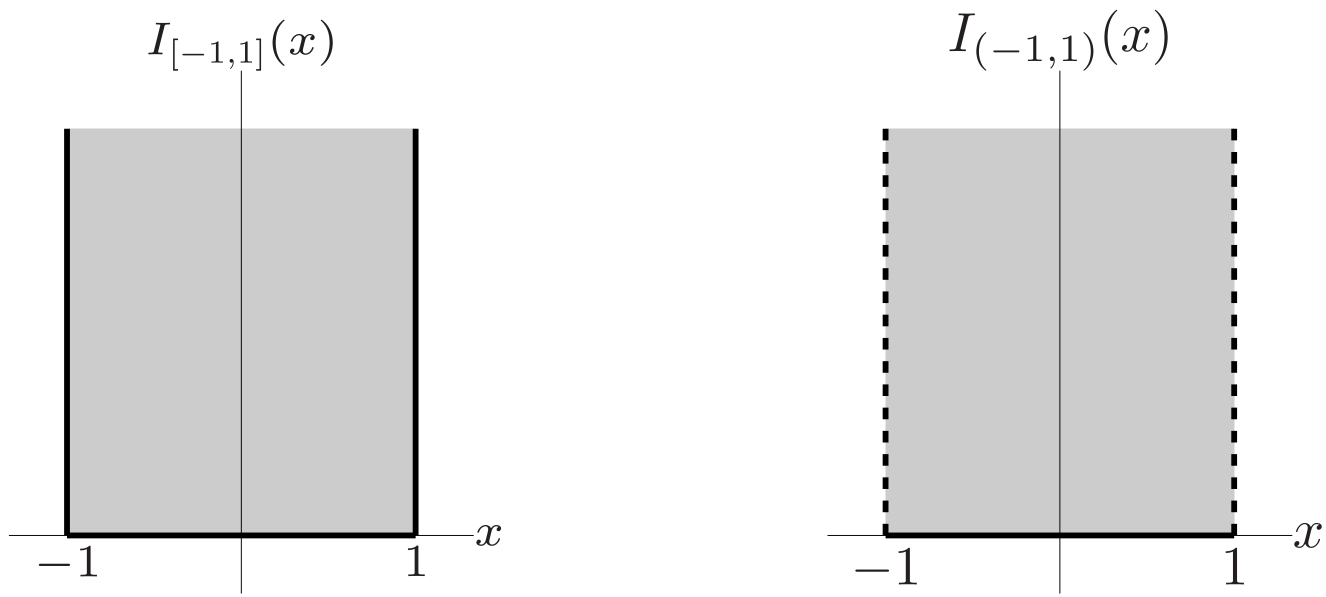
hence, for closed convex f the following three statements are equivalent:

$$x^Ty = f(x) + f^*(y) \quad \Longleftrightarrow \quad y \in \partial f(x) \quad \Longleftrightarrow \quad x \in \partial f^*(y)$$

Indicator function

the indicator function of a set C is

$$I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & \text{otherwise} \end{cases}$$

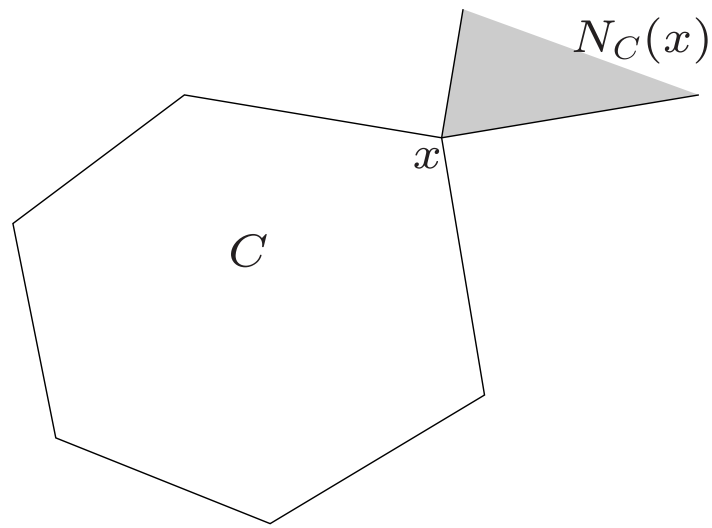


the indicator function of a (closed) convex set is a (closed) convex function

Subgradients of indicator function

subdifferential of $I_C(x)$ is the *normal cone* to C at x (notation: $N_C(x)$)

$$\partial I_C(x) = N_C(x) = \{s \mid s^T(y-x) \leq 0 \text{ for all } y \in C\}$$



Dual norm

the *dual norm* of a norm $\|\cdot\|$ is

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

(the support function of the unit ball for $\|\cdot\|$)

- common pairs of dual vector norms

$$\|x\|_2,\ \|y\|_2, \qquad \|x\|_1,\ \|y\|_\infty, \qquad \sqrt{x^TQx},\ \sqrt{y^TQ^{-1}y} \quad (Q \succ 0)$$

- common pairs of dual matrix norms (for inner product $\mathbf{tr}(X^TY)$)

$$\|X\|_F,\ \|Y\|_F, \qquad \|X\|_2 = \sigma_{\max}(X),\ \ \|Y\|_* = \sum_i \sigma_i(Y)$$

Conjugate of norm

conjugate of $f = \|x\|$ is the indicator function of the dual unit norm ball

$$f^*(y) = \sup_x \left(y^T x - \|x\|\right) = \begin{cases} 0 & \|y\|_* \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

proof

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

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

  and equality holds if $x = 0$; therefore $\sup(y^T x - \|x\|) = 0$

- if $\|y\|_* > 1$, there exists an x with $\|x\| \le 1$, $y^T x > 1$; then

$$f^*(y) \ge y^T(tx) - \|tx\| = t(y^T x - \|x\|) \to \infty \text{ if } t \to \infty$$

Outline

- introduction
- review of conjugate functions
- **proximal mapping**

Proximal mapping

definition: proximal mapping associated with closed convex h

$$\mathbf{prox}_h(x) = \underset{u}{\operatorname{argmin}} \left(h(u) + \frac{1}{2}\|u - x\|_2^2 \right)$$

it can be shown that $\mathbf{prox}_h(x)$ exists and is unique for all x

subgradient characterization

from optimality conditions of minimization in the definition:

$$u = \mathbf{prox}_h(x) \quad \Longleftrightarrow \quad x - u \in \partial h(u)$$

Projection

proximal mapping of indicator function I_C is the Euclidean projection on C

$$\mathbf{prox}_{I_C}(x) = \underset{u \in C}{\operatorname{argmin}} \|u - x\|_2^2 = P_C(x)$$

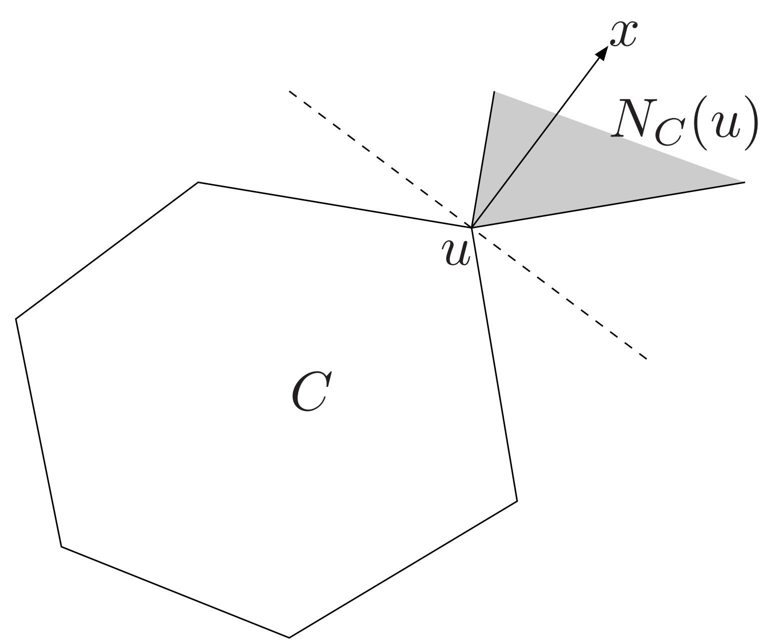
subgradient characterization

$u = P_C(x)$ satisfies

$$x - u \in \partial I_C(u) = N_C(u)$$

in other words,

$$(x - u)^T (y - u) \leq 0 \quad \forall y \in C$$



we will see that proximal mappings have many properties of projections

Nonexpansiveness

if $u = \mathbf{prox}_h(x)$, $v = \mathbf{prox}_h(y)$, then

$$(u - v)^T (x - y) \geq \|u - v\|_2^2$$

- follows from characterization of p.6–15 and monotonicity (p.4-8)

$$x - u \in \partial h(u), \quad y - v \in \partial h(v) \quad \Longrightarrow \quad (x - u - y + v)^T (u - v) \geq 0$$

- implies (from Cauchy-Schwarz inequality)

$$\|\mathbf{prox}_h(x) - \mathbf{prox}_h(y)\|_2 \leq \|x - y\|_2$$

$\mathbf{prox}_h$ is *nonexpansive* or *Lipschitz continuous* with constant 1

Moreau decomposition

$$\mathbf{prox}_{h^*}(x) = x - \mathbf{prox}_h(x)$$

proof: define $u = \mathbf{prox}_h(x)$, $v = x - u$

- from subgradient characterization on p. 6–15: $v \in \partial h(u)$
- hence (from p. 6–10), $u \in \partial h^*(v)$
- therefore (again from p. 6–15), $v = \mathbf{prox}_{h^*}(x)$

interpretation: decomposition of x in two components

$$x = \mathbf{prox}_h(x) + \mathbf{prox}_{h^*}(x)$$

example: $h(u) = I_L(u)$, L a subspace of $\mathbf{R}^n$

- conjugate is the indicator function of the orthogonal complement $L^\perp$

$$\begin{aligned} h^*(v) = \sup_{u \in L} v^T u &= \begin{cases} 0 & v \in L^\perp \\ +\infty & \text{otherwise} \end{cases} \\ &= I_{L^\perp}(v) \end{aligned}$$

- Moreau decomposition is orthogonal decomposition

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

Projection on affine sets

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

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

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

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

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

Projection on simple polyhedral sets

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

$$P_C(x) = x + \frac{b - a^T x}{\|a\|_2^2} a \quad \text{if } a^T x > b, \qquad P_C(x) = x \quad \text{if } a^T x \leq b$$

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

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

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

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

probability simplex: $C = \{x \mid \mathbf{1}^T x = 1, x \succeq 0\}$

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

where λ is the solution of the equation

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

intersection of hyperplane and rectangle: $C = \{x | a^T x = b, l \preceq x \preceq u\}$

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

where λ is the solution of

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

Projection on norm balls

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

$$P_C(x) = \frac{1}{\|x\|_2} x \quad \text{if } \|x\|_2 > 1, \qquad P_C(x) = x \quad \text{if } \|x\|_2 \le 1$$

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

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

$\lambda = 0$ if $\|x\|_1 \le 1$; 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 $C = \{(x,t) \in \mathbf{R}^{n\times 1} \mid \|x\|_2 \leq t\}$

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

and

$$P_C(x,t) = \frac{t + \|x\|_2}{2\|x\|_2} \left[\begin{array}{c} x \\ \|x\|_2 \end{array} \right] \quad \text{if } -t < \|x\|_2 < t,\ x \neq 0$$

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

Other examples of proximal mappings

quadratic function

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

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

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

logarithmic barrier

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

Some calculus rules

separable sum: $h(x_1, x_2) = h_1(x_1) + h_2(x_2)$

$$\mathbf{prox}_h(x_1, x_2) = \big(\mathbf{prox}_{h_1}(x_1), \mathbf{prox}_{h_2}(x_2)\big)$$

scaling and translation of argument: $h(x) = f(\lambda x + a)$ with $\lambda \neq 0$

$$\mathbf{prox}_h(x) = \frac{1}{\lambda}\big(\mathbf{prox}_{\lambda^2 f}(\lambda x + a) - a\big)$$

prox-operator of conjugate: for $t > 0$,

$$\mathbf{prox}_{th^*}(x) = x - t\,\mathbf{prox}_{h/t}(x/t)$$

proof

- the conjugate of $g(u) = h(u)/t$ is

$$g^*(v) = \sup_u \left(v^T u - h(u)/t\right) = \frac{1}{t}\sup_u \left((tv)^T u - h(u)\right) = \frac{1}{t}h^*(tv)$$

- by the Moreau decomposition and the second property on page 6–26

$$\mathbf{prox}_g(y) = y - \mathbf{prox}_{g^*}(y) = y - \frac{1}{t}\mathbf{prox}_{th^*}(ty)$$

- a change of variables $x = ty$ gives

$$\mathbf{prox}_{h/t}(x/t) = \frac{1}{t}x - \frac{1}{t}\mathbf{prox}_{th^*}(x)$$

Norms

$$h(x) = \|x\|, \qquad h^*(y) = I_B(y)$$

where $B = \{y \mid \|y\|_* \leq 1\}$, the unit norm ball for the dual norm (p.6–28)

prox-operator: from page 6–27,

$$\begin{aligned} \mathbf{prox}_{th}(x) &= x - t\,\mathbf{prox}_{h^*/t}(x/t) \\ &= x - tP_B(x/t) \end{aligned}$$

a useful formula for $\mathbf{prox}_{th}$ when the projection P_B is inexpensive

examples

- $h(x) = \|x\|_1$, $B = \{y \mid \|y\|_\infty \leq 1\}$ (gives soft-threshold operator; p.6–2)
- $h(x) = \|x\|_2$, $B = \{y \mid \|y\|_2 \leq 1\}$ (gives formula of page 6–25)

Distance to a point

distance in general norm

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

prox-operator: from p.6–26, with $g(x) = \|x\|$

$$\begin{aligned}
\mathbf{prox}_{th}(x) &= a + \mathbf{prox}_g(x - a) \\
&= a + x - a - tP_B(\frac{x - a}{t}) \\
&= x - tP_B\left(\frac{x - a}{t}\right)
\end{aligned}$$

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

Euclidean distance to a set

Euclidean distance (C is a closed convex set)

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

prox-operator

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

prox-operator of squared distance: $h(x) = \mathbf{dist}(x)^2/2$

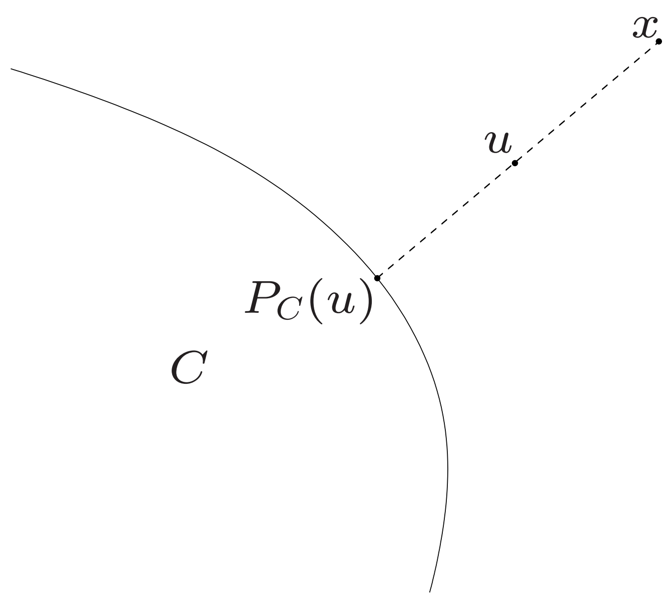
$$\mathbf{prox}_{th}(x) = \frac{1}{1+t}x + \frac{t}{1+t}P_C(x)$$

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

- if $u = \mathbf{prox}_{td}(x) \notin C$, then from page 6–15 and page 5-15

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

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 of expression for $\mathbf{prox}_{th}(x)$ when $h(x) = d(x)^2/2$

$$\begin{aligned}
\mathbf{prox}_{th}(x) &= \underset{u}{\operatorname{argmin}} \left(\frac{1}{2} d(u)^2 + \frac{1}{2t} \|u - x\|_2^2 \right) \\
&= \underset{u}{\operatorname{argmin}} \inf_{v \in C} \left(\frac{1}{2} \|u - v\|_2^2 + \frac{1}{2t} \|u - x\|_2^2 \right)
\end{aligned}$$

optimal u as a function of v is

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

optimal v minimizes

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

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

References

- this lecture is a modified version of: L. Vandenberghe, *Lecture notes for EE236C - Optimization Methods for Large-Scale Systems* (Spring 2011), UCLA.

- 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*, `arxiv.org/abs/0912.3522v4` (2010)