Introduction to Convex Optimization

Lecture 6: Duality

Silin DU

Department of Management Science and Engineering
Tsinghua University
dsl21@mails.tsinghua.edu.cn

TSINGHUA UNIVERSITY
清华经管学院

August 23, 2022

Contents

Lecture Overview

In this lecture, we cover Lagrangian duality, which plays a central role in convex optimization.

1. Lagrange dual problem.
2. Weak and strong duality.
3. Geometric interpretation.
4. Optimality conditions.
5. Examples of primal and dual problem.

We put some proofs in appendix.

Contents

The Lagrangian

Consider a standard form problem (not necessarily convex)

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, ..., m \\ & h_i(x) = 0, & i = 1, ..., p \end{array} \tag{1}$$

variable $x \in \mathbb{R}^n$, domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} f_i \cap \bigcap_{i=0}^{p} \mathbf{dom} h_i$, optimal value $p^\star$.
Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\mathbf{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$
- λ and ν are called the dual variables associated with Problem 1

Definition 1 (Lagrange Dual Function)

Lagrange dual function (or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is the minimum value of the Lagrangian over x

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) + \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) = \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$.
- The dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when Problem 1 is not convex.

Lagrangian Dual Function II

Theorem 1 (Lower Bounds on Optimal Value)

The dual function yields lower bounds on the optimal value $p^\star$ of Problem 1. For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \leq p^\star$$

- If $\tilde{x}$ is feasible for Problem 1 and $\lambda \succeq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0$$

  Therefore,

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

- Minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda, \nu)$.

Examples of Dual Functions I

Least-norm solution of linear equations

$$\begin{array}{ll} \min & x^\top x \\ \text{s.t.} & Ax = b \end{array}$$

- Lagrangian is $L(x, \nu) = x^\top x + v^\top (Ax - b)$
- To minimize L over x, set gradient equal to zero:

$$\nabla_x L(x, \nu) = 2x + A^\top \nu = 0 \Longrightarrow x = -(1/2)A^\top \nu$$

- Plug in L to obtain g:

$$g(v) = L\left(-(1/2)A^\top \nu, \nu\right) = -\frac{1}{4}\nu^\top AA^\top \nu - b^\top \nu$$

  a concave function of ν
- $p^\star \geq -\frac{1}{4}\nu^\top AA^\top \nu - b^\top \nu$ for all ν.

Examples of Dual Functions II

Standard form LP

$$
\begin{array}{ll}
\min & c^\top x \\
\text{s.t.} & Ax = b, \quad \succeq 0
\end{array}
$$

- Lagrangian is

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

- L is affine in x, hence

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

  g is linear on affine domain $\left\{(\lambda, \nu) \mid A^\top \nu - \lambda + c = 0\right\}$, hence concave

- lower bound property: $p^\star \geq -b^\top \nu$ if $A^\top \nu + c \succeq 0$

Examples of Dual Functions III

Equality constrained norm minimization

$$\begin{array}{ll} \min & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

- dual function

$$g(\nu) = \inf_x \left(\|x\| - \nu^T Ax + b^\top \nu \right) = \begin{cases} b^\top \nu, & \left\|A^\top \nu\right\|_* \leq 1 \\ -\infty, & \text{otherwise} \end{cases}$$

  where $\|v\|_* = \sup_{\|u\| \leq 1} u^\top v$ is dual norm of $\| \cdot \|$

- if $\|y\|_* \leq 1$, then $\|x\| - y^\top x \geq 0$ for all x, with equality if $x = 0$
- if $\|y\|_* > 1$, choose $x = tu$ where $\|u\| \leq 1, u^\top y = \|y\|_* > 1$:

$$\|x\| - y^\top x = t\left(\|u\| - \|y\|_*\right) \to -\infty \quad \text{as } t \to \infty$$

- lower bound property: $p^\star \geq b^\top \nu$ if $\left\|A^\top \nu\right\|_* \leq 1$

Examples of Dual Functions IV

Two-way partitioning problem

$$\begin{array}{ll} \min & x^\top W x \\ \text{s.t.} & x_i^2 = 1, \quad i = 1, \cdots, n \end{array}$$

- a nonconvex problem; feasible set contains 2^n discrete points
- interpretation: partition $\{1, \ldots, n\}$ in two sets; W_{ij} is cost of assigning i, j to the same set; $-W_{ij}$ is cost of assigning to different sets
- dual function

$$\begin{aligned} g(\nu) &= \inf_x \left(x^\top W x + \sum_i \nu_i \left(x_i^2 - 1 \right) \right) \\ &= \inf_x x^\top (W + \operatorname{diag}(\nu)) x - \mathbf{1}^\top \nu \\ &= \begin{cases} -\mathbf{1}^\top \nu, & W + \operatorname{diag}(\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

- lower bound property: $p^\star \geq -1^\top \nu$ if $W + \operatorname{diag}(\nu) \succeq 0$
- example: $\nu = -\lambda_{\min}(W)\mathbf{1}$ gives bound $p^\star \geq n\lambda_{\min}(W)$

Lagrange Dual and Conjugate Function I

Recall that the conjugate f^* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f^*(y) = \sup_{x \in \mathbf{dom} f} \left(y^\top x - f(x) \right)$$

The conjugate function and Lagrange dual function are closely related.

- Consider the following problem with linear ineuqality and equality constraints

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax \preceq b \\ & Cx = d \end{array}$$

The associated dual function is

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left(f_0(x) + \lambda^\top (Ax - b) + \nu^\top (Cx - d) \right) \\ &= -b^\top \lambda - d^\top \nu + \inf_x \left(f_0(x) + \left(A^\top \lambda + C^\top \nu \right)^\top x \right) \\ &= -b^\top \lambda - d^\top \nu - f_0^* \left(-A^\top \lambda - C^\top \nu \right) \end{aligned} \tag{2}$$

The domain of g follows from the domain of f_0^*:

$$\mathbf{dom} g = \left\{ (\lambda, \nu) \mid -A^\top \lambda - C^\top \nu \in \mathbf{dom} f_0^* \right\}$$

Lagrange Dual and Conjugate Function II

- Equality constrained norm minimization

$$\begin{array}{ll} \min & \|x\| \\ \text{subject to} & Ax = b \end{array}$$

where $\|\cdot\|$ is any norm. Recall that the conjugate of $f_0 = \|\cdot\|$ is given by

$$f_0^*(y) = \begin{cases} 0, & \|y\|_* \leq 1 \\ \infty, & \text{otherwise} \end{cases}$$

the indicator function of the dual norm unit ball. Using the result from Eq. 2, the dual function is

$$g(\nu) = -b^\top \nu - f_0^*\left(-A^\top \nu\right) = \begin{cases} -b^\top \nu, & \left\|A^\top \nu\right\|_* \leq 1 \\ -\infty, & \text{otherwise} \end{cases}$$

- Entropy maximization

$$\begin{array}{ll} \min & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{s.t.} & Ax \preceq b \\ & \mathbf{1}^\top x = 1 \end{array}$$

where $\mathbf{dom} f_0 = \mathbb{R}^n_{++}$. The conjugate of the negative entropy function $u \log u$ with scalar variable u, is $e^{\nu-1}$. Since f_0 is a sum of negative entropy functions of different variables, we conclude that its conjugate is

$$f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

with $\mathbf{dom} f_0^* = \mathbb{R}^n$. Using the result from Eq. 2, the dual function is

$$g(\lambda, \nu) = -b^\top - \nu - f_0^*\left(-A^\top \lambda - \nu\right) = -b^\top - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^\top \lambda}$$

where a_i is the i-th column of A.

- Minimum volume covering ellipsoid

$$\begin{array}{lll} \min & f_0(x) = \log\det X^{-1} & \\ \text{s.t.} & a_i^\top X a_i \le 1, & i = 1, \ldots, m \end{array}$$

where $\mathbf{dom} f_0 = \mathcal{S}_{++}^n$. With each $X \in \mathcal{S}_{++}^n$ we associated the ellipsoid, centered at the origin,

$$\mathcal{E}_X = \left\{ z \mid z^\top X z \le 1 \right\}$$

The volume of this ellipsoid is proportional to $\left(\det X^{-1}\right)^{-1/2}$. The constraints of the problem are $a_i \in \mathcal{E}_X$. Thus the problem is to determine the minimum volume ellipsoid, centered at the origin, that includes the points $a_1, \ldots, a_m$.
The inequality constraints are affine.

$$\operatorname{tr}\left(\left(a_i a_i^\top\right) X\right) \le 1$$

The conjugate of f_0 is

$$f_0^*(Y) = \log\det(-Y)^{-1} - n$$

with $\mathbf{dom} f_0^* = -\mathcal{S}_{++}^n$. Applying the result from Eq. 2, the dual function is

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^n \lambda_i a_i a_i^\top\right) - \mathbf{1}^\top \lambda + n, & \sum_{i=1}^n \lambda_i a_i a_i^\top \succ 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Contents

Dual Problem

We go back to Problem 1. The dual function $g(\lambda, \nu)$ gives a lower bound on the optimal value $p^\star$ that depends on some parameters λ, ν. To find the best lower bound, we obtain the Lagrange dual problem

$$\operatorname{tr}\left(\left(a_i a_i^\top\right) X\right) \leq 1$$

- a convex optimization problem ($g(\lambda, \nu)$ is concave); optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0, (\lambda, \nu) \in \mathbf{dom} g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \mathbf{dom} g$ explicit

Making Dual Constraints Explicit

- Standard form LP and its dual

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax = b \\ & x \succeq 0 \end{array} \qquad\qquad \begin{array}{ll} \max & -b^\top \nu \\ \text{s.t.} & A^\top \nu + c \succeq 0 \end{array}$$

- Inequality form LP

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Ax \succeq b \end{array} \qquad\qquad \begin{array}{ll} \max & -b^\top \lambda \\ \text{s.t.} & A^\top \lambda + c = 0 \\ & \lambda \succeq 0 \end{array}$$

Weak Duality

Property 1

The optimal value of the Lagrange dual problem, which we denote $d^\star$, is, by definition, the best lower bound on $p^\star$ that can be obtained from the Lagrange dual function. In particular, we have

$$d^\star \leq p^\star \tag{3}$$

which holds even if the original problem is not convex. This property is called weak duality.

- The weak duality inequality always holds.
- If the primal problem is unbounded below ($p^\star = -\infty$), we must have $d^\star = -\infty$, i.e., the dual problem is infeasible.
- Conversely, if the dual problem is unbounded above ($d^\star = \infty$), we must $p^\star = \infty$, i.e., the primal problem is infeasible.
- This property can be used to find nontrivial lower bounds for difficult problems.
- $p^\star - d^\star$ is the optimality gap.

Strong Duality

Property 2 (Strong Duality)

If the equality

$$d^\star = p^\star$$

holds, i.e., the optimal duality gap is zero, then we say that strong duality holds.

- Strong duality does not, in general, hold.
- Strong duality usually but not always holds for convex problems.
- Conditions that guarantee strong duality in convex problems are called constraint qualifications.

Slater's Condition

Theorem 2 (Slater's theorem)

Consider the standard form convex optimization problem.

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \cdots, m \\ & Ax = b \end{aligned}$$

The Slater's condition is that there exists an $x \in \text{relint}\mathcal{D}$ such that

$$f_i(x) < 0, i = 1, ..., m, \quad Ax = b$$

Strong duality holds, if Slater's condition holds (and the problem is convex).

- If the first k constraint functions $f_1, ..., f_k$ are affine, then strong duality holds provided the following weaker condition holds: there exists an $x \in \textbf{relint}\mathcal{D}$ with

  $$f_i(x) \leq 0, i = 1, ..., k, \quad f_i(x) < 0, i = k+1, ..., m, \quad Ax = b$$

  The affine inequalities do not need to hold with strict inequality.
- Slater's condition also implies that the dual optimal value is attained when $d^\star > -\infty$, i.e., there exists a dual feasible $(\lambda^\star, \nu^\star)$ with $g(\lambda^\star, \nu^\star) = d^\star = p^\star$.

Examples of Strong Duality I

Least-squares solution of linear equalities

$$\begin{array}{ll} \min & x^\top x \\ \text{s.t.} & Ax = b \end{array}$$

The associated dual problem is

$$\max \quad -(1/4)\nu^\top AA^\top \nu - b^\top \nu$$

- Slater's condition is simply that the primal problem is feasible, so $p^\star = d^\star$ provided $b \in \mathcal{R}(A)$, i.e., $p^\star < \infty$.
- For this problem we always have strong duality, even when $p^\star = \infty$. This is the case when $b \notin \mathcal{A}$, so there is a z with $A^\top z = 0, b^\top z \neq 0$. It follows that the dual function is unbounded above along the line $\{tz \mid t \in \mathbb{R}\}$, so $d^\star = \infty$ as well.

Examples of Strong Duality II

Inequality form LP

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax \preceq b
\end{aligned}
$$

Dual problem

$$
\begin{aligned}
\max \quad & -b^\top \lambda \\
\text{s.t.} \quad & A^\top \lambda + c = 0 \\
& \lambda \succeq 0
\end{aligned}
$$

- From Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$.
- In fact $p^\star = d^\star$ except when primal and dual are infeasible.