Duality (I)

Lecture 9, Convex Optimization

National Taiwan University

May 6, 2021

Table of contents

The Lagrangian of an optimization problem (1/2)

- Consider an optimization problem, not necessarily convex, in the standard form:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, p, \end{array}$$

with variable $x \in \mathbf{R}^n$, with a nonempty domain

$$\mathcal{D} = \left(\bigcap_{i=0}^{m} \textbf{dom } f_i \right) \cap \left(\bigcap_{i=1}^{p} \textbf{dom } h_i \right),$$

and the optimal value being p^*.

The Lagrangian of an optimization problem (2/2)

- The **Lagrangian** associated with the problem is defined as $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

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

  with **dom** $L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$.
- We refer to λ_i and ν_i as the Lagrange multiplier associated with the ith inequality constraint $f_i(x) \leq 0$ and that with the ith equality constraint $h_i(x) = 0$, respectively.
- The vectors λ and ν are called the **dual variables** or Lagrange multiplier vectors associated with the original problem.

The Lagrange dual function

- The **Lagrange dual function** (or just dual function) is defined as the minimum value of the Lagrangian over x: $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ and for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

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

- When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$.
- The domain of the dual function is set to be

$$\mathbf{dom}\ g = \{(\lambda, \nu) \mid g(\lambda, \nu) > -\infty\}.$$

- The dual function is always concave.

The Dual Function Gives Lower Bounds on Optimal Value

- For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \leq p^*.$$

- Proof: Suppose $\tilde{x}$ is a feasible point for the original problem, i.e., $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \succeq 0$. Then we have

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

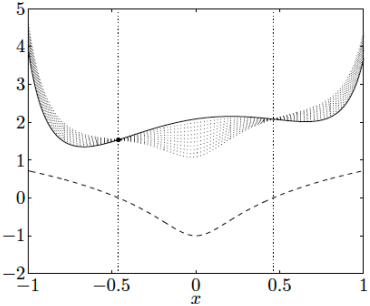
and therefore

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

So, $g(\lambda, \nu) \leq f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$.

Example

- A simple problem with $x \in \mathbf{R}$ and one inequality constraint.



- The objective function f_0: in solid curve.
- The constraint function f_1: in dashed curve.
- The feasible set $= [-0.46, 0.46]$.
- The optimal point and value: $x^* = -0.46$, $p^* = 1.54$.
- The dotted curves show $L(x, \lambda)$ for $\lambda = 0.1, 0.2, ..., 1.0$.

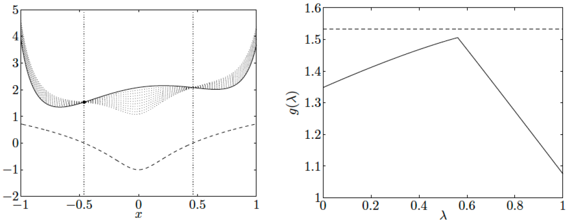
The Lagrange dual function

- When $g(\lambda, \nu) = -\infty$, the inequality

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

  holds, but is vacuous.
- The dual function gives a nontrivial lower bound on p^* only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \textbf{dom}\ g$, i.e., $g(\lambda, \nu) > -\infty$.
- We refer to a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \textbf{dom}\ g$ as **dual feasible**.

Example



- Left figure: The objective function f_0: in solid curve. The constraint function f_1: in dashed curve. Neither f_0 nor f_1 is convex.
- Right figure: The dual function g for the problem in the left figure. It is concave. The horizontal dashed line shows p^*, the optimal value of the problem.

Linear approximation interpretation (1/3)

- The Lagrangian and lower bound property can be given a simple interpretation, based on a linear approximation of the **indicator functions** of the sets $\{0\}$ and $-\mathbf{R}_+$ (defined below).
- We rewrite the original problem as an unconstrained problem,

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{i=1}^{p} I_0(h_i(x)),$$

where $I_- : \mathbf{R} \to \mathbf{R}$ is the indicator function for $-\mathbf{R}_+$, with $\mathbf{dom}\ I_- = -\mathbf{R}_+$ and its extended-value extension being

$$\tilde{I}_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases},$$

and similarly, I_0 is the indicator function of $\{0\}$, with $\mathbf{dom}\ I_0 = \{0\}$ and its extended-value extension being

$$\tilde{I}_0(u) = \begin{cases} 0 & u = 0 \\ \infty & u \neq 0 \end{cases}.$$

Linear approximation interpretation (2/3)

- Now, in the unconstrained problem

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x)),$$

  suppose we replace
  - the function $I_-(u)$ with the linear function $\lambda_i u$, where $\lambda_i \geq 0$,
  - and the function $I_0(u)$ with $\nu_i u$.
- Then, the objective becomes the Lagrangian function $L(x, \lambda, \nu)$, and the dual function value $g(\lambda, \nu)$ is the optimal value of the problem

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

Linear approximation interpretation (3/3)

- In the above formulation, we use a linear or "soft" displeasure function in place of I_- and I_0.
- For an inequality constraint (i.e., $f_i(x)$ for some $i > 0$), our displeasure is zero when $f_i(x) = 0$, and is positive when $f_i(x) > 0$ (assuming $\lambda_i > 0$); our displeasure grows as the constraint becomes "more violated". Further, we derive pleasure from constraints that have margin, i.e., from $f_i(x) < 0$.
- Although such a linear approximation of the indicator function $I_-(u)$ is rather poor, it is at least an underestimator of the indicator function.
- Since $\lambda_i u \leq I_-(u)$ and $\nu_i u \leq I_0(u)$ for all u, we see immediately that the dual function yields a lower bound on the optimal value of the original problem.

Least-squares solution of linear equations (1/2)

- Consider the problem

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

  where $A \in \mathbf{R}^{p \times n}$. This problem has no inequality constraints and p (linear) equality constraints.
- The Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$, with domain $\mathbf{R}^n \times \mathbf{R}^p$.
- The dual function is given by $g(\nu) = \inf_x L(x, \nu)$. Since $L(x, \nu)$ is a convex quadratic function of x, we can find the minimizing x from the optimality condition

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0,$$

  which yields $x = -(1/2)A^T \nu$.

Least-squares solution of linear equations (2/2)

- Therefore, the dual function is

$$g(\nu) = L(-(1/2)A^T\nu, \nu) = -(1/4)\nu^T AA^T\nu - b^T\nu,$$

which is a concave quadratic function, with domain $\mathbf{R}^p$.

- The lower bound property

$$g(\lambda, \nu) \le p^*$$

states that for any $\nu \in \mathbf{R}^p$, we have

$$-(1/4)\nu^T AA^T\nu - b^T\nu \le \inf\left\{x^T x \mid Ax = b\right\}.$$

Standard form LP (1/2)

- Consider an LP in standard form,

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

  which has inequality constraint functions $f_i(x) = -x_i, i = 1, ..., n$. Here $A \in \mathbf{R}^{p \times n}$.

- The Lagrangian is

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

Standard form LP (2/2)

- The dual function is

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

- Note that the dual function g is finite only on a proper affine subset of $\mathbf{R}^n \times \mathbf{R}^p$.
- The lower bound property $g(\lambda,\nu) \le p^*$ is nontrivial (i.e., $(\lambda,\nu) \in \mathbf{dom}\ g$) only when λ and ν satisfy $\lambda \succeq 0$ and $A^T\nu - \lambda + c = 0$.
- When this occurs, $-b^T\nu$ is a lower bound on the optimal value of the LP in standard form.

Two-way partitioning problem (1/3)

- We consider the (nonconvex) problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, i = 1, \ldots, n, \end{array}$$

  where $W \in \mathbf{S}^n$.
- The constraints restrict the values of x_i to 1 or -1, so the problem is equivalent to finding the vector x with components ± 1 that minimizes $x^T W x$.
- The feasible set here is finite (it contains 2^n points), so this problem can in principle be solved by simply checking the objective value of each feasible point.
  - However, for a large n, it is very difficult to solve.
- We can interpret the problem as a two-way partitioning problem on a set of n elements, say, $\{1, \ldots, n\}$: A feasible x corresponds to the partition

$$\{1, \ldots, n\} = \{i \mid x_i = -1\} \bigcup \{i \mid x_i = 1\}.$$

Two-way partitioning problem (2/3)

- The Lagrangian is

$$L(x,\nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \textbf{diag}\ (\nu)) x - \mathbf{1}^T \nu.$$

- We obtain the Lagrange dual function by minimizing over x:

$$\begin{aligned} g(\nu) &= \inf_x x^T (W + \textbf{diag}\ (\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu, & W + \textbf{diag}\ (\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}, \end{aligned}$$

where we use the fact that the infimum of a quadratic form $u^T A u$ is either zero (if $A \succeq 0$) or $-\infty$ (if $A \not\succeq 0$).

Two-way partitioning problem (3/3)

- This dual function

$$g(\nu) = \begin{cases} -\mathbf{1}^T\nu, & W + \textbf{diag}\ (\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

  provides lower bounds on the optimal value of the two-way partitioning problem.

- For example, we can take the specific value of the dual variable $\nu = -\lambda_{min}(W)\mathbf{1}$, which is dual feasible, since

$$W + \textbf{diag}\ (\nu) = W - \lambda_{min}(W)I \succeq 0.$$

  This yields the bound on the optimal value p^*

$$p^* \geq -\mathbf{1}^T\nu = n\lambda_{min}(W).$$

The Lagrange dual function and conjugate functions (1/2)

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

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

- The conjugate function and Lagrange dual function are closely related. As a simple example, consider the trivial problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x = 0. \end{array}$$

- This problem has Lagrangian $L(x, \nu) = f(x) + \nu^T x$, and dual function

$$g(\nu) = \inf_x \left(f(x) + \nu^T x \right) = -\sup_x \left((-\nu)^T x - f(x) \right) = -f^*(-\nu),$$

with $\mathbf{dom}\ g = -\mathbf{dom}\ f^*$.

The Lagrange dual function and conjugate functions (2/2)

- More generally, consider an optimization problem with linear inequality and equality constraints,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b \\ & Cx = d. \end{array}$$

- Using the conjugate of f_0 we can write the dual function for the problem as

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

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

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

Example – Equality constrained norm minimization

- Consider the problem

$$\begin{array}{ll} \text{minimize} & ||x||_2 \\ \text{subject to} & Ax = b. \end{array}$$

- The conjugate of $f_0 = ||\cdot||_2$ is given by

$$f_0^*(y) = \begin{cases} 0, & ||y||_2 \leq 1 \\ \infty, & \text{otherwise} \end{cases}.$$

- The dual function for the problem is given by

$$g(\nu) = -b^T\nu - f_0^*(-A^T\nu) = \begin{cases} -b^T\nu, & ||A^T\nu||_2 \leq 1 \\ -\infty, & \text{otherwise} \end{cases}.$$

The Lagrange dual problem (1/2)

- For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function $g(\lambda, \nu)$ gives us a lower bound on the optimal value p^* of the optimization problem (called the **primal problem**)

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, ..., m \\ & h_i(x) = 0, i = 1, ..., p. \end{aligned}$$

- The **Lagrange dual problem** associated with the original problem, defined as

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succeq 0, \end{aligned}$$

would give the best lower bound that can be obtained from the Lagrange dual function.

The Lagrange dual problem (2/2)

- We use the term **dual feasible** to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$.
- We refer to (λ^*, ν^*) as **dual optimal** or optimal Lagrange multipliers if they are optimal for the dual problem.
- The Lagrange dual problem is a convex optimization problem, since the objective to be maximized is concave and the constraint is convex, regardless of the convexity of the primal problem.

Making dual constraints explicit

- It is not uncommon for the domain of the dual function,

$$\textbf{dom}\ g = \{(\lambda, \nu) \mid g(\lambda, \nu) > -\infty\}\,,$$

  to have an affine dimension smaller than $m + p$.

- In many cases we can identify the affine hull of **dom** g, and describe it as a set of linear equality constraints. This means we can identify the equality constraints that are 'hidden' or 'implicit' in the objective g of the dual problem

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

- In this case we can form an equivalent problem, in which these equality constraints are given explicitly as constraints.

Example – Lagrange dual of standard form LP (1/2)

- The Lagrange dual function for the standard form LP

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

is given by

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

- Strictly speaking, the Lagrange dual problem of the standard form LP is to maximize this dual function g subject to $\lambda \succeq 0$:

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) = \begin{cases} -b^T\nu, & A^T\nu - \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

Example – Lagrange dual of standard form LP (2/2)

- Here g is finite only when $A^T\nu - \lambda + c = 0$. So the Lagrange dual problem is equivalent to (but not the same as)

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

- This problem, in turn, can be expressed as

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

  which is an LP in inequality form.
- With some abuse of terminology, the above two problems are also referred to as the Lagrange dual of the standard form LP.

Example – Lagrange dual of inequality form LP (1/2)

- Conversely, let's find the Lagrange dual problem of an LP in inequality form

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

- The Lagrangian is

$$L(x, \lambda) = c^T x + \lambda^T (Ax - b) = -b^T \lambda + (A^T \lambda + c)^T x,$$

so the dual function is

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x.$$

- The infimum of a linear function is $-\infty$, except in the special case when it is identically zero, so the dual function is

$$g(\lambda) = \begin{cases} -b^T \lambda, & A^T \lambda + c = 0 \\ -\infty, & \text{otherwise.} \end{cases}$$

Example – Lagrange dual of inequality form LP (2/2)

- The dual variable λ is dual feasible if $\lambda \succeq 0$ and $A^T\lambda + c = 0$.
- So, the dual problem can be written as

$$\begin{array}{ll} \text{maximize} & -b^T\lambda \\ \text{subject to} & A^T\lambda + c = 0 \\ & \lambda \succeq 0, \end{array}$$

  which is an LP in standard form.
- Note that the dual of a standard form LP is an inequality form LP (i.e., LP with only inequality constraints), and vice versa.

Weak Duality (1/3)

- The optimal value of the Lagrange dual problem, which we denote d^*, is the best lower bound on p^* that can be obtained from the Lagrange dual function.
- Therefore, we have the simple but important inequality

$$d^* \leq p^*,$$

which holds even if the original problem is not convex. This property is called **weak duality**.
- The weak duality inequality holds even when d^* and p^* are infinite.
    - For example, if the primal problem is unbounded below, so that $p^* = -\infty$, we must have $d^* = -\infty$, i.e., the Lagrange dual problem is infeasible.
    - Conversely, if the dual problem is unbounded above, so that $d^* = \infty$, we must have $p^* = \infty$, i.e., the primal problem is infeasible.

Weak Duality (2/3)

- The difference $p^* - d^*$ is called the **optimal duality gap** of the original problem, which is always nonnegative.
- The bound can sometimes be used to find a lower bound on the optimal value of a problem that is difficult to solve, since the dual problem is always convex, and in many cases can be solved efficiently, to find d^*.
  - As an example, consider the two-way partitioning problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, i = 1, \ldots, n, \end{array}$$

    where $W \in \mathbf{S}^n$.
  - The dual problem is a semidefinite program (SDP),

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}\,(\nu) \succeq 0, \end{array}$$

    with variable $\nu \in \mathbf{R}^n$.

Weak Duality (3/3)

- This problem can be solved efficiently, even for relatively large values of n, such as $n = 1000$.
- The optimal value of the dual problem

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}\,(\nu) \succeq 0, \end{array}$$

is a lower bound on the optimal value of the two-way partitioning problem, and is always at least as good as the lower bound given as

$$p^* \geq d^* \geq n\lambda_{min}(W).$$

Strong duality

- If the equality $d^* = p^*$ holds, i.e., the optimal duality gap is zero, then we say that **strong duality** holds.
- This means that the best bound that can be obtained from the Lagrange dual function is tight.
- Strong duality does not always hold. But if the primal problem is convex, i.e., of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & Ax = b, \end{aligned}$$

with $f_0, \dots, f_m$ convex, we usually (but not always) have strong duality.

Slater's Condition

Slater's Condition

Consider a convex optimization problem

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b, \end{aligned}$$

with the domain $\mathcal{D} = (\bigcap_{i=0}^{m} \textbf{dom } f_i) \subseteq \mathbf{R}^n$. The problem is said to satisfy the **Slater's condition** if there exists an $x \in \textbf{relint } \mathcal{D} \subseteq \mathbf{R}^n$ such that

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

Such a point is sometimes called strictly feasible.

Slater's Theorem

Slater's Condition

A convex optimization problem is said to satisfy the **Slater's condition** if there exists an $x \in \textbf{relint}\ \mathcal{D} \subseteq \mathbf{R}^n$ such that

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

Such a point is sometimes called strictly feasible.

Slater's Theorem

If the problem is convex and Slater's condition holds, then the strong duality holds.

Refined Slater's Condition

- Slater's condition can be refined when some of the inequality constraint functions f_i are affine.

Refined (Weaker) Slater's Condition

Suppose the first k constraint functions $f_1, \dots, f_k$ are affine in an optimization problem. Then the problem is said to satisfy the **refined (weak) Slater's condition** if there exists an $x \in \textbf{relint } \mathcal{D} \subseteq \mathbf{R}^n$ such that

$$f_i(x) \leq 0, i = 1, \dots, k,$$

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

(Refined) Slater's Theorem

If the problem is convex and the refined Slater's condition holds, then the strong duality holds.

Strong duality and Slater's Condition

- Note that the refined (weaker) Slater's condition reduces to feasibility when
  - the constraints are all linear equalities and inequalities, and,
  - **dom** f_0 is open.
- Slater's condition not only implies strong duality for convex problems. It also implies that the dual optimal value is attained when $d^* > -\infty$, i.e., there exists a dual feasible (λ^*, ν^*) with $g(\lambda^*, \nu^*) = d^* = p^*$.
- We will prove later on that strong duality obtains when the primal problem is convex and Slater's condition holds.

Examples – Least-squares solution for linear equations

- Recall the least-squares problem

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

  whose associated dual problem is

$$\text{maximize} \quad -(1/4)\nu^T AA^T \nu - b^T \nu,$$

  an unconstrained concave quadratic maximization problem.
- We always have strong duality. Slater's condition is simply that the primal problem is feasible, so $p^* = d^*$ provided $b \in \mathcal{R}(A)$, i.e., $p^* < \infty$.

Lagrange dual of LP

- By the weaker form of Slater's condition, we find that strong duality holds for any LP (in standard or inequality form) provided the primal problem is feasible (i.e., $p^* < \infty$).
- Applying this result to the dual problem, we conclude that strong duality holds for LPs if the dual problem is feasible (i.e., $d^* > -\infty$).
- This leaves only one possible situation in which strong duality for LPs can fail: both the primal and dual problems are infeasible.
  - Example:

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} x \preceq \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \end{array}$$

Lagrange dual of QCQP (1/2)

- We consider the QCQP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0, i = 1, \ldots, m, \end{array}$$

with $P_0 \in \mathbf{S}^n_{++}$, and $P_i \in \mathbf{S}^n_{+}, i = 1, \ldots, m$.

- The Lagrangian is

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

where

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

Lagrange dual of QCQP (2/2)

- If $\lambda \succeq 0$, we have $P(\lambda) \succ 0$ and the dual function can be written as

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

- We can therefore express the dual problem as

$$\begin{array}{ll} \text{maximize} & -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

- The Slater's condition says that strong duality holds if the quadratic inequality constraints are strictly feasible, i.e., there exists an x with

$$(1/2)x^T P_i x + q_i^T x + r_i < 0, i = 1, \ldots, m.$$

Lagrange Dual of Inequality-form SDP

- Consider the inequality-form SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 A_1 + \cdots + x_n A_n \preceq B, \end{array}$$

  with variable $x \in \mathbf{R}^n$, and parameters $B, A_1, \ldots, A_n \in \mathbf{S}^k$, $c \in \mathbf{R}^n$.
- What is the Lagrange dual problem of the problem above?

Weak and strong duality via set of values (1/5)

In this section, we will prove a special case of the Slater's theorem via a geometric interpretation of optimization problems.

- We first consider any given optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p. \end{array}$$

  For this problem, we define the set

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

- Then, the optimal value p^* of the problem is easily expressed in terms of $\mathcal{G}$ as

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

Weak and strong duality via set of values (2/5)

- The value of the dual function at (λ, ν) can be written as

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

by minimizing the affine function

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

over $(u, v, t) \in \mathcal{G}$.

- In particular, we see that if the infimum is finite, then the inequality

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

defines a supporting hyperplane to $\mathcal{G}$. This is sometimes referred to as a nonvertical supporting hyperplane, because the last component of the normal vector is nonzero.

Weak and strong duality via set of values (3/5)

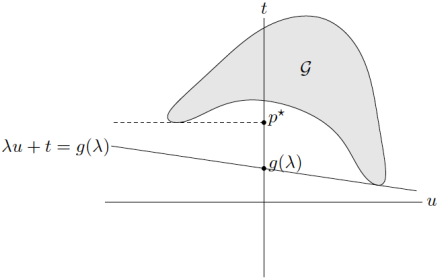
- Now suppose $\lambda \succeq 0$. Then, $t \geq (\lambda, \nu, 1)^T(u, v, t)$ if $u \preceq 0$ and $v = 0$.
- Therefore,

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

i.e., we have weak duality.

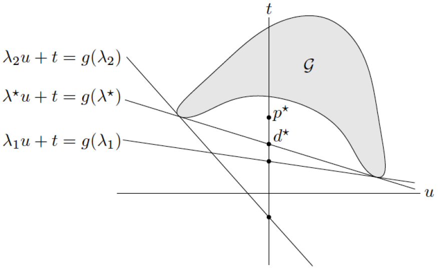
Weak and strong duality via set of values (4/5)

- Consider a problem with one (inequality) constraint. Given λ, we minimize $(\lambda, 1)^T(u, t)$ over $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$.
- This yields a supporting hyperplane with slope $-\lambda$. The intersection of this hyperplane with the $u = 0$ axis gives $g(\lambda)$.



Weak and strong duality via set of values (5/5)

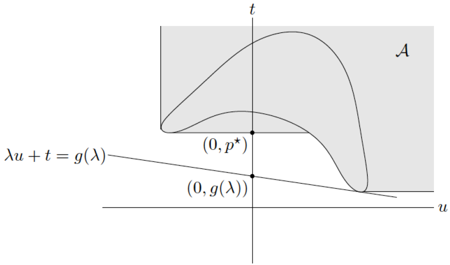
- The figure below depicts supporting hyperplanes corresponding to three dual feasible values of λ, including the optimum λ^*. Strong duality does not hold; the optimal duality gap $p^* - d^*$ is positive.



Epigraph variation (1/3)

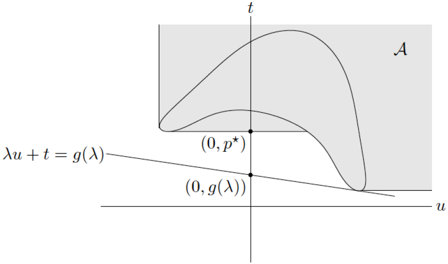
- We define the set $\mathcal{A} \subseteq \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R}$ as

$$\begin{aligned}\mathcal{A} &= \mathcal{G} + \left(\mathbf{R}_+^m \times \{0\} \times \mathbf{R}_+\right) \\ &= \{(u, v, t) \mid \exists x \in \mathcal{D},\ f_i(x) \leq u_i, i = 1, \ldots, m, \\ &\quad h_i(x) = v_i, i = 1, \ldots, p, \quad f_0(x) \leq t\}.\end{aligned}$$



- The set $\mathcal{A}$ is considered as an epigraph-like form of $\mathcal{G}$.

Epigraph variation (2/3)



- We can express the optimal value in terms of $\mathcal{A}$ as

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

- The dual function at a point (λ, ν) with $\lambda \succeq 0$ is

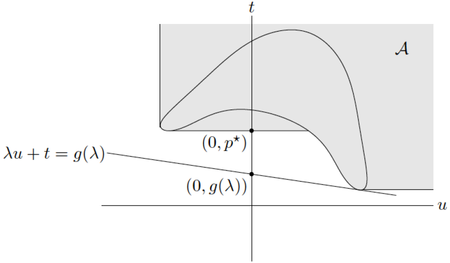
$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf \left\{ (\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in \mathcal{A} \right\}.$$

Epigraph variation (3/3)

- If $g(\lambda, \nu)$ is finite, then

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

defines a nonvertical supporting hyperplane to $\mathcal{A}$.



- In particular, since $(0, 0, p^*) \in \mathbf{bd}\ \mathcal{A}$, we have

$$p^* = (\lambda, \nu, 1)^T(0, 0, p^*) \geq g(\lambda, \nu),$$

the weak duality lower bound.
- Strong duality holds if and only if the equality holds for some dual feasible (λ, ν).

Strong Duality Under Slater's Constraint Qualification

- We want to prove that Slater's constraint qualification, i.e., there exists an $\tilde{x} \in \mathbf{relint}\ \mathcal{D}$ such that

$$f_i(\tilde{x}) < 0, \quad i = 1, \dots, m, \qquad A\tilde{x} = b,$$

  guarantees strong duality (and that the dual optimum is attained) for a convex problem.
- The following assumptions are made to simplify the proof:
  - $\mathcal{D}$ has nonempty interior (hence, $\mathbf{relint}\ \mathcal{D} = \mathbf{int}\ \mathcal{D}$).
  - $\mathrm{rank}\ A = p$. (Note that $A \in \mathbf{R}^{p \times n}$.)
  - p^* is finite. (Since there is a feasible point, we can only have $p^* = -\infty$ or p^* finite; if $p^* = -\infty$, then $d^* = -\infty$ by weak duality.)

The Proof (1/5)

- The set $\mathcal{A}$ defined as

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

is convex if the underlying problem is convex.

  - Proof idea: Any two points in $\mathcal{A}$ can be expressed as $(f(x_1) + r_1, h(x_1), f_0(x_1) + s_1)$ and $(f(x_2) + r_2, h(x_2), f_0(x_2) + s_2)$, where $f : \mathbf{R}^n \to \mathbf{R}^m$ with its ith component being $f_i(x)$ (convex) and $h : \mathbf{R}^n \to \mathbf{R}^p$ with its ith component being $h_i(x)$ (affine), and $r_1, r_2 \succeq 0, s_1, s_2 \geq 0$.

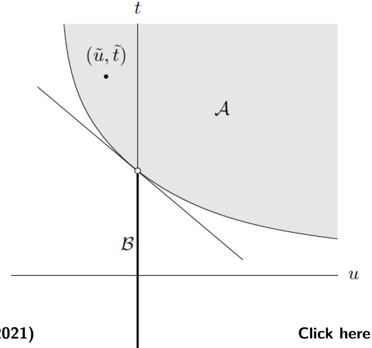
- We define a second convex set $\mathcal{B}$ as

$$\mathcal{B} = \left\{(0, 0, s) \in \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R} \mid s < p^*\right\}.$$

- We have that $\mathcal{A} \cap \mathcal{B} = \phi$.

The Proof (2/5)

- To see that $\mathcal{A} \cap \mathcal{B} = \phi$, suppose $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$. Since $(u, v, t) \in \mathcal{B}$ we have $u = 0$, $v = 0$, and $t < p^*$.
- Since $(u, v, t) \in \mathcal{A}$, there exists an $x \in \mathcal{D}$ with $f_i(x) \leq u_i = 0, i = 1, \ldots, m$, $Ax - b = v = 0$, and $f_0(x) \leq t < p^*$, which is impossible since p^* is the optimal value of the primal problem.



The Proof (3/5)

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

$$(u, v, t) \in \mathcal{A} \Longrightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq \alpha,$$

and

$$(u, v, t) \in \mathcal{B} \Longrightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq \alpha.$$

- The first condition implies that $\tilde{\lambda} \succeq 0$ and $\mu \geq 0$.
- The second condition means that $\mu t \leq \alpha$ for all $t < p^*$, and hence, $\mu p^* \leq \alpha$ *(Hint: suppose the otherwise: $\mu p^* \leq \alpha$ is wrong, then some contradiction can be found).*
- Therefore, we conclude that for any $x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^*.$$

The Proof (4/5)

- For any $x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) + \mu f_0(x) \geq \alpha \geq \mu p^*.$$

- Assume that $\mu > 0$. Then

$$\mu L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) = \sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) + \mu f_0(x) \geq \mu p^*$$

  for all $x \in \mathcal{D}$, from which it follows, by minimizing over x, that $g(\lambda, \nu) \geq p^*$, where we define $\lambda = \tilde{\lambda}/\mu, \nu = \tilde{\nu}/\mu$.
- By weak duality we have $g(\lambda, \nu) \leq p^*$, so in fact $g(\lambda, \nu) = p^*$. This shows that strong duality holds, and that the dual optimum is attained, when $\mu > 0$.

The Proof (5/5)

- Now consider the case $\mu = 0$. We have that for all $x \in \mathcal{D}$,

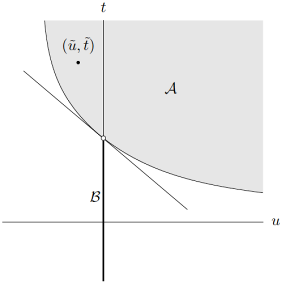
$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \geq 0.$$

- For any point $\tilde{x}$ that satisfies the Slater's condition, we have

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

  - Since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$, we obtain that $\tilde{\lambda} = 0$. Since $\tilde{\lambda} = 0, \mu = 0$ but $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$, we conclude that $\tilde{\nu} \neq 0$.
- Then we have that for all $x \in \mathcal{D}$, $\tilde{\nu}^T (Ax - b) \geq 0$.
  - But $\tilde{x}$ satisfies $\tilde{\nu}^T (A\tilde{x} - b) = 0$, and since $\tilde{x} \in \textbf{int } \mathcal{D}$, there are points in $\mathcal{D}$ with $\tilde{\nu}^T (Ax - b) < 0$ unless $A^T \tilde{\nu} = 0$. This contradicts the assumption that rank $A = p$.

Geometric idea behind the proof



- The above figure illustrates a simple problem with one inequality constraint.
- The hyperplane separating $\mathcal{A}$ and $\mathcal{B}$ defines a supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$.
- Slater's constraint qualification is used to establish that the hyperplane must be nonvertical (i.e., has a normal vector of the form $(\lambda^*, 1)$).

An example in which Slater's condition does not hold

- Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq 0 \end{array}$$

  with variables x and y, and domain $\mathcal{D} = \{(x, y) \mid y > 0\}$.
- It is a convex optimization problem with the optimal value 1.
- Slater's condition does not hold for this problem since there does not exist $(x, y) \in \textbf{int}\, \mathcal{D}$ s.t. $x^2/y < 0$.
- The Lagrange dual function is

$$g(\lambda) = \inf_{x,y} L(x, y, \lambda) = \inf_{x,y} \left(e^{-x} + \lambda(x^2/y)\right) = \begin{cases} 0, & \lambda \geq 0 \\ -\infty, & \lambda < 0 \end{cases}.$$

- The optimal solution and optimal value of the dual problem is $\lambda^* \in [0, \infty)$ and $d^* = 0$, respectively.
- The optimal duality gap is $p^* - d^* = 1$.

Summary of dual problems and Slater's condition

- Given an optimization problem (convex or nonconvex) with an optimal value p^*, we defined
  - The Lagrangian $L(x, \lambda, \nu)$.
  - The dual function $g(\lambda, \nu)$.
  - The dual problem.
- Weak duality: The optimal value of the dual problem is always a lower bound of the optimal value of the primal problem: $d^* \leq p^*$.
- Strong duality: When certain conditions are met, we will have $d^* = p^*$. A sufficient (but not necessary) condition for strong duality is
  - The primal problem is convex.
  - The primal problem meets the Slater's condition.