

CS675: Convex and Combinatorial Optimization  
Fall 2019  
Duality of Convex Optimization Problems

Instructor: Shaddin Dughmi

# Outline

- 1 The Lagrange Dual Problem
- 2 Duality
- 3 Optimality Conditions

## Recall: Optimization Problem in Standard Form

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

- For convex optimization problems in standard form,  $f_i$  is convex and  $h_i$  is affine.
- Let  $\mathcal{D}$  denote the domain of all these functions (i.e. when their value is finite)

# Recall: Optimization Problem in Standard Form

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

- For convex optimization problems in standard form,  $f_i$  is convex and  $h_i$  is affine.
- Let  $\mathcal{D}$  denote the domain of all these functions (i.e. when their value is finite)

## This Lecture + Next

We will develop duality theory for convex optimization problems, generalizing linear programming duality.

# Running Example: Linear Programming

We have already seen the standard form LP below

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} -\text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

# Running Example: Linear Programming

We have already seen the standard form LP below

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} -\text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

Along the way, we will recover the following standard form dual

$$\begin{array}{ll} \text{minimize} & y^\top b \\ \text{subject to} & A^\top y \succeq c \\ & y \succeq 0 \end{array}$$

# The Lagrangian

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

Basic idea of Lagrangian duality is to relax/soften the constraints by replacing each with a linear “penalty term” or “cost” in the objective.

# The Lagrangian

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

Basic idea of Lagrangian duality is to relax/soften the constraints by replacing each with a linear “penalty term” or “cost” in the objective.

## The Lagrangian Function

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

- $\lambda_i$  is **Lagrange Multiplier** for  $i$ 'th inequality constraint
  - Required to be nonnegative
- $\nu_i$  is **Lagrange Multiplier** for  $i$ 'th equality constraint
  - Allowed to be of arbitrary sign



# The Lagrange Dual Function

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

The Lagrange dual function gives the optimal value of the primal problem subject to the softened constraints

# The Lagrange Dual Function

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

The Lagrange dual function gives the optimal value of the primal problem subject to the softened constraints

## The Lagrange Dual Function

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

- Observe:  $g$  is a concave function of the Lagrange multipliers
- We will see: Its quite common for the Lagrange dual to be unbounded ( $-\infty$ ) for some  $\lambda$  and  $\nu$
- By convention, domain of  $g$  is  $(\lambda, \nu)$  s.t.  $g(\lambda, \nu) > -\infty$

# Lagrange Dual of LP

$$\begin{array}{ll} \text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

First, the Lagrangian function

$$\begin{aligned} L(x, \lambda) &= -c^\top x + \lambda_1^\top (Ax - b) - \lambda_2^\top x \\ &= (A^\top \lambda_1 - c - \lambda_2)^\top x - \lambda_1^\top b \end{aligned}$$

# Langrange Dual of LP

$$\begin{array}{ll} \text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

First, the Lagrangian function

$$\begin{aligned} L(x, \lambda) &= -c^\top x + \lambda_1^\top (Ax - b) - \lambda_2^\top x \\ &= (A^\top \lambda_1 - c - \lambda_2)^\top x - \lambda_1^\top b \end{aligned}$$

And the Lagrange Dual

$$\begin{aligned} g(\lambda) &= \inf_x L(x, \lambda) \\ &= \begin{cases} -\infty & \text{if } A^\top \lambda_1 - c - \lambda_2 \neq 0 \\ -\lambda_1^\top b & \text{if } A^\top \lambda_1 - c - \lambda_2 = 0 \end{cases} \end{aligned}$$

# Langrange Dual of LP

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

First, the Lagrangian function

$$\begin{aligned} L(x, \lambda) &= -c^T x + \lambda_1^T (Ax - b) - \lambda_2^T x \\ &= (A^T \lambda_1 - c - \lambda_2)^T x - \lambda_1^T b \end{aligned}$$

And the Lagrange Dual

$$\begin{aligned} g(\lambda) &= \inf_x L(x, \lambda) \\ &= \begin{cases} -\infty & \text{if } A^T \lambda_1 - c - \lambda_2 \neq 0 \\ -\lambda_1^T b & \text{if } A^T \lambda_1 - c - \lambda_2 = 0 \end{cases} \end{aligned}$$

So we restrict the domain of  $g$  to  $\lambda$  satisfying  $A^T \lambda_1 - c - \lambda_2 = 0$

# Interpretation: “Soft” Lower Bound

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

## The Lagrange Dual Function

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

# Interpretation: “Soft” Lower Bound

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

## The Lagrange Dual Function

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

## Fact

$g(\lambda, \nu)$  is a lowerbound on  $\text{OPT}(\text{primal})$  for every  $\lambda \succeq 0$  and  $\nu \in \mathbb{R}^k$ .

# Interpretation: “Soft” Lower Bound

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

## The Lagrange Dual Function

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

## Fact

$g(\lambda, \nu)$  is a lowerbound on  $\text{OPT}(\text{primal})$  for every  $\lambda \succeq 0$  and  $\nu \in \mathbb{R}^k$ .

## Proof

- Every primal feasible  $x$  incurs nonpositive penalty by  $L(x, \lambda, \nu)$
- Therefore,  $L(x^*, \lambda, \nu) \leq f_0(x^*)$
- So  $g(\lambda, \nu) \leq f_0(x^*) = \text{OPT}(\text{Primal})$



# Interpretation: “Soft” Lower Bound

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

## The Lagrange Dual Function

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

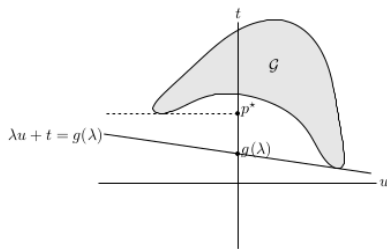
## Interpretation

- A “hard” feasibility constraint can be thought of as imposing a penalty of  $+\infty$  if violated, and a penalty/reward of 0 if satisfied
- Lagrangian imposes a “soft” linear penalty for violating a constraint, and a reward for slack
- Lagrange dual finds the optimal subject to these soft constraints

# Interpretation: Geometric

Most easily visualized in the presence of a single inequality constraint

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$

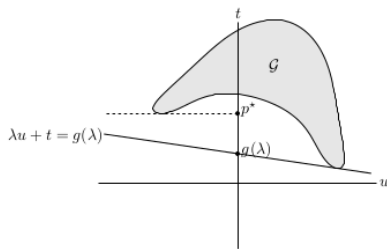


- Let  $\mathcal{G}$  be attainable constraint/objective function value tuples
  - i.e.  $(u, t) \in \mathcal{G}$  if there is an  $x$  such that  $f_1(x) = u$  and  $f_0(x) = t$
- $p^* = \inf \{t : (u, t) \in \mathcal{G}, u \leq 0\}$
- $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{G}\}$

# Interpretation: Geometric

Most easily visualized in the presence of a single inequality constraint

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$

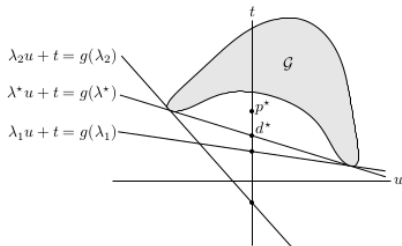


- Let  $\mathcal{G}$  be attainable constraint/objective function value tuples
    - i.e.  $(u, t) \in \mathcal{G}$  if there is an  $x$  such that  $f_1(x) = u$  and  $f_0(x) = t$
  - $p^* = \inf \{t : (u, t) \in \mathcal{G}, u \leq 0\}$
  - $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{G}\}$
- $\lambda u + t = g(\lambda)$  is a supporting hyperplane to  $\mathcal{G}$  pointing northeast
- Must intersect vertical axis below  $p^*$
- Therefore  $g(\lambda) \leq p^*$

# The Lagrange Dual Problem

This is the problem of finding the best lower bound on  $\text{OPT}(\text{primal})$  implied by the Lagrange dual function

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



- Note: this is a convex optimization problem, regardless of whether primal problem was convex
- By convention, sometimes we add “dual feasibility” constraints to impose “nontrivial” lowerbounds (i.e.  $g(\lambda, \nu) \geq -\infty$ )
- $(\lambda^*, \nu^*)$  solving the above are referred to as the **dual optimal** solution

# Lagrange Dual Problem of LP

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} -\text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

## Recall

Our Lagrange dual function for the above minimization LP (to the right), defined over the domain  $A^\top \lambda_1 - c - \lambda_2 = 0$ .

$$g(\lambda) = -\lambda_1^\top b$$

# Lagrange Dual Problem of LP

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} -\text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

## Recall

Our Lagrange dual function for the above minimization LP (to the right), defined over the domain  $A^\top \lambda_1 - c - \lambda_2 = 0$ .

$$g(\lambda) = -\lambda_1^\top b$$

The Lagrange dual problem can then be written as

$$\begin{array}{ll} -\text{maximize} & -\lambda_1^\top b \\ \text{subject to} & A^\top \lambda_1 - c - \lambda_2 = 0 \end{array}$$

$$\lambda \succeq 0$$

# Langrange Dual Problem of LP

$$\begin{array}{ll} \text{maximize} & c^\top x \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0 \end{array}$$

$$\begin{array}{ll} -\text{minimize} & -c^\top x \\ \text{subject to} & Ax - b \preceq 0 \\ & -x \preceq 0 \end{array}$$

## Recall

Our Lagrange dual function for the above minimization LP (to the right), defined over the domain  $A^\top \lambda_1 - c - \lambda_2 = 0$ .

$$g(\lambda) = -\lambda_1^\top b$$

The Lagrange dual problem can then be written as

$$\begin{array}{ll} -\text{maximize} & -\lambda_1^\top b \\ \text{subject to} & \cancel{A^\top \lambda_1 - c - \lambda_2 = 0} \\ & A^\top \lambda_1 \succeq c \\ & \lambda \succeq 0 \end{array}$$

# Langrange Dual Problem of LP

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

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

## Recall

Our Lagrange dual function for the above minimization LP (to the right), defined over the domain  $A^T \lambda_1 - c - \lambda_2 = 0$ .

$$g(\lambda) = -\lambda_1^T b$$

The Lagrange dual problem can then be written as

$$\begin{array}{ll} \text{minimize} & y^T b \\ \text{subject to} & A^T y \succeq c \\ & y \succeq 0 \end{array}$$

$$\begin{array}{ll} \text{--maximize} & -\lambda_1^T b \\ \text{subject to} & \cancel{A^T \lambda_1 - c - \lambda_2 = 0} \\ & A^T \lambda_1 \succeq c \\ & \lambda \succeq 0 \end{array}$$



# Another Example: Conic Optimization Problem

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

- $x \in K$  can equivalently be written as  $z^\top x \leq 0, \forall z \in K^\circ$

$$\begin{aligned} L(x, \lambda, \nu) &= c^\top x + \nu^\top (Ax - b) + \sum_{z \in K^\circ} \lambda_z \cdot z^\top x \\ &= (c - A^\top \nu + \sum_{z \in K^\circ} \lambda_z \cdot z)^\top x + \nu^\top b \end{aligned}$$

# Another Example: Conic Optimization Problem

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

- $x \in K$  can equivalently be written as  $z^\top x \leq 0, \forall z \in K^\circ$

$$\begin{aligned} L(x, \lambda, \nu) &= c^\top x + \nu^\top (Ax - b) + \sum_{z \in K^\circ} \lambda_z \cdot z^\top x \\ &= (c - A^\top \nu + \sum_{z \in K^\circ} \lambda_z \cdot z)^\top x + \nu^\top b \end{aligned}$$

- Can think of  $\lambda \succeq 0$  as choosing some  $s \in K^\circ$

$$L(x, s, \nu) = (c - A^\top \nu + s)^\top x + \nu^\top b$$

# Another Example: Conic Optimization Problem

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && Ax = b \\ &&& x \in K \end{aligned}$$

- $x \in K$  can equivalently be written as  $z^\top x \leq 0, \forall z \in K^\circ$

$$\begin{aligned} L(x, \lambda, \nu) &= c^\top x + \nu^\top (Ax - b) + \sum_{z \in K^\circ} \lambda_z \cdot z^\top x \\ &= (c - A^\top \nu + \sum_{z \in K^\circ} \lambda_z \cdot z)^\top x + \nu^\top b \end{aligned}$$

- Can think of  $\lambda \succeq 0$  as choosing some  $s \in K^\circ$

$$L(x, s, \nu) = (c - A^\top \nu + s)^\top x + \nu^\top b$$

- Lagrange dual function  $g(s, \nu)$  is bounded when coefficient of  $x$  is zero, in which case it has value  $\nu^\top b$

## Another Example: Conic Optimization Problem

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

$$\begin{array}{ll} \text{maximize} & \nu^\top b \\ \text{subject to} & A^\top \nu - c \in K^\circ \end{array}$$

- $x \in K$  can equivalently be written as  $z^\top x \leq 0, \forall z \in K^\circ$

$$\begin{aligned} L(x, \lambda, \nu) &= c^\top x + \nu^\top (Ax - b) + \sum_{z \in K^\circ} \lambda_z \cdot z^\top x \\ &= (c - A^\top \nu + \sum_{z \in K^\circ} \lambda_z \cdot z)^\top x + \nu^\top b \end{aligned}$$

- Can think of  $\lambda \succeq 0$  as choosing some  $s \in K^\circ$

$$L(x, s, \nu) = (c - A^\top \nu + s)^\top x + \nu^\top b$$

- Lagrange dual function  $g(s, \nu)$  is bounded when coefficient of  $x$  is zero, in which case it has value  $\nu^\top b$

# Outline

1 The Lagrange Dual Problem

**2 Duality**

3 Optimality Conditions

# Weak Duality

## Primal Problem

$$\min f_0(x)$$

s.t.

$$f_i(x) \leq 0, \quad \forall i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \forall i = 1, \dots, k.$$

## Dual Problem

$$\max g(\lambda, \nu)$$

s.t.

$$\lambda \succeq 0$$

# Weak Duality

## Primal Problem

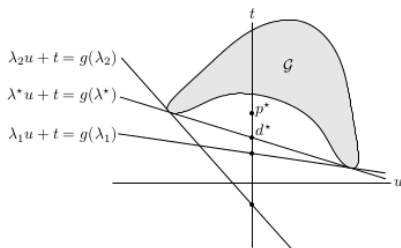
$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

## Weak Duality

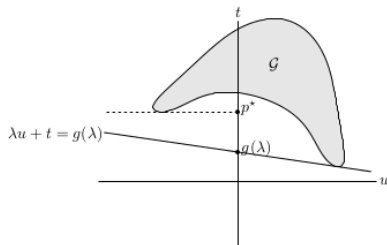
$$OPT(dual) \leq OPT(primal).$$



- We have already argued holds for every optimization problem
- **Duality Gap**: difference between optimal dual and primal values

# Recall: Geometric Interpretation of Weak Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$

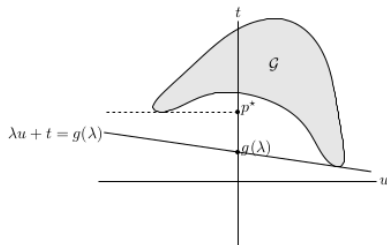


- Let  $\mathcal{G}$  be attainable constraint/objective function value tuples
  - i.e.  $(u, t) \in \mathcal{G}$  if there is an  $x$  such that  $f_1(x) = u$  and  $f_0(x) = t$
- $p^* = \inf \{t : (u, t) \in \mathcal{G}, u \leq 0\}$
- $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{G}\}$



# Recall: Geometric Interpretation of Weak Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



- Let  $\mathcal{G}$  be attainable constraint/objective function value tuples
  - i.e.  $(u, t) \in \mathcal{G}$  if there is an  $x$  such that  $f_1(x) = u$  and  $f_0(x) = t$
- $p^* = \inf \{t : (u, t) \in \mathcal{G}, u \leq 0\}$
- $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{G}\}$

## Fact

The equation  $\lambda u + t = g(\lambda)$  defines a supporting hyperplane to  $\mathcal{G}$ , intersecting  $t$  axis at  $g(\lambda) \leq p^*$ .

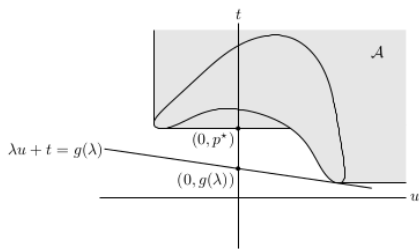
## Strong Duality

We say strong duality holds if  $OPT(dual) = OPT(primal)$ .

- Equivalently: there exists a setting of Lagrange multipliers so that  $g(\lambda, \nu)$  gives a tight lowerbound on primal optimal value.
- In general, does not hold for non-convex optimization problems
- Usually, but not always, holds for convex optimization problems.
  - Mild assumptions, such as **Slater's condition**, needed.

# Geometric Proof of Strong Duality

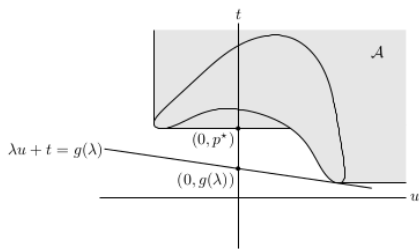
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



- Let  $\mathcal{A}$  be everything northeast (i.e. “worse”) than  $\mathcal{G}$ 
  - i.e.  $(u, t) \in \mathcal{A}$  if there is an  $x$  such that  $f_1(x) \leq u$  and  $f_0(x) \leq t$
- $p^* = \inf \{t : (0, t) \in \mathcal{A}\}$
- $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{A}\}$

# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



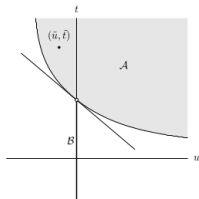
- Let  $\mathcal{A}$  be everything northeast (i.e. “worse”) than  $\mathcal{G}$ 
  - i.e.  $(u, t) \in \mathcal{A}$  if there is an  $x$  such that  $f_1(x) \leq u$  and  $f_0(x) \leq t$
- $p^* = \inf \{t : (0, t) \in \mathcal{A}\}$
- $g(\lambda) = \inf \{\lambda u + t : (u, t) \in \mathcal{A}\}$

## Fact

The equation  $\lambda u + t = g(\lambda)$  defines a supporting hyperplane to  $\mathcal{A}$ , intersecting  $t$  axis at  $g(\lambda) \leq p^*$ .

# Geometric Proof of Strong Duality

minimize  $f_0(x)$   
subject to  $f_1(x) \leq 0$

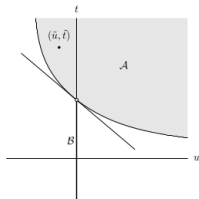


## Fact

When  $f_0$  and  $f_1$  are convex,  $\mathcal{A}$  is convex.

# Geometric Proof of Strong Duality

minimize  $f_0(x)$   
subject to  $f_1(x) \leq 0$



## Fact

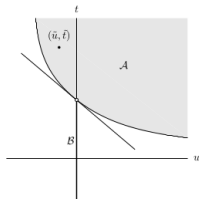
When  $f_0$  and  $f_1$  are convex,  $\mathcal{A}$  is convex.

## Proof

- Assume  $(u, t)$  and  $(u', t')$  are in  $\mathcal{A}$

# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



## Fact

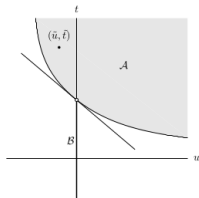
When  $f_0$  and  $f_1$  are convex,  $\mathcal{A}$  is convex.

## Proof

- Assume  $(u, t)$  and  $(u', t')$  are in  $\mathcal{A}$
- $\exists x, x'$  with  $(f_1(x), f_0(x)) \leq (u, t)$  and  $(f_1(x'), f_0(x')) \leq (u', t')$ .

# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



## Fact

When  $f_0$  and  $f_1$  are convex,  $\mathcal{A}$  is convex.

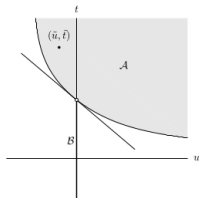
## Proof

- Assume  $(u, t)$  and  $(u', t')$  are in  $\mathcal{A}$
- $\exists x, x'$  with  $(f_1(x), f_0(x)) \leq (u, t)$  and  $(f_1(x'), f_0(x')) \leq (u', t')$ .
- By Jensen's inequality  
 $(f_1(\alpha x + (1-\alpha)x'), f_0(\alpha x + (1-\alpha)x')) \leq (\alpha u + (1-\alpha)u', \alpha t + (1-\alpha)t')$



# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



## Fact

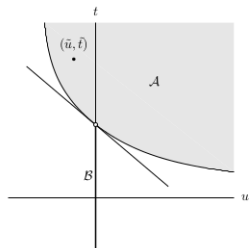
When  $f_0$  and  $f_1$  are convex,  $\mathcal{A}$  is convex.

## Proof

- Assume  $(u, t)$  and  $(u', t')$  are in  $\mathcal{A}$
- $\exists x, x'$  with  $(f_1(x), f_0(x)) \leq (u, t)$  and  $(f_1(x'), f_0(x')) \leq (u', t')$ .
- By Jensen's inequality  
 $(f_1(\alpha x + (1-\alpha)x'), f_0(\alpha x + (1-\alpha)x')) \leq (\alpha u + (1-\alpha)u', \alpha t + (1-\alpha)t')$
- Therefore, segment connecting  $(u, t)$  and  $(u', t')$  also in  $\mathcal{A}$ .

# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$

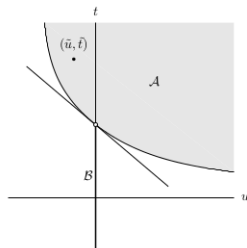


## Theorem (Informal)

There is a choice of  $\lambda$  so that  $g(\lambda) = p^*$ . Therefore, strong duality holds.

# Geometric Proof of Strong Duality

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



## Theorem (Informal)

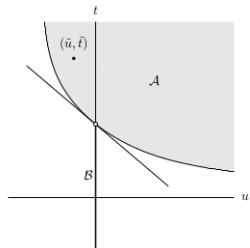
There is a choice of  $\lambda$  so that  $g(\lambda) = p^*$ . Therefore, strong duality holds.

## Proof

- Recall  $(0, p^*)$  is on the boundary of  $\mathcal{A}$
- By the supporting hyperplane theorem, there is a supporting hyperplane to  $\mathcal{A}$  at  $(0, p^*)$
- Direction of the supporting hyperplane gives us an appropriate  $\lambda$

# I Lied (A little)

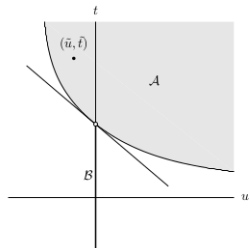
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



- In our proof, we ignored a technicality that can prevent strong duality from holding.

# I Lied (A little)

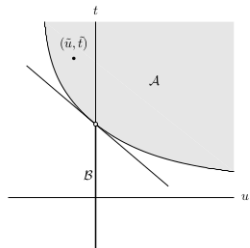
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



- In our proof, we ignored a technicality that can prevent strong duality from holding.
- What if our supporting hyperplane  $H$  at  $(0, p^*)$  is **vertical**?
  - The normal to  $H$  is perpendicular to the  $t$  axis
- In this case, no finite  $\lambda$  exists such that  $(\lambda, 1)$  is normal to  $H$ .

# I Lied (A little)

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_1(x) \leq 0 \end{array}$$



- In our proof, we ignored a technicality that can prevent strong duality from holding.
- What if our supporting hyperplane  $H$  at  $(0, p^*)$  is **vertical**?
  - The normal to  $H$  is perpendicular to the  $t$  axis
- In this case, no finite  $\lambda$  exists such that  $(\lambda, 1)$  is normal to  $H$ .
- Somewhat counterintuitively, this can happen even in simple convex optimization problems (though its somewhat rare in practice)

# Violation of Strong Duality

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

- Let domain of constraint be region  $y \geq 1$
- Problem is convex, with feasible region given by  $x = 0$
- Optimal value is 1, at  $x = 0$  and  $y = 1$

# Violation of Strong Duality

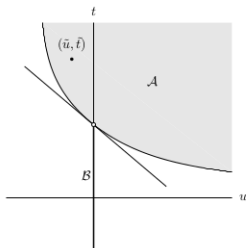
$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & \frac{x^2}{y} \leq 0 \end{array}$$

- Let domain of constraint be region  $y \geq 1$
- Problem is convex, with feasible region given by  $x = 0$
- Optimal value is 1, at  $x = 0$  and  $y = 1$
- $\mathcal{A} = \mathbb{R}_{++}^2 \cup (\{0\} \times [1, \infty])$
- Therefore, any supporting hyperplane to  $\mathcal{A}$  at  $(0, 1)$  must be vertical.
- Optimal dual value is 0; a duality gap of 1.



## Slater's Condition

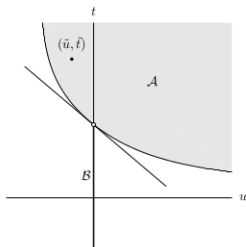
There exists a point  $x \in \mathcal{D}$  where all inequality constraints are strictly satisfied (i.e.  $f_i(x) < 0$ ). I.e. the optimization problem is **strictly feasible**.



- A sufficient condition for strong duality.
- Forces supporting hyperplane to be non-vertical

## Slater's Condition

There exists a point  $x \in \mathcal{D}$  where all inequality constraints are strictly satisfied (i.e.  $f_i(x) < 0$ ). I.e. the optimization problem is **strictly feasible**.



- A sufficient condition for strong duality.
- Forces supporting hyperplane to be non-vertical
- Can be weakened to requiring strict feasibility only of non-affine constraints

# Outline

- 1 The Lagrange Dual Problem
- 2 Duality
- 3 Optimality Conditions**

# Recall: Lagrangian Duality

## Primal Problem

$$\min f_0(x)$$

s.t.

$$f_i(x) \leq 0, \quad \forall i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \forall i = 1, \dots, k.$$

## Dual Problem

$$\max g(\lambda, \nu)$$

s.t.

$$\lambda \succeq 0$$

# Recall: Lagrangian Duality

## Primal Problem

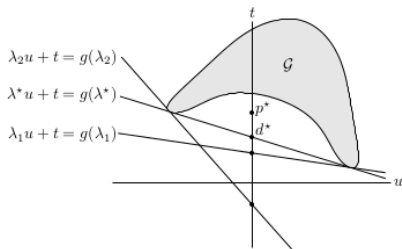
$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

## Weak Duality

$$OPT(dual) \leq OPT(primal).$$



# Recall: Lagrangian Duality

## Primal Problem

$$\min f_0(x)$$

s.t.

$$f_i(x) \leq 0, \quad \forall i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \forall i = 1, \dots, k.$$

## Dual Problem

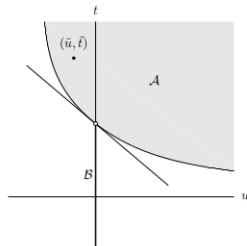
$$\max g(\lambda, \nu)$$

s.t.

$$\lambda \succeq 0$$

## Strong Duality

$$OPT(dual) = OPT(primal).$$



# Dual Solution as a Certificate

## Primal Problem

$$\min f_0(x)$$

s.t.

$$f_i(x) \leq 0, \quad \forall i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \forall i = 1, \dots, k.$$

## Dual Problem

$$\max g(\lambda, \nu)$$

s.t.

$$\lambda \succeq 0$$

- Dual solutions serves as a **certificate** of optimality
- If  $f_0(x) = g(\lambda, \nu)$ , and both are feasible, then both are optimal.

# Dual Solution as a Certificate

## Primal Problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

- Dual solutions serves as a **certificate** of optimality
- If  $f_0(x) = g(\lambda, \nu)$ , and both are feasible, then both are optimal.
- If  $f_0(x) - g(\lambda, \nu) \leq \epsilon$ , then both are within  $\epsilon$  of optimality.
  - $\text{OPT}(\text{primal})$  and  $\text{OPT}(\text{dual})$  lie in the interval  $[g(\lambda, \nu), f_0(x)]$



# Dual Solution as a Certificate

## Primal Problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

- Dual solutions serves as a **certificate** of optimality
- If  $f_0(x) = g(\lambda, \nu)$ , and both are feasible, then both are optimal.
- If  $f_0(x) - g(\lambda, \nu) \leq \epsilon$ , then both are within  $\epsilon$  of optimality.
  - $\text{OPT}(\text{primal})$  and  $\text{OPT}(\text{dual})$  lie in the interval  $[g(\lambda, \nu), f_0(x)]$
- **Primal-dual** algorithms use dual certificates to recognize optimality, or bound sub-optimality.

# Implications of Strong Duality

## Primal Problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

## Facts

If strong duality holds, and  $x^*$  and  $(\lambda^*, \nu^*)$  are feasible & optimal, then

- $x^*$  minimizes  $L(x, \lambda^*, \nu^*)$  over all  $x$ .
- $\lambda_i^* f_i(x^*) = 0$  for all  $i = 1, \dots, m$ . (Complementary Slackness)

# Implications of Strong Duality

## Primal Problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

## Facts

If strong duality holds, and  $x^*$  and  $(\lambda^*, \nu^*)$  are feasible & optimal, then

- $x^*$  minimizes  $L(x, \lambda^*, \nu^*)$  over all  $x$ .
- $\lambda_i^* f_i(x^*) = 0$  for all  $i = 1, \dots, m$ . (Complementary Slackness)

## Proof

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) = \min_x L(x, \lambda^*, \nu^*) \\ &\leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^k \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

# Implications of Strong Duality

## Primal Problem

$$\begin{aligned} \min & f_0(x) \\ \text{s.t.} & \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

## Dual Problem

$$\begin{aligned} \max & g(\lambda, \nu) \\ \text{s.t.} & \\ & \lambda \succeq 0 \end{aligned}$$

## Facts

If strong duality holds, and  $x^*$  and  $(\lambda^*, \nu^*)$  are feasible & optimal, then

- $x^*$  minimizes  $L(x, \lambda^*, \nu^*)$  over all  $x$ .
- $\lambda_i^* f_i(x^*) = 0$  for all  $i = 1, \dots, m$ . (Complementary Slackness)

## Interpretation

- Lagrange multipliers  $(\lambda^*, \nu^*)$  “simulate” the primal feasibility constraints
- Interpreting  $\lambda_i$  as the “value” of the  $i$ 'th constraint, at optimality only the binding constraints are “valuable”
  - Recall economic interpretation of LP

$$\min f_0(x)$$

s.t.

$$f_i(x) \leq 0, \quad \forall i = 1, \dots, m.$$

$$h_i(x) = 0, \quad \forall i = 1, \dots, k.$$

$$\max g(\lambda, \nu)$$

s.t.

$$\lambda \succeq 0$$

## KKT Conditions

Suppose the primal problem is convex and defined on an open domain, and moreover the constraint functions are differentiable everywhere in the domain. If strong duality holds, then  $x^*$  and  $(\lambda^*, \nu^*)$  are optimal iff:

- $x^*$  and  $(\lambda^*, \nu^*)$  are feasible
- $\lambda_i^* f_i(x^*) = 0$  for all  $i$  (Complementary Slackness)
- $\nabla_x L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^k \nu_i^* \nabla h_i(x^*) = 0$

$$\begin{aligned} & \min f_0(x) \\ & \text{s.t.} \\ & f_i(x) \leq 0, \quad \forall i = 1, \dots, m. \\ & h_i(x) = 0, \quad \forall i = 1, \dots, k. \end{aligned}$$

$$\begin{aligned} & \max g(\lambda, \nu) \\ & \text{s.t.} \\ & \lambda \succeq 0 \end{aligned}$$

## KKT Conditions

Suppose the primal problem is convex and defined on an open domain, and moreover the constraint functions are differentiable everywhere in the domain. If strong duality holds, then  $x^*$  and  $(\lambda^*, \nu^*)$  are optimal iff:

- $x^*$  and  $(\lambda^*, \nu^*)$  are feasible
- $\lambda_i^* f_i(x^*) = 0$  for all  $i$  (Complementary Slackness)
- $\nabla_x L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^k \nu_i^* \nabla h_i(x^*) = 0$

## Why are KKT Conditions Useful?

- Derive an analytical solution to some convex optimization problems
- Gain structural insights

# Example: Equality-constrained Quadratic Program

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Px + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

- KKT Conditions:  $Ax^* = b$  and  $Px^* + q + A^T \nu^* = 0$
- Simply a solution of a linear system with variables  $x^*$  and  $\nu^*$ .
  - $m + n$  constraints and  $m + n$  variables

## Example: Market Equilibria (Fisher's Model)

- Buyers  $B$ , and goods  $G$ .
- Buyer  $i$  has utility  $u_{ij}$  for each unit of good  $G$ .
- Buyer  $i$  has budget  $m_i$ , and there's one divisible unit of each good.



## Example: Market Equilibria (Fisher's Model)

- Buyers  $B$ , and goods  $G$ .
- Buyer  $i$  has utility  $u_{ij}$  for each unit of good  $G$ .
- Buyer  $i$  has budget  $m_i$ , and there's one divisible unit of each good.
- Does there exist a **market equilibrium**?
  - Prices  $p_j$  on items, such that each player can buy his favorite bundle that he can afford and the market clears (supply = demand).

# Example: Market Equilibria (Fisher's Model)

- Buyers  $B$ , and goods  $G$ .
- Buyer  $i$  has utility  $u_{ij}$  for each unit of good  $G$ .
- Buyer  $i$  has budget  $m_i$ , and there's one divisible unit of each good.
- Does there exist a **market equilibrium**?
  - Prices  $p_j$  on items, such that each player can buy his favorite bundle that he can afford and the market clears (supply = demand).

## Eisenberg-Gale Convex Program

$$\begin{array}{ll} \text{maximize} & \sum_i m_i \log \sum_j u_{ij} x_{ij} \\ \text{subject to} & \sum_i x_{ij} \leq 1, \quad \text{for } j \in G. \\ & x \succeq 0 \end{array}$$

# Example: Market Equilibria (Fisher's Model)

- Buyers  $B$ , and goods  $G$ .
- Buyer  $i$  has utility  $u_{ij}$  for each unit of good  $G$ .
- Buyer  $i$  has budget  $m_i$ , and there's one divisible unit of each good.
- Does there exist a **market equilibrium**?
  - Prices  $p_j$  on items, such that each player can buy his favorite bundle that he can afford and the market clears (supply = demand).

## Eisenberg-Gale Convex Program

$$\begin{array}{ll} \text{maximize} & \sum_i m_i \log \sum_j u_{ij} x_{ij} \\ \text{subject to} & \sum_i x_{ij} \leq 1, \quad \text{for } j \in G. \\ & x \succeq 0 \end{array}$$

Using KKT conditions, we can prove that the dual variables corresponding to the item supply constraints are market-clearing prices!