Introduction to Convex Optimization

Lecture 6: Duality

Silin DU

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

August 28, 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)

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

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 $f_i(x) \le 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$
- λ and ν are called the dual variables associated with Problem 1

Lagrange Dual Function I

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.

Lagrange 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.

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

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

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

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

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

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

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^*:

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

- 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

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

where a_i is the i-th column of A.

Lagrange Dual and Conjugate Function IV

- Minimum volume covering ellipsoid

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

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

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

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 $a_1, ..., a_m$.
The inequality constraints are affine.

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

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

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

- Inequality form LP

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

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.

Theorem 2 (Slater's theorem)

Consider the standard form convex optimization problem.

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

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

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

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

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

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

  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

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

Dual problem

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

- 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.

Examples of Strong Duality III

Quadratic programming

$$\begin{array}{ll} \min & x^\top P x \qquad (P \in \mathcal{S}^n_{++}) \\ \text{s.t.} & Ax \preceq b \end{array}$$

Dual function

$$\begin{aligned} g(\lambda) &= -b^\top \lambda + \inf_x \left(x^\top P x + \lambda^\top A x \right) \\ &= -b^\top \lambda - \frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda \end{aligned} \qquad (x = -\tfrac{1}{2} P^{-1} A^\top \lambda)$$

Dual problem

$$\begin{array}{ll} \max & -b^\top \lambda - \frac{1}{4} \lambda^\top A P^{-1} A^\top \lambda \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- From Slater's condition: $p^\star = d^\star$ if $A\tilde{x} \prec b$ for some $\tilde{x}$
- In fact, $p^\star = d^\star$ always

Examples of Strong Duality IV

QCQP

$$\begin{array}{ll} \min & (1/2)x^\top P_0 x + q_0^\top x + r_0 \\ \text{s.t.} & (1/2)x^\top P_i x + q_i^\top x + r_i \leq 0, \quad i = 1, ..., m \\ & P_0 \in \mathcal{S}_{++}^n, P_i \in \mathcal{S}_+^n \end{array}$$

The Lagrangian is

$$L(x, \lambda) = (1/2)x^\top P(\Lambda)x + q(\lambda)^\top x + r(\lambda)$$

where

$$P(\lambda) = P_0 + \sum_{i=1}^m \lambda_i P_i, \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$$

For $\lambda \succeq 0$,

$$g(\lambda) = \inf_x L(x, \lambda) = -(1/2)q(\lambda)^\top P(\lambda)^{-1} q(\lambda) + r(\lambda)$$

Dual problem

$$\begin{array}{ll} \max & -(1/2)q(\lambda)^\top P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

Strong duality holds if the quadratic inequality constraints are strictly feasible.

Examples of Strong Duality V

Entropy maximization

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

Dual problem

$$\begin{array}{ll} \max & -b^\top \lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^\top \lambda} \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- From weaker Slater condition, $d^\star = p^\star$ if there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^\top x = 1$.
- We can simplify this problem by maximizing over the dual varaible ν. By FOC,

$$\nu = \log \sum_{i=1}^n e^{-a_i^\top \lambda} - 1$$

Then we get

$$\begin{array}{ll} \max & -b^\top \lambda - \log \left(\sum_{i=1}^n e^{-a_i^\top \lambda} \right) \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

Examples of Strong Duality VI

Minimum volume covering ellipsoid

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

Dual problem

$$\begin{array}{ll} \min & f_0(x) = \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^\top\right) - \mathbf{1}^\top \lambda + n \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

- The weaker Slater's condition is that there exists an $X \in \mathcal{S}_{++}^n$ with $a_i^\top X a_i \leq 1$. This is always satisfied, so strong duality always holds.

A nonconvex problem with strong duality I

Minimize a nonconvex quadratic function over the unit ball

$$\begin{array}{ll} \min & f_0(x) = \log\det\left(\sum_{i=1}^{m} \lambda_i a_i a_i^\top\right) - \mathbf{1}^\top \lambda + n \\ \text{s.t.} & \lambda \succeq 0 \end{array}$$

where $A \in \mathcal{S}^n, A \not\succeq 0$.
Dual function

$$g\left(\lambda\right) = \inf_x L(\lambda, x) = \inf_x \left(x^\top (A + \lambda I)x + 2b^\top x - \lambda\right)$$

- First-order derivative: $\nabla_x L(\lambda, x) = 2(A + \lambda I) + 2b$.
- $g(\lambda)$ is unbounded below if $A + \lambda I \not\succeq 0$ or $A + \lambda I \succeq 0, b \notin \mathcal{R}(A + \lambda I)$
- $x = -(A + \lambda I)^{+} b, \quad g(\lambda) = -b^\top (A + \lambda I)^{+} b - \lambda$

A nonconvex problem with strong duality II

Dual problem

$$
\begin{array}{ll}
\max & -b^\top (A+\lambda I)^+ b - \lambda \\
\text{s.t.} & A+\lambda I \succeq 0 \\
& b \in \mathcal{R}(A+\lambda I)
\end{array}
\qquad
\begin{array}{ll}
\max & -t - \lambda \\
\text{s.t.} & \begin{bmatrix} A+\lambda I & b \\ b^\top & t \end{bmatrix} \succeq 0
\end{array}
$$

Note that by Schur complement

$$
\begin{bmatrix} A+\lambda I & b \\ b^\top & t \end{bmatrix} \succeq 0 \Longleftrightarrow t - b^\top (A+\lambda I)^+ b \geq 0, \text{ and } t \geq 0
$$

- Strong duality holds although primal problem is not convex.
- In fact, a more general result holds: strong duality holds for any optimization problem with quadratic objective and one quadratic inequality constraint, provided Slater's condition holds.

Mixed Strategies for Matrix Games I

We use strong duality to derive a basic result for zero-sum matrix game.

- Consider a game with two players. Player 1 makes a choice $k \in \{1, ...n\}$, and player 2 makes a choice $l \in \{1, ..., m\}$.
- Player 1 then makes a payment of P_{kl} to player 2, where $P \in \mathbb{R}^{n \times m}$ is the payoff matrix for the game.
- The goal of player 1 is to make the payment as small as possible, while the goal of player 2 is to maximize it.
- The players use randomized or mixed strategies, which means that each player makes his or her choice randomly and independently of the other player's choice, according to a probability distribution:

$$\mathbf{prob}(k = i) = u_i, i = 1, ..., n, \quad \mathbf{prob}(l = i) = v_i, i = 1, ..., m$$

- The expected payoff from player 1 to player 2 is then

$$\sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl} = u^\top P v$$

- First consider the point of view player 1. Assuming her strategy u is known to player 2. Player 2 will choose v to maximize $u^\top Pv$, which results in the expected payoff

$$\sup\left\{u^\top Pv \mid v \succeq 0, \mathbf{1}v = 1\right\} = \max_{i=1,\dots,m} \left(P^\top u\right)_i$$

Then player 1 needs to minimize the worst-case payoff to player 2, i.e.,

$$\begin{array}{ll} \min & \max_{i=1,\dots,m} \left(P^\top u\right)_i \\ \text{s.t.} & u \succeq 0, \quad \mathbf{1}^\top u = 1 \end{array} \tag{4}$$

which is a piecewise-linear convex optimization problem. Denote the optimal value of this problem as $p_1^\star$.

- Similarly, player 2 wants to maximize the payoff, i.e.,

$$\begin{array}{ll} \max & \min_{i=1,\dots,n} \left(Pv\right)_i \\ \text{s.t.} & v \succeq 0, \quad \mathbf{1}^\top v = 1 \end{array} \tag{5}$$

which is another convex optimization problem. Denote the optimal value of this problem as $p_2^\star$.

Using duality, we can establish $p_1^\star = p_2^\star$. Formulate Problem 4 as an LP

$$\begin{array}{ll} \min_{t,u} & t \\ \text{s.t.} & P^\top u \preceq t\mathbf{1}, \mathbf{1}^\top u = 1, u \succeq 0 \end{array}$$

The Lagrangian is $(\alpha, \beta \succeq 0)$

$$\begin{aligned} L(t, u, \alpha, \beta, \nu) &= t + \alpha^\top \left(P^\top u - t\mathbf{1}\right) + \beta^\top(-u) + v\left(\mathbf{1}^\top u - 1\right) \\ &= -\nu + t - t\alpha^\top \mathbf{1} + (P\alpha)^\top u - \beta^\top u + \nu \mathbf{1}^\top u \end{aligned}$$

Dual function

$$g(\alpha, \beta, \nu) = \inf_{t,u} L(t, u, \alpha, \beta, \nu) = \begin{cases} -\nu, & P\alpha + \nu\mathbf{1} - \beta = 0 \text{ and } \alpha^\top \mathbf{1} = 1 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem

$$\begin{array}{llcll} \max & -\nu & \iff & \max & \nu \\ \text{s.t.} & P\alpha + \nu\mathbf{1} = \beta & & \text{s.t.} & P\alpha \succeq \nu\mathbf{1} \\ & \mathbf{1}^\top\alpha = 1 & & & \mathbf{1}^\top\alpha = 1 \\ & \alpha \succeq 0, \beta \succeq 0 & & & \alpha \succeq 0 \end{array}$$

which is equivalent to Problem 5.

- Strong duality holds for LPs. Thus, $p_1^\star = p_2^\star$.
- In other words, in a matrix game with mixed strategies, there is no advantage to knowing your opponent's strategy.

Contents

Weak and Strong Duality via Set of Values I

We can give a simple geometric interpretation of the dual function in terms of the set

$$\mathcal{G} = \{(f_1(x), ..., f_m(x), h_1(x), ..., h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid x \in \mathcal{D}\}$$

which is the set of values taken on by the constraint and objective functions. The optimal value $p^\star$ is

$$p^\star = \inf \{t \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\}$$

The Lagrangian is

$$(\lambda, \nu, 1)^\top (u, v, t) = \sum_{i=1}^m \lambda_i u_i + \sum_{i=1}^p \nu_i v_i + t$$

Minimize the affine function over $(u, \nu, t) \in \mathcal{G}$

$$g(\lambda, \nu) = \inf \left\{(\lambda, \nu, 1)^\top (u, v, t) \mid (u, v, t) \in \mathcal{G}\right\}$$

If the infimum is finite, then the inequality

$$(\lambda, \nu, 1)^\top (u, v, t) \geq g(\lambda, \nu)$$

defines a supporting hyperplane to $\mathcal{G}$

Now suppose $\lambda \succeq 0$. Then $t \geq (\lambda, \nu, 1)^{\top}(u, v, t)$ if $u \preceq 0, v = 0$.

$$\begin{aligned} p^{\star} &= \inf\{t \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\} \\ &\geq \inf\left\{(\lambda, \nu, 1)^{\top}(u, v, t) \mid (u, v, t) \in \mathcal{G}, u \preceq 0, v = 0\right\} \\ &\geq \inf\left\{(\lambda, \nu, 1)^{\top}(u, v, t) \mid (u, v, t) \in \mathcal{G}\right\} \\ &= g(\lambda, \nu) \end{aligned}$$

Weak duality holds. This interpretation is illustrated in Fig 1.

Weak and Strong Duality via Set of Values III

$$g(\lambda) = \inf_{(u,t)\in\mathcal{G}} (t + \lambda u), \qquad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$

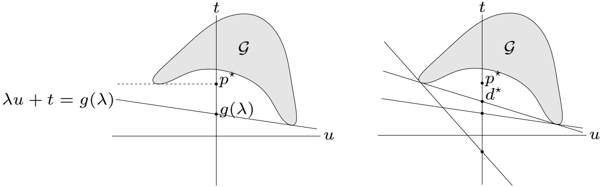

Fig. 1: (Left.) Geometric interpretation of dual function and lower bound $g(\lambda) \le p^\star$, for a problem with one inequality constraint. (Right.) Supporting hyperplanes corresponding to three dual feasible values of λ, including the optimum $\lambda^\star$.

- $\lambda u + t = g(\lambda)$ is a nonvertical supporting hyperplane (since the last element) with slope $-\lambda$. The intersection of this hyperplane with the $u = 0$ axis gives $g(\lambda)$.
- Strong duality does not hold.

Epigraph Variation I

We define the set $\mathcal{A} \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$ as

$$\mathcal{A} = \mathcal{G} + \left(\mathbb{R}_+^m \times \{0\}^p \times \mathbb{R}_+\right)$$

or, more explicitly

$$\mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D}, f_i(x) \leq u_i, i = 1, ..., m, h_i(x) = v_i, i = 1, ..., p, f_0(x) \leq t\}$$

The optimal value of $\mathcal{A}$ is

$$p^\star = \inf \{t \mid (0, 0, t) \in \mathcal{A}\}$$

Dual function

$$g(\lambda, \nu) = \inf \left\{(\lambda, \nu, 1)^\top (u, v, t) \mid (u, v, t) \in \mathcal{A}\right\}$$

If the infinum is finite then the inequality

$$(\lambda, \nu, 1)^\top (u, v, t) \geq g(\lambda, \nu)$$

defines a nonvertical supporting hyperplane to $\mathcal{A}$. In particular, since $(0, 0, p^\star) \in \mathbf{bd}\mathcal{A}$, we have

$$p^\star = (\lambda, \nu, 1)^\top (0, 0, p^\star) \geq g(\lambda, \nu)$$

the weak duality bound.
Strong duality holds if and only if we have equality above for some dual feasible (λ, ν), i.e., there exists a nonvertical supporting hyperplane to $\mathcal{A}$ at its boundary point $(0, 0, p^\star)$.

Epigraph Variation III

$$\mathcal{A} = \{(u,t) \mid f_1(x) \le u, f_0(x) \le t, \text{ for some } x \in \mathcal{D}\}$$

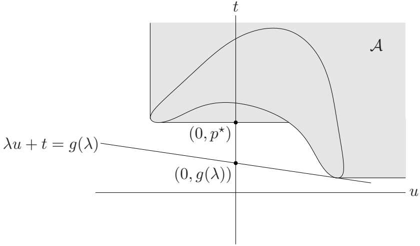

Fig. 2: Geometric interpretation of dual function and lower bound $g(\lambda) \le p^\star$, for a problem with one inequality constraint.

- Strong duality holds if there is a nonvertical supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$
- For convex problem, $\mathcal{A}$ is convex, hence has supporting hyperplane at $(0, p^\star)$
- Slater's condition: if there exists $(\tilde{u}, \tilde{t}) \in \mathcal{A}$ with $\tilde{u} < 0$, then supporting hyperplanes at $(0, p^\star)$ must be nonvertical.

Proof of Strong Duality under Constraint Qualification I

Consider the standard form convex optimization problem

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

Two additional assumptions:

- $\mathcal{D}$ has nonempty interior ($\mathrm{int}\mathcal{D} \neq \emptyset$)
- A is full rank, i.e., $\mathrm{rank}(A) = p$

Assume $p^\star$ is finite. Define a convex set $\mathcal{B}$ as

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^\star\}$$

1. $\mathcal{A} \cap \mathcal{B} = \emptyset$. Suppose $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$.

$$(u, v, t) \in \mathcal{B} \Longrightarrow u = 0, v = 0, t < p^\star$$
$$(u, v, t) \in \mathcal{B} \Longrightarrow \text{there exists an } x \text{ with } f_i(x) \le 0, Ax - b = 0, f_0(x) \le t \le p^\star$$

which contradicts with $p^\star$ is optimal.

2. By the separating hyperplane theorem, there exists $\left(\tilde{\lambda}, \tilde{\nu}, \mu\right) \neq 0$ and α such that

$$(u, v, t) \in \mathcal{A} \Rightarrow \left(\tilde{\lambda}, \tilde{\nu}, \mu\right)^\top (u, v, t) = \tilde{\lambda}^\top u + \tilde{\nu}^\top v + \mu t \geq \alpha \tag{6}$$

$$(u, v, t) \in \mathcal{B} \Rightarrow \left(\tilde{\lambda}, \tilde{\nu}, \mu\right)^\top (u, v, t) = \tilde{\lambda}^\top u + \tilde{\nu}^\top v + \mu t \leq \alpha \tag{7}$$

From Eq. 6 we conclude that $\tilde{\lambda} \succeq 0$ and $\mu \geq 0$. (Otherwise $\tilde{\lambda}^\top u + \mu t$ is unbounded below over $\mathcal{A}$)

Eq. 7 means that $\mu t \leq \alpha$ for all $t < p^\star$, and hence $\mu p^\star \leq \alpha$. Together with Eq. 6 we have for any $x \in \mathcal{D}$

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^\top (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^\star \tag{8}$$

3. Assume that $\mu = 0$. From Eq. 8, we conclude that for all $x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^\top (Ax - b) \geq 0 \tag{9}$$

Applying this to the point $\tilde{x}$ that satisfies the Slater's condition, we have

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) \geq 0$$

Since $f_i(\tilde{x})$ and $\tilde{\lambda}_i \geq 0$, $\tilde{\lambda} = 0$. From $\left(\tilde{\lambda}, \tilde{\nu}, \mu\right) \neq 0$ and $\tilde{\lambda} = 0, \mu = 0$, we conclude that $\nu \neq 0$.
Then Eq. 9 implies that for all $x \in \mathcal{D}$, $\tilde{\nu}^\top (Ax - b) \geq 0$. But $\tilde{x}$ satisfies $\tilde{\nu}^\top (A\tilde{x} - b) = 0$, and since $\tilde{x} \in \text{int}\mathcal{D}$, there are points in $\mathcal{D}$ with $\tilde{\nu}^\top (Ax - b) < 0$ unless $A^\top \tilde{\nu} = 0$, which means

$$\nu_1 a_1 + \nu_2 a_2 + ... + \nu_p a_p = 0$$

This contradicts with $\text{rank}(A) = p$.

4. Now consider $\mu > 0$. We divide Eq. 8 by ν to obtain

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^\star$$

for all $x \in \mathcal{D}$, from which it follows, by minimizing over x, that $g(\lambda, \nu) \geq p^\star$, where we define

$$\lambda = \tilde{\lambda}/\mu, \quad \nu = \tilde{\nu}/\mu$$

By weak duality we have $g(\lambda, \nu) \leq p^\star$, so in fact $g(\lambda, \nu) = p^\star$. Strong duality holds.

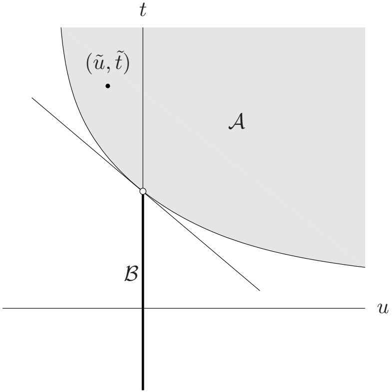

Fig. 3: Illustration of strong duality proof, for a convex problem that satisfies Slater's constraint qualification.

- The set $\mathcal{A}$ is shown shaded, and the set $\mathcal{B}$ is the thick vertical line segment, not including the point $(0, p^\star)$, shown as a small open circle.
- The two sets are convex and do not intersect, so they can be separated by a hyperplane.
- Slater's constraint qualification guarantees that any separating hyperplane must be nonvertical, since it must pass to the left of the point $(\tilde{u}, \tilde{t}) = (f_1(\tilde{x}), f_0(\tilde{x}))$, where $\tilde{x}$ is strictly feasible.

Contents

Max-Min Characterization of Weak and Strong Duality I

Assume there are no equality constraints. First note that

$$\sup_{\lambda \succeq 0} L(x,\lambda) = \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$
$$= \begin{cases} f_0(x) & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ \infty & \text{otherwise.} \end{cases}$$

This means that we can express the optimal value of the primal problem as

$$p^\star = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda)$$

By the definition of the dual function, we also have

$$d^\star = \sup_{\lambda \succeq 0} \inf_x L(x,\lambda)$$

Then weak duality can be expressed as

$$\sup_{\lambda \succeq 0} \inf_x L(x,\lambda) \leq \inf_x \sup_{\lambda \succeq 0} L(x,\lambda) \tag{10}$$

and strong duality as the equality

$$\sup_{\lambda \succeq 0} \inf_{x} L(x, \lambda) = \inf_{x} \sup_{\lambda \succeq 0} L(x, \lambda)$$

Strong duality means that the order of the minimization over x and the maximization over $\lambda \succeq 0$ can be switched without affecting the result.
In fact, Eq 10 does not depend on ant properties of L: We have

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

for any $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ (and any $W \subseteq \mathbb{R}^n, Z \subseteq \mathbb{R}^m$). This general inequality is called the max-min inequality. When equality holds, i.e.,

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z) \tag{11}$$

we say that f (and W and Z) satisfy the strong max-min property or the saddle-point property.

Saddle-Point Interpretation

We refer to a point $\tilde{w} \in W, \tilde{z} \in Z$ as a saddle-point for f (and W and Z) if

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\tilde{w}$ minimizes $f(w, \tilde{z})$ (over $w \in W$) and $\tilde{z}$ maximize $f(\tilde{w}, z)$ (over $z \in Z$):

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z)$$

This implies that the strong max-min property (Eq. 11) holds, and that the common value is $f(\tilde{w}, \tilde{z})$.

Returning to the discussion of Lagrange duality, we see that if $x^\star$ and $\lambda^\star$ are primal and dual optimal points for a problem in which strong duality obtains, they form a saddle-point for the Lagrangian. The converse is also true: If (x, λ) is a saddle-point of the Lagrangian, then x is primal optimal, λ is dual optimal, and the optimal duality gap is zero.

Game Interpretation I

We consider a continuous zero-sum game with two players.

- If player 1 chooses $w \in W$, and player 2 selects $z \in Z$, then player 1 pays an amount $f(w, z)$ to player 2. Player 1 therefore wants to minimize f, while player 2 wants to maximize f. (The game is called continuous since the choices are vectors, and not discrete.)
- Suppose that player 1 makes his choice first, and then player 2, after learning the choice of player 1, makes her selection.
- Player 2 wants to maximize the payoff $f(w, z)$, and so will choose $z \in Z$ to maximize $f(w, z)$. The resulting payoff will be $\sup_{z \in Z} f(w, z)$. which will depends on the choice of player 1.
- Player 1 knows (or assumes) that player 2 will follow this strategy, and so will choose $w \in W$ to make this worst-case payoff to player 2 as small as possible, which results in the payoff

$$\inf_{w \in W} \sup_{z \in Z} f(w, z)$$

Game Interpretation II

- Now suppose the order of play is reversed: Player 2 must choose $z \in Z$ first, and then player 1 chooses $w \in W$ (with knowledge of z). The payoff from player 1 to player 2 is

$$\sup_{z \in Z} \inf_{w \in W} f(w, z)$$

- The max-min inequality (Eq. 10) states the fact that it is better for a player to go second, or more precisely, for a player to know his or her opponent's choice before choosing.
- When the saddle-point property (Eq. 11) holds, there is no advantage to playing second.
- If $(\tilde{w}, \tilde{z})$ is a saddle-point for f (and W and Z), then it is called a solution of the game; $\tilde{w}$ is called the optimal choice or strategy for player 1, and $\tilde{z}$ is the optimal choice for player 2.
- Now consider the special case where the payoff function is the Lagrangian, $W = \mathbb{R}^n$ and $Z = \mathbb{R}_+^m$. Here player 1 chooses the primal variable x, while player 2 chooses the dual variable $\lambda \succeq 0$.

Game Interpretation III

- If player 2 chooses first, the optimal choice is any $\lambda^\star$ which is dual optimal. It results in a payoff to player 2 of $d^\star$.
- Conversely, if player 1 must choose first, the optimal choice is any primal optimal $x^\star$, which results in a payoff of $p^\star$.
- The optimal duality gap for the problem is exactly equal to the advantage afforded the player who goes second, i.e., the player who has the advantage of knowing his or her opponent's choice before choosing.
- If strong duality holds, then there is no advantage to the players of knowing their opponent's choice.

Theorem 3 (Von-neumann Min-Max Theorem)

Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be compact convex sets. If f : $X \times Y \to \mathbb{R}$ is a continuous function and:

1. $f(\cdot, y) : X \to \mathbb{R}$ is convex for fixed y
2. $f(x, \cdot) : Y \to \mathbb{R}$ is concave for fixed x

Then $\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y)$.

Contents

Certificate of Suboptimality and Stopping Criteria I

- Dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of $p^\star$.
- Indeed, if x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu)$$

  This establishes that x is ϵ-suboptimal, with $\epsilon = f_0(x) - g(\lambda, \nu)$.
- A primal dual feasible pair $x, (\lambda, \nu)$ localizes the optimal value of the primal (and dual) problems to an interval:

$$p^\star \in [g(\lambda, \nu), f_0(x)], \quad d^\star \in [g(\lambda, \nu), f_0(x)]$$

  the width of which is the duality gap.
- If the duality gap of the primal dual feasible pair $x, (\lambda, \nu)$ is zero, i.e., $f_0(x) = g\,(\lambda, \nu)$, then x is primal optimal and (λ, ν) is dual optimal.
- These observations can be used in optimization algorithms to provide nonheuristic stopping criteria.

- Suppose an algorithm produces a sequence of primal feasible $x^{(k)}$ and dual feasible $\left(\lambda^{(k)}, \nu^{(k)}\right)$, for $k = 1, 2, \ldots$, and $\epsilon_{abs} > 0$ is a given required absolute accuracy. Then the stopping criterion (i.e., the condition for terminating the algorithm)

$$f_0\left(x^{(k)}\right) - g\left(\lambda^{(k)}, \nu^{(k)}\right) \leq \epsilon_{\text{abs}}$$

  guarantees that when the algorithm terminates, $x^{(k)}$is ϵ_{abs}-suboptimal.

- Of course strong duality must hold if this method is to work for arbitrarily small tolerances ϵ_{abs}.

Complementary Slackness I

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

$$
\begin{aligned}
f_0\left(x^\star\right) &= g\left(\lambda^\star, \nu^\star\right) \\
&= \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\
&\leq f_0\left(x^\star\right) + \sum_{i=1}^m \lambda_i^\star f_i\left(x^\star\right) + \sum_{i=1}^p \nu_i^\star h_i\left(x^\star\right) \\
&\leq f_0\left(x^\star\right)
\end{aligned}
$$

Hence, the two inequalities hold with equality.

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x. The Lagrangian $L(x, \lambda^\star, \nu^\star)$ can have other minimizers; $x^\star$ is simply a minimizer.

Complementary Slackness II

- Another important conclusion is that

$$\sum_{i=1}^{m} \lambda_i^\star f_i\left(x^\star\right) = 0$$

Since each term in this sum is nonpositive, we conclude that

$$\lambda^\star f_i\left(x^\star\right) = 0, \quad i = 1, \ldots, m$$

This condition is known as complementary slackness. It holds for any primal optimal $x^\star$ and any dual optimal $(\lambda^\star, \nu^\star)$ (when strong duality holds).

$$\lambda_i^\star > 0 \Rightarrow f_i\left(x^\star\right) = 0, \quad f_i\left(x^\star\right) < 0 \Rightarrow \lambda_i^\star = 0$$

KKT Condition I

We now assume that the functions $f_0, ..., f_m, h_1, ..., h_p$ are differentiable (and therefore have open domains). Note that we make no assumptions about convexity yet.

- Assume that strong duality holds. Let $x^\star$ and $(\lambda^\star, \nu^\star)$ be any primal and dual optimal points with zero duality gap.
- Since $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x, it follows that its gradient vanishes at x, i.e.,

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0$$

Then the following four conditions hold

1. primal feasible: $f_i(x^\star) \leq 0, i = 1, ..., m, h_i(x) = 0, i = 1, ..., p$
2. dual feasible: $\lambda_i^\star \geq 0, i = 1, ..., m$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, ..., m$
4. gradient of Lagrange with respect to x vanishes (first-order condition)

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0$$

which are called Karush-Kuhn-Tucker (KKT) conditions.

KKT Condition II

For any optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions. (necessary condition)

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.

If f_i are convex and h_i are affine (and differentiable), and $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are any points that satisfy the KKT conditions, then $\tilde{x}$ and $\left(\tilde{\lambda}, \tilde{\nu}\right)$ are primal and dual feasible, with zero duality gap.

- Note that the first condition states that $\tilde{x}$ is feasible. Since $\tilde{\lambda}_i \geq 0, L\left(x, \tilde{\lambda}, \tilde{\nu}\right)$ is convex in x.
- The last KKT condition states that the gradient of $L\left(x, \tilde{\lambda}, \tilde{\nu}\right)$ with respect to x vanishes at $x = \tilde{x}$, so $\tilde{x}$ minimizes $L\left(x, \tilde{\lambda}, \tilde{\nu}\right)$

KKT Condition III

- Then we conclude that

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\ &= f_0(\tilde{x}) \end{aligned}$$

  where in the last line we use $h_i(\tilde{x}) = 0$ and $\tilde{\lambda}_i f_i(\tilde{x}) = 0$ (complementary slackness).

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality

Examples of KKT Conditions I

Equality constrained convex quadratic minimization.

$$\begin{array}{ll} \min & (1/2)x^\top P x + q^\top x + r \\ \text{s.t.} & Ax = b \end{array}$$

where $P \in \mathcal{S}_+^n$. The KKT condition for this problem are

$$Ax^\star = b, \quad Px^\star + q + A^\top \nu^\star = 0$$

which we can write as

$$\begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$$

Solving this set of equations gives the optimal primal and dual variables.

Water-filling

$$\begin{array}{ll} \min_x & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{s.t.} & x \succeq 0 \\ & \mathbf{1}^\top x = 1 \end{array}$$

where $\alpha_i > 0$. The Lagrangian is

$$L(x, \lambda, \nu) = -\sum_{i=1}^n \log(\alpha_i + x_i) + \lambda^\top(-x) + \nu\left(\mathbf{1}^\top x - 1\right)$$

Then the KKT conditions are

$$\begin{aligned} & x^\star \succeq 0, \quad \mathbf{1}^\top x^\star - 1 = 0 \\ & \lambda^\star \succeq 0 \\ & \lambda_i^\star x_i^\star = 0 \\ & \frac{-1}{\alpha_i + x_i^\star} - \lambda_i^\star + \nu^\star = 0 \end{aligned}$$

which is equivalent to (eliminate $\lambda^\star$)

$$
\begin{aligned}
& x^\star \succeq 0, \quad \mathbf{1}^\top x^\star - 1 = 0 \\
& \left(\nu^\star - \frac{1}{\alpha_i + x_i^\star}\right) x_i^\star = 0 \\
& \nu \geq \frac{1}{\alpha_i + x_i^\star}
\end{aligned}
$$

- If $\nu^\star < \frac{1}{\alpha_i}$, then $x_i^\star > 0$ and $x_i^\star = \frac{1}{\nu^\star} - \alpha_i$.
- If $\nu^\star \geq \frac{1}{\alpha_i}$, $x_i^\star$ must be zero due to the complementary slackness condition.
- Thus we have

$$
x^\star = \begin{cases} \frac{1}{\nu^\star} - \alpha_i, & \nu^\star < \frac{1}{\alpha_i} \\ 0, & \nu^\star \geq \frac{1}{\alpha_i} \end{cases} = \max\left\{0, \frac{1}{\nu^\star} - \alpha_i\right\}
$$

Substituting this expression for $x^\star$ into the condition $\mathbf{1}^\top x^\star = 1$ we obtain

$$
\sum_{i=1}^{n} \max\left\{0, \frac{1}{\nu^\star} - \alpha_i\right\} = 1
$$

The lefthand side is a piecewise-linear increasing function of $1/\nu^\star$, with breakpoints at α_i , so the equation has a unique solution which is readily determined.

Examples of KKT Conditions IV

This solution method is called water-filling for the following reason.

- We think of α_i as the ground level above patch i, and then flood the region with water to a depth $1/\nu$, as illustrated in Fig. 4 .
- The total amount of water used is $\sum_{i=1}^{n} \max\left\{0, \frac{1}{\nu^\star} - \alpha_i\right\}$.
- We then increase the flood level until we have used a total amount of water equal to one.
- The depth of water above patch i is then the optimal value $x_i^\star$.

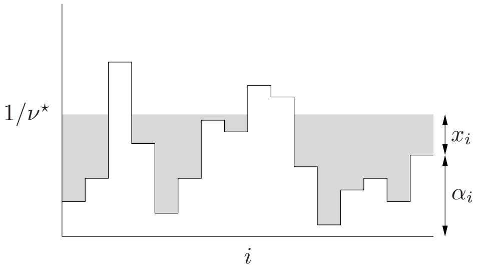

Fig. 4: Illustration of water-filling algorithm.

Solve the Primal Problem via the Dual

If strong duality holds and a dual optimal solution $(\lambda^\star, \nu^\star)$ exists, then any primal optimal point is also a minimizer of $L(x, \lambda^\star, \nu^\star)$. This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution.

- Suppose we have strong duality and an optimal $(\lambda^\star, \nu^\star)$ is known.
- Suppose that the minimizer of $L(x, \lambda^\star, \nu^\star)$, i.e., the solution of

$$\min_x \quad f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \tag{12}$$

  is unique. For a convex problem this occurs.
- Then if the solution of Problem. 12 is primal feasible, it must be primal optimal; if it is not primal feasible, then no primal optimal point can exist, i.e., we can conclude that the primal optimum is not attained.

Contents

Introducing New Variables and Equality Constraints

Contents

asd