Duality correspondences

Geoff Gordon & Ryan Tibshirani
Optimization 10-725 / 36-725

Remember KKT conditions

Recall that for the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0, \;\; i = 1, \ldots m \\ & \ell_j(x) = 0, \;\; j = 1, \ldots r \end{aligned}$$

the **KKT conditions** are

- $0 \in \partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_i \partial \ell_j(x)$ (stationarity)
- $u_i \cdot h_i(x) = 0$ for all i (complementary slackness)
- $h_i(x) \leq 0, \; \ell_j(x) = 0$ for all i, j (primal feasibility)
- $u_i \geq 0$ for all i (dual feasibility)

These are necessary for optimality (of a primal-dual pair $x^\star$ and $u^\star, v^\star$) under strong duality, and sufficient for convex problems

Remember solving the primal via the dual

An important consequence of stationarity: under strong duality, given a dual solution $u^\star, v^\star$, any primal solution $x^\star$ solves

$$\min_{x \in \mathbb{R}^n} \; f(x) + \sum_{i=1}^m u_i^\star h_i(x) + \sum_{j=1}^r v_i^\star \ell_j(x)$$

Often, solutions of this unconstrained problem can be expressed explicitly, giving an explicit **characterization** of primal solutions (from dual solutions)

Furthermore, suppose the solution of this problem is unique; then it must be the primal solution $x^\star$

This can be very helpful when the dual is easier to solve than the primal

Consider as an example (from B & V page 249):

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} f_i(x_i) \;\text{ subject to }\; a^T x = b$$

where each $f_i : \mathbb{R} \to \mathbb{R}$ is a strictly convex function. Dual function:

$$\begin{aligned} g(v) &= \min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} f_i(x_i) + v(b - a^T x) \\ &= bv + \sum_{i=1}^{n} \min_{x_i \in \mathbb{R}} \left(f_i(x_i) - a_i v x_i \right) \\ &= bv - \sum_{i=1}^{n} f_i^*(a_i v) \end{aligned}$$

where f_i^* is the conjugate of f_i, to be defined shortly

Therefore the dual problem is

$$\max_{v \in \mathbb{R}} \; bv - \sum_{i=1}^{n} f_i^*(a_i v)$$

or equivalently

$$\min_{v \in \mathbb{R}} \; \sum_{i=1}^{n} f_i^*(a_i v) - bv$$

This is a convex minimization problem with scalar variable—much easier to solve than primal

Given v^*, the primal solution $x^\star$ solves

$$\min_{x \in \mathbb{R}^n} \; \sum_{i=1}^{n} (f_i(x_i) - a_i v^\star x_i)$$

Strict convexity of each f_i implies that this has a unique solution, namely $x^\star$, which we compute by solving $\partial f_i(x_i) \ni a_i v^\star$ for each i

Dual subtleties

- Often, we will transform the dual into an equivalent problem and still call this the dual. Under strong duality, we can use solutions of the (transformed) dual problem to characterize or compute primal solutions

  Warning: the optimal value of this transformed dual problem is not necessarily the optimal primal value

- A common trick in deriving duals for unconstrained problems is to first transform the primal by adding a dummy variable and an equality constraint

  Usually there is **ambiguity** in how to do this, and different choices lead to different dual problems!

Lasso dual

Recall the lasso problem:

$$\min_{x \in \mathbb{R}^p} \frac{1}{2}\|y - Ax\|^2 + \lambda \|x\|_1$$

Its dual function is just a constant (equal to $f^\star$). Therefore we redefine the primal as

$$\min_{x \in \mathbb{R}^p,\, z \in \mathbb{R}^n} \frac{1}{2}\|y - z\|^2 + \lambda \|x\|_1 \text{ subject to } z = Ax$$

so dual function is now

$$\begin{aligned} g(u) &= \min_{x \in \mathbb{R}^p,\, z \in \mathbb{R}^n} \frac{1}{2}\|y - z\|^2 + \lambda \|x\|_1 + u^T(z - Ax) \\ &= \frac{1}{2}\|y\|^2 - \frac{1}{2}\|y - u\|^2 - I_{\{v \,:\, \|v\|_\infty \le 1\}}(A^T u/\lambda) \end{aligned}$$

This calculation will make sense once we learn conjugates, shortly

Therefore the **lasso dual** problem is

$$\max_{u \in \mathbb{R}^n} \frac{1}{2}\Big(\|y\|^2 - \|y-u\|^2\Big) \text{ subject to } \|A^T u\|_\infty \le \lambda$$

or equivalently

$$\min_{u \in \mathbb{R}^n} \|y-u\|^2 \text{ subject to } \|A^T u\|_\infty \le \lambda$$

Note that strong duality holds here (Slater's condition), but the optimal value of the last problem is not necessarily the optimal lasso objective value

Further, note that given $u^\star$, any lasso solution $x^\star$ satisfies (from the z block of the stationarity condition) $z^\star - y + u^\star = 0$, i.e.,

$$Ax^\star = y - u^\star$$

So the lasso fit is just the dual residual

Outline

Today:

- Conjugate function
- Dual cones
- Dual polytopes
- Polar sets

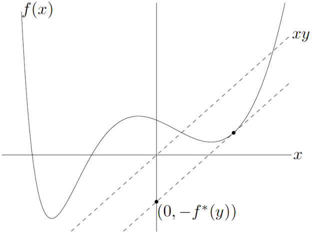
(And there are lots more duals—e.g., dual graphs, alebgraic dual, analytic dual—all related in some way...)

Conjugate function

Given a function $f : \mathbb{R}^n \to \mathbb{R}$, define its **conjugate** $f^* : \mathbb{R}^n \to \mathbb{R}$,

$$f^*(y) = \max_{x \in \mathbb{R}^n} y^T x - f(x)$$

Note that f^* is always convex, since it is the pointwise maximum of convex (affine) functions in y (f need not be convex)



$f^*(y)$: maximum gap between linear function $y^T x$ and $f(x)$

(From B & V page 91)

For differentiable f, conjugation is called the Legendre transform

Properties:

- Fenchel's inequality: for any x, y,

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

- Hence conjugate of conjugate f^{**} satisfies $f^{**} \leq f$
- If f is closed and convex, then $f^{**} = f$
- If f is closed and convex, then for any x, y,

$$\begin{aligned} x \in \partial f^*(y) \quad &\Leftrightarrow \quad y \in \partial f(x) \\ &\Leftrightarrow \quad f(x) + f^*(y) = x^T y \end{aligned}$$

- If $f(u, v) = f_1(u) + f_2(v)$ (here $u \in \mathbb{R}^n$, $v \in \mathbb{R}^m$), then

$$f^*(w, z) = f_1^*(w) + f_2^*(z)$$

Examples:

- Simple quadratic: let $f(x) = \frac{1}{2}x^T Qx$, where $Q \succ 0$. Then $y^T x - \frac{1}{2}x^T Qx$ is strictly concave in y and is maximized at $y = Q^{-1}x$, so
$$f^*(y) = \frac{1}{2}y^T Q^{-1} y$$
Note that Fenchel's inequality gives:
$$\frac{1}{2}x^T Qx + \frac{1}{2}y^T Q^{-1}y \geq x^T y$$

- Indicator function: if $f(x) = I_C(x)$, then its conjugate is
$$f^*(y) = I_C^*(y) = \max_{x \in C} \; y^T x$$
called the **support function** of C; we'll revisit this later

- Norm: if $f(x) = \|x\|$, then its conjugate is

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_* \le 1 \\ \infty & \text{else} \end{cases}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$ (recall that we defined $\|y\|_* = \max_{\|z\| \le 1} z^T y$).

Why? Note that if $\|y\|_* > 1$, then there exists $\|z\| \le 1$ with $z^T y = \|y\|_* > 1$, so

$$(tz)^T y - \|tz\| = t(z^T y - \|z\|) \to \infty, \quad \text{as } t \to \infty$$

i.e., $f^*(y) = \infty$

On the other hand, if $\|y\|_* \le 1$, then

$$z^T y - \|z\| \le \|z\| \|y\|_* - \|z\| \le 0$$

and $= 0$ when $z = 0$, so $f^*(y) = 0$

Conjugates and dual problems

Conjugates appear frequently in derivation of dual problems, via

$$-f^*(u) = \min_{x \in \mathbb{R}^n} \; f(x) - u^T x$$

in minimization of the Lagrangian. E.g., consider

$$\min_{x \in \mathbb{R}^n} \; f(x) + g(x)$$

$$\Leftrightarrow \quad \min_{x \in \mathbb{R}^n,\, z \in \mathbb{R}^n} \; f(x) + g(z) \;\text{ subject to }\; x = z$$

Lagrange dual function:

$$g(u) = \min_{x \in \mathbb{R}^n,\, z \in \mathbb{R}^n} \; f(x) + g(z) + u^T(z - x) = -f^*(u) - g^*(-u)$$

Hence dual problem is

$$\max_{u \in \mathbb{R}^n} \; -f^*(u) - g^*(-u)$$

Examples of this last calculation:

- Indicator function: dual of

$$\min_{x \in \mathbb{R}^n} \; f(x) + I_C(x)$$

is

$$\max_{u \in \mathbb{R}^n} \; -f(u) - I_C^*(-u)$$

where I_C^* is the support function of C

- Norms: the dual of

$$\min_{x \in \mathbb{R}^n} \; f(x) + \|x\|$$

is

$$\max_{u \in \mathbb{R}^n} \; -f^*(u) \;\; \text{subject to} \;\; \|u\|_* \leq 1$$

where $\| \cdot \|_*$ is the dual norm of $\| \cdot \|$

Double dual

Consider general minimization problem with linear constraints:

$$\min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{subject to } \; Ax \leq b, \;\; Cx = d$$

The Lagrangian is

$$L(x, u, v) = f(x) + (A^T u + C^T v)^T x - b^T u - d^T v$$

and hence the dual problem is

$$\max_{u \in \mathbb{R}^m, \, v \in \mathbb{R}^r} \quad -f^*(-A^T u - C^T v) - b^T u - d^T v$$
$$\text{subject to} \quad u \geq 0$$

Recall property: $f^{**} = f$ if f is closed and convex. Hence in this case, we can show that the **dual of the dual** is the primal

Actually, the connection (between duals of duals and conjugates) runs much deeper than this, beyond linear constraints. Consider

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) \le 0, \;\; i = 1, \ldots m \\ & \ell_j(x) = 0, \;\; j = 1, \ldots r \end{aligned}$$

If f and $h_1, \ldots h_m$ are closed and convex, and $\ell_1, \ldots \ell_r$ are affine, then the **dual of the dual** is the primal

This is proved by viewing the minimization problem in terms of a bifunction. In this framework, the dual function corresponds to the conjugate of this bifunction (for more, read Chapters 29 and 30 of Rockafellar)

Cones

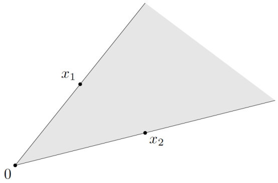
A set $K \in \mathbb{R}^n$ is called a **cone** if

$$x \in K \quad \Rightarrow \quad \theta x \in K \quad \text{for all } \theta \geq 0$$

It is called a **convex cone** if

$$x_1, x_2 \in C \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 \in C \quad \text{for all } \theta_1, \theta_2 \geq 0$$

i.e., K is convex and a cone



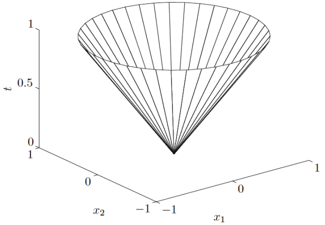
(From B & V page 26)

Examples:

- Linear subspace: any linear subspace is a convex cone
- Norm cone: if $\|\cdot\|$ is a norm then

$$K = \{(x, t) \in \mathbb{R}^{n+1} : \|x\| \leq t\}$$

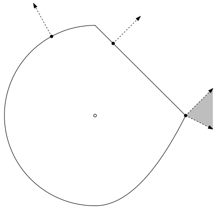
is a convex cone, called a norm cone (epigraph of norm function). Under 2-norm, called second-order cone, e.g.,



(From B & V page 31)

- Normal cone: given a set C, recall we defined its normal cone at a point $x \in C$ as

$$\mathcal{N}_C(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in C\}$$



This is always a convex cone, regardless of C

- Positive semidefinite cone: consider the set of (symmetric) positive semidefinite matrices

$$\mathbb{S}^n_+ = \{X \in \mathbb{R}^{n \times n} : X = X^T,\ X \succeq 0\}$$

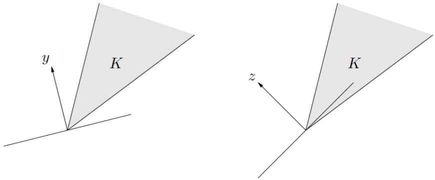
This is a convex cone, because for $A, B \succeq 0$ and $\theta_1, \theta_2 \geq 0$, $x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0$

Dual cones

For a cone $K \in \mathbb{R}^n$,

$$K^* = \{y \in \mathbb{R}^n : y^T x \geq 0 \text{ for all } x \in K\}$$

is called its **dual cone**. This is always a convex cone (even if K is not convex)



Note that $y \in K^* \Leftrightarrow$ the halfspace $\{x \in \mathbb{R}^n : y^T x \geq 0\}$ contains K

(From B & V page 52)

Important property: if K is a closed convex cone, then $K^{**} = K$

Examples:

- Linear subspace: the dual cone of a linear subspace V is $V^\perp$, its orthogonal complement. E.g., $(\mathrm{row}(A))^* = \mathrm{null}(A)$

- Norm cone: the dual cone of the norm cone

$$K = \{(x,t) \in \mathbb{R}^{n+1} : \|x\| \le t\}$$

is the norm cone of its dual norm

$$K^* = \{(y,s) \in \mathbb{R}^{n+1} : \|y\|_* \le s\}$$

- Positive semidefinite cone: the convex cone $\mathbb{S}^n_+$ is **self-dual**, meaning $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$. Why? Check that

$$Y \succeq 0 \quad \Leftrightarrow \quad \mathrm{tr}(YX) \ge 0 \text{ for all } X \succeq 0$$

by looking at the eigenvalue decomposition of X

Dual cones and dual problems

Consider the constrained problem

$$\min_{x \in K} f(x)$$

Recall that its dual problem is

$$\max_{u \in \mathbb{R}^n} -f^*(u) - I_K^*(-u)$$

where recall $I_K^*(y) = \max_{z \in K} z^T y$, the support function of K. If K is a cone, then this is simply

$$\max_{u \in K^*} -f^*(u)$$

where K^* is the dual cone of K, because $I_K^*(-u) = I_{K^*}(u)$

This is quite a useful observation, because many different types of constraints can be posed as cone constraints

Generalized inequalities

If $K \in \mathbb{R}^n$ is a proper cone (convex cone, closed, solid, pointed), then it induces a **generalized inequality** $\leq_K$ over $\mathbb{R}^n$ via

$$x \leq_K y \quad \text{if} \quad y - x \in K$$

Examples:

- Componentwise inequality: the nonnegative orthant is a proper cone, $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ all } i\}$, and it induces the generalized inequality: $x \leq_{\mathbb{R}^n_+} y$ if and only if $x_i \leq y_i$ for all i (we have been writing this as $x \leq y$)
- Matrix inequality: $\mathbb{S}^n_+$ is a proper cone, and it induces the generalized inequality: $X \leq_{\mathbb{S}^n_+} Y$ if and only if $Y - X$ is positive semidefinite (we have been writing this as $X \preceq Y$)

Hence any set of generalized inequalities can be posed in terms of cone constraints

Conic solvers

Two general suites of solvers, that rely on transforming a convex problem into conic form (i.e., one with cone constraints) are **CVX**[1] and **TFOCS**[2]

- Transformation to conic form is not necessarily unique, and different transformations yield different problems, possibly of varying difficulty
- CVX is more general; TFOCS is less general but can be a lot faster (apparently close to state of the art)
- Both are freely available (implemented in MATLAB)

[1] M. Grant and S. Boyd (2008), *Graph implementations for nonsmooth convex problems*, `http://cvxr.com/cvx`

[2] S. Becker and E. Candes and M. Grant (2010), *Templates for convex cone problems with applications to sparse signal recovery*, `http://cvxr.com/tfocs`

Given a problem in conic form, TFOCS (Templates for First-Order Conic Solvers) derives and solves the dual problem[3], and then computes a primal solution relying on strong duality. Consider:

$$\min_{x \in \mathbb{R}^n} \ f(x)$$
$$\text{subject to } \ Ax + b \in K$$

for a convex cone K. The dual problem is

$$\max_{u \in \mathbb{R}^n} \ -f^*(A^T u) - b^T u$$
$$\text{subject to } \ u \in K^*$$

Important point: **projection** onto K^* is quite often a lot easier than projection onto $\{x \in \mathbb{R}^n : Ax + b \in K\}$, so we can employ a a first-order method on the dual

[3]Actually, in TFOCS the dual problem is often smoothed before being solved, but we haven't covered smoothing yet

E.g., consider the problem

$$\min_{x \in \mathbb{R}^p} f(x) \;\; \text{subject to} \;\; \|y - Ax\|_2 \leq \sigma$$

where the parameter $\sigma > 0$ is a known fixed quantity. This can be transformed into desired conic form by writing the constraint as

$$\begin{bmatrix} A \\ 0 \end{bmatrix} x + \begin{bmatrix} -y \\ \sigma \end{bmatrix} \in \{(z,t) \in \mathbb{R}^{n+1} : \|z\|_2 \leq t\}$$

i.e., K is the second-order cone. Note that $K^* = K$, self-dual, and projection onto K is easy:

$$P_K(z,t) = \begin{cases} (z,t) & \text{if } \|z\|_2 \leq t \\ \dfrac{\|z\|_2 + t}{2\|z\|_2} \cdot (z,t) & \text{if } -t \leq \|z\|_2 \leq t \\ (0,0) & \text{if } t \leq -\|z\|_2 \end{cases}$$

Polytopes

A polytope $P \in \mathbb{R}^n$ is the convex hull of a finite number of points in $\mathbb{R}^n$:

$$P = \text{conv}\{x_1, \dots x_k\}$$

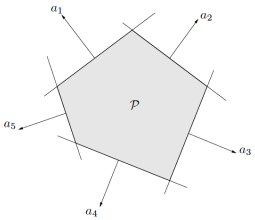
This is called the **V-representation** of P. Fundamental result: P is a polytope $\Leftrightarrow$ P is a bounded polyhedron, i.e., P is bounded and

$$P = \bigcap_{i=1}^{m} \{x \in \mathbb{R}^n : a_i^T x \leq b_i\}$$

This is called the **H-representation** of P. These representations also called primal and dual representations, we'll see why shortly

H-representation

(From B & V page 32)



Faces of polytopes

A **face** of a polytope P is a set F such that

$$x, y \in P \text{ and } \frac{x+y}{2} \in F \quad \Rightarrow \quad x, y \in F$$

The set of faces of P written $\mathcal{F}(P)$. Properties and definitions:

- Each face F of P satisfies $F = \emptyset$, $F = P$, or $F = P \cap H$ for a supporting hyperplane H to P
- Faces $F \neq \emptyset, P$ are called proper
- A face F is said to have dimension d (or, called a d-face) if $\mathrm{aff}(F)$ is d-dimensional
- If $F = \{x\}$ is a 0-face, then x is called a **vertex**. Moreover,

  $$P = \mathrm{conv}\{x_1, \ldots x_k\}$$

  for the vertices $x_1, \ldots x_k$ of P. Conversely, if $P = \mathrm{conv}(A)$, then A contains the vertices of P

- If F is an $(n-1)$-face, then it is called a **facet**.[4] If $F_1, \dots F_m$ are the facets of P, then

$$P = \bigcap_{i=1}^{m} H_i$$

for halfspaces H_i such that $\operatorname{bd}(H_i) = \operatorname{aff}(F_i)$. Conversely, if

$$P = \bigcap_{i=1}^{m} H_i$$

for halfspaces H_i, then $\{\operatorname{bd}(H_i) \cap P : i = 1, \dots m\}$ contains the facets of P

- The set of faces $\mathcal{F}(P)$ can be partially ordered by inclusion. Note that, with respect to this ordering, vertices are minimal proper faces, and facets are maximal proper faces

[4]This is assuming, without a loss of generality, that $\operatorname{aff}(P) = \mathbb{R}^n$. Otherwise we just reparametrize to $\mathbb{R}^d$, where $d = \dim(\operatorname{aff}(P))$

Dual polytopes

Given a polytope $P \in \mathbb{R}^n$, a polytope $P^* \in \mathbb{R}^n$ is called its **dual polytope** if there exists a one-to-one mapping $\Psi : \mathcal{F}(P) \to \mathcal{F}(P^*)$ that is inclusion-reversing:

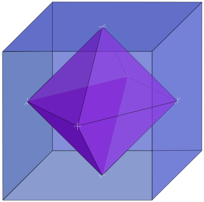
$$F_1 \subseteq F_2 \quad \Leftrightarrow \quad \Psi(F_1) \supseteq \Psi(F_2), \quad \text{all } F_1, F_2 \in \mathcal{F}(P)$$

This implies that

$$\dim(F) + \dim(\Psi(F)) = n - 1, \quad \text{all } F \in \mathcal{F}(P)$$

E.g., cross-polytope (1-norm ball) and hypercube (∞-norm ball) are dual

(From `http://en.wikipedia.org/wiki/Dual_polyhedron`)



Does every polytope have a dual? As we'll see shortly, answer is yes

One use of polytope dualilty (among many) is that it allows us to compute (in theory) one type of representation from the other:

- Suppose we had an H-representation for P^*. From this we can enumerate facets of $F_1^*, \ldots F_k^*$ of P^*, and hence vertices

$$x_1 = \Psi^{-1}(F_1^*), \ \ldots \ x_k = \Psi^{-1}(F_k^*)$$

  of P. Therefore $\mathrm{conv}\{x_1, \ldots x_k\}$ is a V-representation for P

- Suppose we had V-representation for P^*. Then we can enumerate vertices $x_1^*, \ldots x_m^*$ of P^*, which yields facets

$$F_1 = \Psi^{-1}(x_1^*), \ \ldots \ F_m = \Psi^{-1}(x_m^*)$$

  of P. Hence $\cup_{i=1}^m H_i$ is an H-representation for P, where H_i are halfspaces with $\mathrm{bd}(H_i) = \mathrm{aff}(F_i)$

Polar sets

Given a set $C \in \mathbb{R}^n$,

$$C^\circ = \{y \in \mathbb{R}^n : y^T x \leq 1 \text{ for all } x \in C\}$$

is called its **polar set**, and is always convex (even when C is not)

Polarity is the most general form of geometric duality. Properties and examples:

- If C is a closed, convex set containing 0, then $C^{\circ\circ} = C$
- If C is a cone, then

  $$C^\circ = \{y \in \mathbb{R}^n : y^T x \leq 0 \text{ for all } x \in C\} = -C^*$$

  where C^* is the dual cone. Here C° is called the **polar cone**
- If C is a polytope, then C° is its dual polytope, and Ψ can be defined by

  $$\Psi(F) = \{y \in C^\circ : y^T x = 1 \text{ for all } x \in C\}$$

- If C is the sublevel set of a norm $\|\cdot\|$,

$$C = \{x \in \mathbb{R}^n : \|x\| \leq t\}$$

for some $t > 0$, then its polar is also a sublevel set,

$$C^\circ = \{y \in \mathbb{R}^n : \|y\|_* \leq 1/t\}$$

where $\|\cdot\|_*$ is the dual norm

- The support function of C satisfies

$$I_C^*(y) \leq 1 \quad \Leftrightarrow \quad y \in C^\circ$$

and if C is a cone, then $I_C^*(y) = I_{C^\circ}(y)$

- Support functions I_C^* and $I_{C^\circ}^*$ are called **dual seminorms**, and satisfy

$$x^T y \leq I_C^*(x) I_{C^\circ}^*(y) \quad \text{for all } x, y \in \mathbb{R}^n$$

References

- S. Boyd and L. Vandenberghe (2004), *Convex Optimization*, Cambridge University Press, Chapters 2, 3, 5
- B. Grunbaum (2003), *Convex Polytopes*, Springer, Chapters 2, 3
- R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press, Chapters 12, 13, 14, 16