9. The proximal mapping

- proximal mapping
- projections
- support functions, norms, distances

Proximal mapping

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

if f is closed and convex then $\operatorname{prox}_f(x)$ **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 (from page 6-7):

$$u = \operatorname{prox}_f(x) \qquad \Longleftrightarrow \qquad x - u \in \partial f(u)$$

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

Examples

quadratic function $(A \succeq 0)$

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

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

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

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

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

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

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

'right' scalar multiplication: with $\lambda > 0$,

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

Addition to linear or quadratic function

linear function

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

quadratic function: with $\mu > 0$

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

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

Moreau decomposition

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

- follows from properties of conjugates and subgradients:

$$\begin{aligned} u = \mathrm{prox}_f(x) \quad &\Longleftrightarrow \quad x - u \in \partial f(u) \\ &\Longleftrightarrow \quad u \in \partial f^*(x - u) \\ &\Longleftrightarrow \quad x - u = \mathrm{prox}_{f^*}(x) \end{aligned}$$

- 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 $\lambda > 0$,

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

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 $(\lambda f)^*(y) = \lambda f^*(y/\lambda)$ and expression on page 9-4

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

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

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

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

proof: $u = \operatorname{prox}_f(x)$ is the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & g(y) + \frac{1}{2}\|u - x\|_2^2 \\ \text{subject to} & Au + b = y \end{array}$$

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 $y = \operatorname{prox}_{\alpha^{-1}g}(Ax + b)$

Outline

- conjugate functions
- proximal mapping
- **projections**
- support functions, norms, distances

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

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

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: $C = \{x \mid a^T x = b, l \preceq x \preceq u\}$

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

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

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

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

Outline

- conjugate functions
- proximal mapping
- projections
- **support functions, norms, distances**

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

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

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

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

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

Norms

conjugate of norm is indicator function of dual norm ball:

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

prox-operator of norm: apply Moreau decomposition

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

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

examples: for $\|\cdot\|_1$, $\|\cdot\|_2$, get expressions on pages 6-2 and 9-3

Distance to a point

distance (in general norm)

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

prox-operator: from page 9-4, with $g(x) = \|x\|$

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

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

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

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

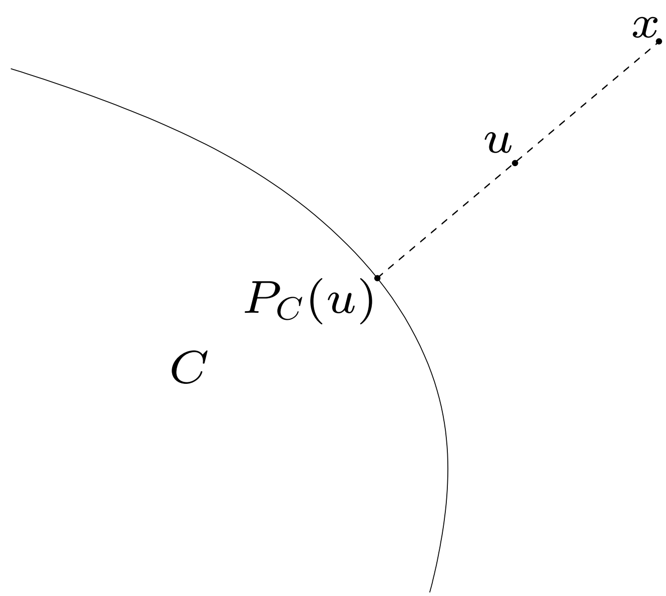
$$\mathrm{prox}_{tf}(x) = \frac{1}{1+t}x + \frac{t}{1+t}P_C(x)$$

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

- if $u = \mathrm{prox}_{td}(x) \notin C$, then from p.9-2 and subgradient for d (p. 4-21)

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

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 $f(x) = d(x)^2/2$)

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

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

- 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).