6. The proximal mapping

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

Proximal mapping

Definition: the proximal mapping of a closed convex function f is

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

Existence and uniqueness: we minimize a closed and strongly convex function

$$g(u) = f(u) + \frac{1}{2}\|u - x\|_2^2$$

- minimizer exists because g is closed with bounded sublevel sets
- minimizer is unique because g is strictly convex

Subgradient characterization (from page 4.7):

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

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

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

Logarithmic barrier

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

Simple calculus rules

Separable sum

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

Scaling and translation of argument: for scalar $a \neq 0$,

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

"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

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

Quadratic function: with $\mu > 0$

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

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

Moreau decomposition

$$x = \mathrm{prox}_f(x) + \mathrm{prox}_{f^*}(x) \quad \text{for all } x$$

- 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 the Moreau decomposition with $f = \delta_L$, $f^* = \delta_{L^\perp}$)

Extended Moreau decomposition

for $\lambda > 0$,

$$x = \text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1} f^*}(x/\lambda) \quad \text{for all } x$$

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 6.4

Composition with affine mapping

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

- for general A, prox-operator of f does not follow easily from prox-operator of g
- however, if $AA^T = (1/\alpha)I$, then

$$\begin{aligned}\operatorname{prox}_f(x) &= (I - \alpha A^T A)x + \alpha A^T(\operatorname{prox}_{\alpha^{-1}g}(Ax + b) - b) \\ &= x - \alpha A^T(Ax + b - \operatorname{prox}_{\alpha^{-1}g}(Ax + b))\end{aligned}$$

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

- write as $f(x) = g(Ax)$ with $A = \begin{bmatrix} I & I & \cdots & I \end{bmatrix}$
- since $AA^T = mI$, we get

$$\operatorname{prox}_f(x_1, \ldots, x_m)_i = x_i - \frac{1}{m}\sum_{j=1}^m x_j + \frac{1}{m}\operatorname{prox}_{mg}(\sum_{j=1}^m x_j), \quad i = 1, \ldots, m$$

Proof: $u = \mathrm{prox}_f(x)$ 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

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

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

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: $C = [l, u] = \{x \in \mathbf{R}^n \mid l \leq x \leq u\}$

$$P_C(x)_k = \begin{cases} l_k & x_k \leq l_k \\ x_k & l_k \leq x_k \leq u_k \\ u_k & x_k \geq u_k \end{cases}$$

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

$$P_C(x) = x_+ = (\max\{0, x_1\}, \max\{0, x_2\}, \ldots, \max\{0, x_n\})$$

Projection on simple polyhedral sets

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 the equation

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

Proof (probability simplex): projection $y = P_C(x)$ solves the optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - x\|_2^2 + \delta_{\mathbf{R}^n_+}(y) \\ \text{subject to} & \mathbf{1}^T y = 1 \end{array}$$

optimality conditions are:

- y minimizes the Lagrangian

$$\begin{aligned} & \frac{1}{2}\|y - x\|_2^2 + \delta_{\mathbf{R}^n_+}(y) + \lambda(\mathbf{1}^T y - 1) \\ = \quad & \sum_{k=1}^{n} \left(\frac{1}{2}(y_k - x_k)^2 + \delta_{\mathbf{R}_+}(y_k) + \lambda y_k \right) - \lambda \end{aligned}$$

  this is a separable function with minimizer $y_k = (x_k - \lambda)_+$ for $k = 1, \ldots, n$

- primal feasibility: requires

$$\sum_{k=1}^{n} y_i = \sum_{k=1}^{n} (x_k - \lambda)_+ = 1$$

Proof (rectangle and hyperplane): $y = P_C(x)$ solves optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - x\|_2^2 + \delta_{[l,u]}(y) \\ \text{subject to} & a^T y = b \end{array}$$

optimality conditions are:

- y minimizes the Lagrangian

$$\begin{aligned} & \frac{1}{2}\|y - x\|_2^2 + \delta_{[l,u]}(y) + \lambda(a^T y - b) \\ &= \sum_{k=1}^{n} \left(\frac{1}{2}(y_k - x_k)^2 + \delta_{[l_k,u_k]}(y_k) + \lambda a_k y_k \right) - \lambda b \end{aligned}$$

  the minimizer is $y_k = P_{[l_k,u_k]}(x_k - \lambda a_k)$ for $k = 1, \ldots, n$

- primal feasibility: requires

$$a^T y = \sum_{k=1}^{n} a_k P_{[l_k,u_k]}(x_k - \lambda a_k) = b$$

Projection on norm balls

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

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

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

projection is $P_C(x) = x$ if $\|x\|_1 \leq 1$; otherwise

$$P_C(x)_k = \operatorname{sign}(x_k) \max\{|x_k| - \lambda, 0\} = \begin{cases} x_k - \lambda & x_k > \lambda \\ 0 & -\lambda \leq x_k \leq \lambda \\ x_k + \lambda & x_k < -\lambda \end{cases}$$

where λ is the solution of the equation

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

Proof (1-norm): projection $y = P_C(x)$ solves the optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|y - x\|_2^2 \\ \text{subject to} & \|y\|_1 \leq 1 \end{array}$$

optimality conditions are:

- y minimizes the Lagrangian

$$\frac{1}{2}\|y - x\|_2^2 + \lambda(\|y\|_1 - \lambda) = \sum_{k=1}^{n} \left(\frac{1}{2}(y_k - x_k)^2 + \lambda|y_k|\right) - \lambda$$

  the minimizer y is obtained by componentwise soft-thresholding:

$$y_k = \text{sign}(x_k) \max\{|x_k| - \lambda, 0\}, \quad k = 1, \ldots, n$$

- primal, dual feasibility and complementary slackness:

$$\lambda = 0, \quad \|y\|_1 = \|x\|_1 \leq 1 \qquad \text{or} \qquad \lambda > 0, \quad \|y\|_1 = \sum_{k=1}^{n} \max\{|x_k| - \lambda, 0\} = 1$$

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

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

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

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

- prox-operator of support function follows from 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

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

prox-operator of f is easily evaluated via projection on C (page 6.12)

Norms

- conjugate of norm is indicator function of dual norm ball:

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

- prox-operator of norm follows from Moreau decomposition

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

- gives $\operatorname{prox}_{t\|\cdot\|}$ when projection on $tB = \{x \mid \|x\|_* \leq t\}$ is cheap

Examples: for $\|\cdot\|_1$, $\|\cdot\|_2$, get expressions on pages 4.2 and 6.3

Distance to a point

Distance (in general norm)

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

Prox-operator: from page 6.4, 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

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

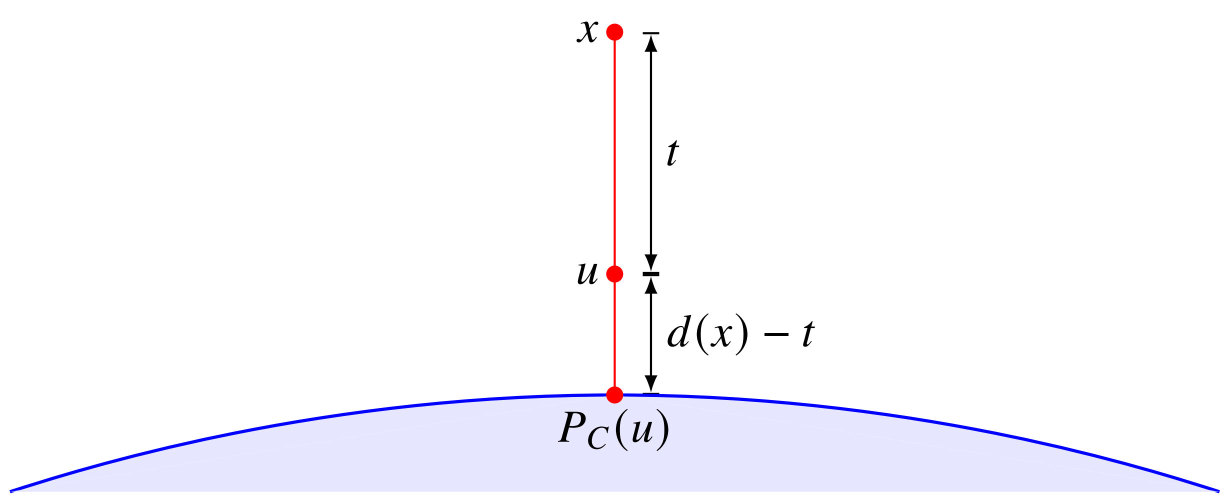
Prox-operator of squared distance: $f(x) = d(x)^2/2$

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

Proof (expression for $\operatorname{prox}_{td}(x)$):

- if $u = \operatorname{prox}_{td}(x) \notin C$, then from page 6.2 and subgradient for d (page 2.20)

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



- if $\operatorname{prox}_{td}(x) \in C$ then the minimizer of

$$d(u) + \frac{1}{2t}\|u - x\|_2^2$$

satisfies $d(u) = 0$ and must be the projection $P_C(x)$

Proof (expression for $\operatorname{prox}_{tf}(x)$ when $f(x) = d(x)^2/2$):

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

- optimal u as a function of v is

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

- optimal v minimizes

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

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

References

- A. Beck, *First-Order Methods in Optimization* (2017), chapter 6.
- P. L. Combettes and J.-Ch. Pesquet, *Proximal splitting methods in signal processing*, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (2011).
- N. Parikh and S. Boyd, *Proximal algorithms* (2013).