7. Conjugate functions

- closed functions
- conjugate function
- duality

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

$$C = \{(x_1, x_2) \in \mathbf{R}^2_+ \mid x_1 x_2 \geq 1\}, \qquad A = \left[\begin{array}{cc} 1 & 0 \end{array}\right], \qquad AC = \mathbf{R}_{++}$$

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.*,

$$Ay = 0, \quad \hat{x} \in C, \quad \hat{x} + \alpha y \in C \text{ for all } \alpha \geq 0 \qquad \Longrightarrow \qquad y = 0$$

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

Closed function

Definition: a function is closed if its epigraph is a closed set

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:

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

Not closed

- $f(x) = x \log x$ with $\mathbf{dom}\, f = \mathbf{R}_{++}$, or with $\mathbf{dom}\, f = \mathbf{R}_+$ 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

Common operations on convex functions that preserve closedness

- *sum:* $f = f_1 + f_2$ is closed if f_1 and f_2 are closed
- *composition with affine mapping:* $f = g(Ax + b)$ is closed if g is closed
- *supremum:* $f(x) = \sup_\alpha f_\alpha(x)$ is closed if each function f_α is closed

in each case, we assume $\mathrm{dom}\, f \neq \emptyset$

Outline

- closed functions
- **conjugate function**
- duality

Conjugate function

the **conjugate** of a function f is

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

f^* is closed and convex (even when f is not)

Fenchel's inequality: the definition implies that

$$f(x) + f^*(y) \geq x^T y \quad \text{for all } x, y$$

this is an extension to non-quadratic convex f of the inequality

$$\frac{1}{2}x^T x + \frac{1}{2}y^T y \geq x^T y$$

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 ($A \succeq 0$)

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

Negative entropy and negative logarithm

Negative entropy

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

Negative logarithm

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

Matrix logarithm

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

Indicator function and norm

Indicator of convex set C: conjugate is the *support function* of C

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

Indicator of convex cone C: conjugate is indicator of polar (negative dual) cone

$$\delta_C^*(y) = \delta_{-C^*}(y) = \delta_{C^*}(-y) = \begin{cases} 0 & y^T x \leq 0 \ \ \forall x \in C \\ +\infty & \text{otherwise} \end{cases}$$

Norm: conjugate is indicator of unit ball for dual norm

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

(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)

  $$y^T x \leq \|x\| \quad \text{for all } x$$

  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 right-hand side goes to infinity if $t \to \infty$

Calculus rules

Separable sum

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

Scalar multiplication ($\alpha > 0$)

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

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

- the operation $f(x) = \alpha g(x/\alpha)$ is sometimes called 'right scalar multiplication'
- a convenient notation is $f = g\alpha$ for the function $(g\alpha)(x) = \alpha g(x/\alpha)$
- conjugates can be written concisely as $(g\alpha)^* = \alpha g^*$ and $(\alpha g)^* = g^*\alpha$

Calculus rules

Addition to affine function

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

Translation of argument

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

Composition with invertible linear mapping (A square and nonsingular)

$$f(x) = g(Ax) \qquad f^*(y) = g^*(A^{-T} y)$$

Infimal convolution

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

The second conjugate

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

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

$$f^{**}(x) \leq f(x) \quad \text{for all } x$$

  equivalently, $\mathbf{epi}\, f \subseteq \mathbf{epi}\, f^{**}$ (for any f)

- if f is closed and convex, then

$$f^{**}(x) = f(x) \quad \text{for all } x$$

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

Proof (by contradiction): assume f is closed and convex, and $\mathbf{epi}\, f^{**} \neq \mathbf{epi}\, f$

suppose $(x, f^{**}(x)) \notin \mathbf{epi}\, f$; then there is a strict separating hyperplane:

$$\left[\begin{array}{c} a \\ b \end{array}\right]^T \left[\begin{array}{c} z - x \\ s - f^{**}(x) \end{array}\right] \leq c < 0 \qquad \forall (z, s) \in \mathbf{epi}\, f$$

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 left-hand side over $(z, s) \in \mathbf{epi}\, f$:

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

  this contradicts Fenchel's inequality

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

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

  now apply the argument for $b < 0$

Conjugates and subgradients

if f is closed and convex, then

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

Proof. if $y \in \partial f(x)$, then $f^*(y) = \sup_u (y^T u - f(u)) = y^T x - f(x)$; hence

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

this holds for all v; therefore, $x \in \partial f^*(y)$

reverse implication $x \in \partial f^*(y) \Longrightarrow y \in \partial f(x)$ follows from $f^{**} = f$

Conjugate of strongly convex function

assume f is closed and strongly convex with parameter $\mu > 0$

- f^* is defined for all y (*i.e.*, $\mathrm{dom}\, f^* = \mathbf{R}^n$)

- f^* is differentiable everywhere, with gradient

$$\nabla f^*(y) = \underset{x}{\operatorname{argmax}}\, (y^T x - f(x))$$

- ∇f^* is Lipschitz continuous with constant $1/\mu$

$$\|\nabla f^*(y) - \nabla f^*(y')\|_2 \leq \frac{1}{\mu}\|y - y'\|_2 \quad \text{for all } y \text{ and } y'$$

Proof: if f is strongly convex and closed

- $y^T x - f(x)$ has a unique maximizer x for every y

- x maximizes $y^T x - f(x)$ if and only if $y \in \partial f(x)$; from page 7-15

$$y \in \partial f(x) \qquad \Longleftrightarrow \qquad x \in \partial f^*(y) = \{\nabla f^*(y)\}$$

  hence $\nabla f^*(y) = \operatorname{argmax}_x (y^T x - f(x))$

- from convexity of $f(x) - (\mu/2)x^T x$:

$$(y - y')^T (x - x') \geq \mu \|x - x'\|_2^2 \qquad \text{if } y \in \partial f(x),\ y' \in \partial f(x')$$

- this is co-coercivity of ∇f^* (which implies Lipschitz continuity)

$$(y - y')^T (\nabla f^*(y) - \nabla f^*(y')) \geq \mu \|\nabla f^*(y) - \nabla f^*(y')\|_2^2$$

Outline

- closed functions
- conjugate function
- **duality**

Duality

primal:	minimize	$f(x) + g(Ax)$
dual:	maximize	$-g^*(z) - f^*(-A^T z)$

- follows from Lagrange duality applied to reformulated primal

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

dual function for the formulated problem is:

$$\inf_{x,y} \left(f(x) + z^T Ax + g(y) - z^T y\right) = -f^*(-A^T z) - g^*(z)$$

- Slater's condition (for convex f, g): strong duality holds if there exists an $\hat{x}$ with

$$\hat{x} \in \mathbf{int}\,\mathbf{dom}\, f, \qquad A\hat{x} \in \mathbf{int}\,\mathbf{dom}\, g$$

this also guarantees that the dual optimum is attained, if optimal value is finite

Set constraint

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

Primal and dual problem

$$\begin{array}{lll} \text{primal:} & \text{minimize} & f(x) + \delta_C(Ax - b) \\ \text{dual:} & \text{maximize} & -b^T z - \delta_C^*(z) - f^*(-A^T z) \end{array}$$

Examples

	constraint	set C	support function $\delta_C^*(z)$
equality	$Ax = b$	$\{0\}$	0
norm inequality	$\|Ax - b\| \leq 1$	unit $\|\cdot\|$-ball	$\|z\|_*$
conic inequality	$Ax \preceq_K b$	$-K$	$\delta_{K^*}(z)$

Norm regularization

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

- take $g(y) = \|y - b\|$ in general problem

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

- conjugate of $\|\cdot\|$ is indicator of unit ball for dual norm

$$g^*(z) = b^T z + \delta_B(z) \quad \text{where } B = \{z \mid \|z\|_* \leq 1\}$$

- hence, dual problem can be written as

$$\begin{array}{ll} \text{maximize} & -b^T z - f^*(-A^T z) \\ \text{subject to} & \|z\|_* \leq 1 \end{array}$$

Optimality conditions

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

assume f, g are convex and Slater's condition holds

Optimality conditions: x is optimal if and only if there exists a z such that

1. primal feasibility: $x \in \mathbf{dom}\, f$ and $y = Ax \in \mathbf{dom}\, g$
2. x and $y = Ax$ are minimizers of the Lagrangian $f(x) + z^T Ax + g(y) - z^T y$:

$$-A^T z \in \partial f(x), \qquad z \in \partial g(Ax)$$

if g is closed, this can be written symmetrically as

$$-A^T z \in \partial f(x), \qquad Ax \in \partial g^*(z)$$

References

- J.-B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algoritms* (1993), chapter X.
- D.P. Bertsekas, A. Nedić, A.E. Ozdaglar, *Convex Analysis and Optimization* (2003), chapter 7.
- R. T. Rockafellar, *Convex Analysis* (1970).