Duality

Goran Banjac

Large-Scale Convex Optimization
ETH Zurich

March 24, 2020

Lagrangian

standard form problem (not necessarily convex)

$$
\begin{array}{ll}
\text{minimize} & f(x) \\
\text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\
 & h_j(x) = 0, \quad j = 1, \ldots, p
\end{array}
$$

variable $x \in \mathbb{R}^n$, domain $\mathcal{D} \subseteq \mathbb{R}^n$, optimal value $p^\star$

Lagrangian: $\mathcal{L} \colon \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$

$$
\mathcal{L}(x, \lambda, \nu) \coloneqq f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)
$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $g_i(x) \leq 0$
- ν_j is Lagrange multiplier associated with $h_j(x) = 0$

Lagrange dual function

Lagrange dual function: $d\colon \mathbb{R}^m \times \mathbb{R}^p \mapsto \mathbb{R}$

$$
\begin{aligned}
d(\lambda, \nu) &\coloneqq \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} \left\{ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \nu_j h_j(x) \right\}
\end{aligned}
$$

d is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \geq 0$, then $d(\lambda, \nu) \leq p^\star$

- indeed, if $\bar{x}$ is feasible and $\lambda \geq 0$, then

$$
f(\bar{x}) \geq \mathcal{L}(\bar{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) = d(\lambda, \nu)
$$

- minimizing over all feasible $\bar{x}$ gives $p^\star \geq d(\lambda, \nu)$

Lagrange dual problem

Lagrange dual problem:

$$\begin{array}{ll} \text{maximize} & d(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- finds best lower bound on $p^\star$ obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, d$
- often simplified by making implicit constraints $(\lambda, \nu) \in \mathbf{dom}\, d$ explicit

Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

- Lagrangian is given by

$$\begin{aligned} \mathcal{L}(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

- $\mathcal{L}$ is affine in x, hence

$$d(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) = \begin{cases} -b^T \nu & c + A^T \nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- dual problem is thus

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \geq 0 \end{array}$$

Weak and strong duality

weak duality: $d^\star \leq p^\star$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

strong duality: $d^\star = p^\star$

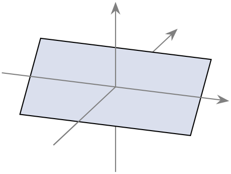
- does not hold in general
- (usually) holds for convex problems
- conditions that guarantee strong duality in convex problems are called *constraint qualification (CQ)*

Relative interior

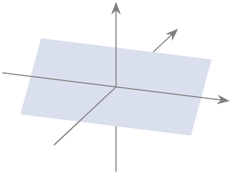
- $\operatorname{relint} \mathcal{C}$ is interior of $\mathcal{C} \subseteq \mathbb{R}^n$ relative to its affine hull
- properties:
  - $\operatorname{relint} \mathcal{C} \subseteq \mathcal{C}$
  - if $\operatorname{aff} \mathcal{C} = \mathbb{R}^n$, then $\operatorname{relint} \mathcal{C} = \operatorname{int} \mathcal{C}$
  - if $\mathcal{C}$ is nonempty and convex, then $\operatorname{relint} \mathcal{C} \neq \emptyset$

$\mathcal{C} \subseteq \mathbb{R}^2$

$\operatorname{int} \mathcal{C} = \operatorname{relint} \mathcal{C}$



$\mathcal{C} \subseteq \mathbb{R}^3$



$\operatorname{int} \mathcal{C} \neq \operatorname{relint} \mathcal{C}$

Slater's constraint qualification

- strong duality holds for a convex problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

  if it is strictly feasible, *i.e.*,

$$\exists x \in \operatorname{int} \mathcal{D} \quad \text{such that} \quad g_i(x) < 0,\ i = 1, \ldots, m, \quad Ax = b$$

- can be sharpened by replacing $\operatorname{int} \mathcal{D}$ with $\operatorname{relint} \mathcal{D}$; linear inequalities need not hold with strict inequality
- there exist many other types of constraint qualification

Complementary slackness

- assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$\begin{aligned} f(x^\star) = d(\lambda^\star, \nu^\star) &= \inf_{x \in \mathcal{D}} \left\{ f(x) + \sum_{i=1}^m \lambda_i^\star g_i(x) + \sum_{j=1}^p \nu_j^\star h_j(x) \right\} \\ &\leq f(x^\star) + \sum_{i=1}^m \lambda_i^\star g_i(x^\star) + \sum_{j=1}^p \nu_j^\star h_j(x^\star) \\ &\leq f(x^\star) \end{aligned}$$

- hence, the two inequalities hold with equality
- $x^\star$ minimizes $\mathcal{L}(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star g_i(x^\star) = 0$ for all $i = 1, \ldots, m$ (*complementary slackness*):

$$\lambda_i^\star > 0 \Longrightarrow g_i(x^\star) = 0, \qquad g_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0$$

Karush-Kuhn-Tucker (KKT) conditions

- the following four conditions are called KKT conditions (for a problem with differentiable f, g_i, h_j):

  1. primal feasibility: $g_i(x) \leq 0$, $i = 1, \ldots, m$, $h_j(x) = 0$, $j = 1, \ldots, p$
  2. dual feasibility: $\lambda \geq 0$
  3. complementary slackness: $\lambda_i g_i(x) = 0$, $i = 1, \ldots, m$
  4. stationarity: gradient of Lagrangian with respect to x vanishes

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{p} \nu_j \nabla h_j(x) = 0$$

- if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

KKT conditions for convex problem

- if $\bar{x}, \bar{\lambda}, \bar{\nu}$ satisfy KKT conditions for a convex problem, then they are optimal
  - from complementary slackness and primal feasibility: $f(\bar{x}) = \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\nu})$
  - from stationarity: $d(\bar{\lambda}, \bar{\nu}) = \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\nu})$
  - hence, $f(\bar{x}) = d(\bar{\lambda}, \bar{\nu})$

- if Slater's condition is satisfied, then x is optimal if and only if there exist λ, ν so that (x, λ, ν) satisfy KKT conditions

Implicit constraints

- **example:** LP with box constraints

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b \\
& -\mathbf{1} \preceq x \preceq \mathbf{1}
\end{array}
\qquad\qquad
\begin{array}{ll}
\text{maximize} & -b^T\nu - \mathbf{1}^T\lambda_1 - \mathbf{1}^T\lambda_2 \\
\text{subject to} & c + A^T\nu + \lambda_1 - \lambda_2 = 0 \\
& \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0
\end{array}
$$

- reformulation: $\mathcal{D} = \{x \in \mathbb{R}^n \mid -\mathbf{1} \preceq x \preceq \mathbf{1}\}$

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b
\end{array}
$$

- dual function:

$$
\begin{aligned}
d(\nu) &= \inf_{x \in \mathcal{D}} \left\{ c^T x + \nu^T (Ax - b) \right\} \\
&= -b^T\nu - \|A^T\nu + c\|_1
\end{aligned}
$$

- dual problem: maximize $-\, b^T\nu - \|A^T\nu + c\|_1$

Standard form conic program

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b \\
& x \geq_{\mathcal{K}} 0
\end{array}
$$

- $\mathcal{K} \subseteq \mathbb{R}^m$ is a proper convex cone
- Lagrange multiplier for $x \geq_{\mathcal{K}} 0$ is vector $\lambda \in \mathbb{R}^m$
- Lagrangian is given by

$$\mathcal{L}(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b)$$

- lower bound property: if $\lambda \geq_{\mathcal{K}^*} 0$, then $d(\lambda, \nu) \leq p^\star$
- dual problem

$$
\begin{array}{ll}
\text{maximize} & d(\lambda, \nu) \\
\text{subject to} & \lambda \geq_{\mathcal{K}^\star} 0
\end{array}
$$

Composite minimization

- consider *primal* composite optimization problem

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

  (f and g are convex closed proper, L linear operator)
- we will derive dual problem and primal-dual optimality conditions

Fenchel duality

- reformulate the composite optimization problem as

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

- equivalent formulation with indicator function

$$\text{minimize} \quad f(x) + g(y) + \mathcal{I}_{\{0\}}(Lx - y)$$

- reformulate the indicator function via

$$h(x, y) = \sup_{\mu} \left\{ \mu^T (Lx - y) \right\}$$

- the problem can now be written as

$$p^\star = \inf_{(x,y)} \sup_{\mu} \left\{ f(x) + g(y) + \mu^T (Lx - y) \right\}$$

Weak duality

- let

$$\mathcal{L}(x, y, \mu) := f(x) + g(y) + \mu^T(Lx - y)$$

- then

$$p^\star = \inf_{(x,y)} \sup_{\mu} \mathcal{L}(x, y, \mu) \geq \sup_{\mu} \inf_{(x,y)} \mathcal{L}(x, y, \mu) = d^\star$$

- the inequality is known as the *min-max inequality*
- holds for general functions (also in nonconvex case)
- in our setting this is called *weak duality*

Fenchel dual problem

- the problem with inf-sup swapped is the Fenchel dual problem

$$
\begin{aligned}
\sup_{\mu} \inf_{(x,y)} \mathcal{L}(x,y,\mu) &= \sup_{\mu} \inf_{(x,y)} \left\{ f(x) + g(y) + \mu^T(Lx - y) \right\} \\
&= \sup_{\mu} -\Big(\sup_{(x,y)} \left\{ -f(x) - g(y) + \mu^T(-Lx + y) \right\} \Big) \\
&= \sup_{\mu} \Big\{ -\sup_{x} \left\{ x^T(-L^T\mu) - f(x) \right\} \\
&\qquad\qquad - \sup_{y} \left\{ \mu^T y - g(y) \right\} \Big\} \\
&= \sup_{\mu} \left\{ -f^*(-L^T\mu) - g^*(\mu) \right\}
\end{aligned}
$$

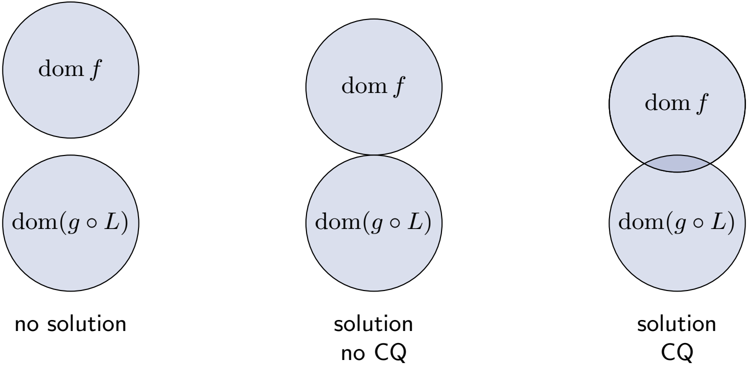
- hence, primal and dual problems are

minimize $\quad f(x) + g(Lx)$ $\qquad\qquad$ maximize $\; -f^*(-L^T\mu) - g^*(\mu)$

Constraint qualification and strong duality

- strong duality holds if f, g are proper closed convex and

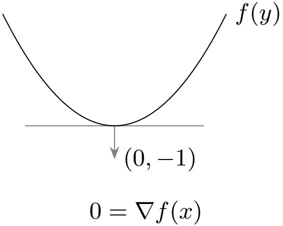
$$\operatorname{relint} \operatorname{dom} f \cap \operatorname{relint} \operatorname{dom}(g \circ L) \neq \emptyset$$

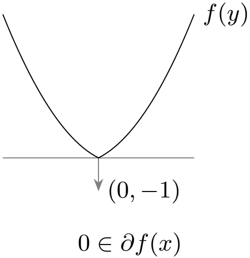


Fermat's rule

- let $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, then $x \in \mathbb{R}^n$ minimizes f if and only if

$$0 \in \partial f(x)$$





- x minimizes f if and only if

$$f(y) \geq f(x) + 0^T(y - x) \quad \text{for all } y \in \mathbb{R}^n$$

which by definition of subdifferential is equivalent to $0 \in \partial f(x)$

Fenchel-Young's equality

- recall the definition of conjugate function

$$f^*(y) = \sup_x \left\{ y^T x - f(x) \right\}$$

- applying Fermat's rule, we obtain the *Fenchel-Young's equality*:

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

- similarly, applying Fermat's rule to biconjugate f^{**} yields

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

- if f is closed and convex, then $f^{**} = f$, and thus

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

(also known as *subdifferential inverse*)

Subdifferential calculus rules

- **subdifferential of sum**: if $x \in \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2$, then

$$\partial(f_1 + f_2)(x) \supseteq \partial f_1(x) + \partial f_2(x)$$

- if f_1, f_2 are closed and convex and $\mathrm{relint}\, \mathrm{dom}\, f_1 \cap \mathrm{relint}\, \mathrm{dom}\, f_1 \neq \emptyset$:

$$\partial(f_1 + f_2) = \partial f_1 + \partial f_2$$

- **subdifferential of composition**: if $Lx \in \mathrm{dom}\, f$, then

$$\partial(f \circ L)(x) \supseteq L^T \partial f(Lx)$$

- if f is closed and convex and $\mathrm{relint}\, \mathrm{dom}(f \circ L) \neq \emptyset$:

$$\partial(f \circ L)(x) = L^T \partial f(Lx)$$

Optimality conditions for composite minimization

- let $f\colon \mathbb{R}^n \mapsto \overline{\mathbb{R}}$, $g\colon \mathbb{R}^m \mapsto \overline{\mathbb{R}}$, $L \in \mathbb{R}^{m\times n}$ with f, g closed convex and CQ holds (see slide 18), then

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

is solved by $x \in \mathbb{R}^n$ if and only if x satisfies

$$0 \in \partial f(x) + L^T \partial g(Lx)$$

- CQ implies subdifferential calculus with equality:

$$0 \in \partial f(x) + L^T \partial g(Lx) = \partial \left(f + (g \circ L)\right)(x)$$

- many algorithms search for x that satisfy $0 \in \partial f(x) + L^T \partial g(Lx)$

Alternative optimality conditions

- introducing dual variable $\mu \in \partial g(Lx)$ and using subdifferential inverse yields primal-dual necessary and sufficient optimality conditions:

$$\begin{cases} -L^T\mu \in \partial f(x) \\ \mu \in \partial g(Lx) \end{cases} \qquad \begin{cases} x \in \partial f^*(-L^T\mu) \\ \mu \in \partial g(Lx) \end{cases}$$

$$\begin{cases} -L^T\mu \in \partial f(x) \\ Lx \in \partial g^*(\mu) \end{cases} \qquad \begin{cases} x \in \partial f^*(-L^T\mu) \\ Lx \in \partial g^*(\mu) \end{cases}$$

- note that the last condition is equivalent to

$$0 \in (-L)\partial f^*(-L^T\mu) + \partial g^*(\mu)$$

which is Fermat's rule for the dual problem

$$\text{maximize } -\big(f^*(-L^T\mu) + g^*(\mu)\big)$$

References

- these lecture notes are based to a large extent on the following material:
  - Stanford EE364a class developed by Stephen Boyd
  - Lund course on Large-Scale Convex Optimization developed by Pontus Giselsson

- the original slides can be downloaded from

  `https://web.stanford.edu/class/ee364a/lectures.html`

  `https://archive.control.lth.se/ls-convex-2015/`