# Module 06
# On Lagrangian Duality

**Ahmad F. Taha**

**CE 5999-02 Special Topics — Intro to Optimization**

*Email:* ahmad.taha@vanderbilt.edu

*Webpage:* http://lab.vanderbilt.edu/taha

VANDERBILT UNIVERSITY

March 21, 2023

# In this lecture

- Lagrangian function and the dual function
- Weak duality
- Strong duality
- KKT conditions
- Dual of SDP
- Necessary conditions for local optimality

# Primal optimization problem

Primal optimization problem with variable $x \in \mathbb{R}^n$ (no convexity assumptions)

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \\ & x \in X \end{aligned}$$

- Constraints $x \in X$ are kept implicit (we called $X$ an abstract constraint set)
- We may include the domain of the problem in $X$
- It is our choice which constraints will be kept implicit

# Lagrangian function

Primal optimization problem with variable $x \in \mathbb{R}^n$ (no convexity assumptions)

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \\ & x \in X \end{aligned}$$

Lagrangian function with multiplier vectors $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$

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

Notice we do not include the constraints $x \in X$ in the Lagrangian function. Dual function can be written as

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

# Dual problem

Dual problem with variables $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$

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

The dual function is concave since it is a pointwise infimum (with respect to $x$) of affine functions of $\lambda$ and $\nu$. So the dual problem is *always convex*. This is truly an *awesome* result.

# Dual of standard form LP

Primal: Standard form LP

$$\begin{aligned} \min \quad & c^T x \\ \text{subj. to} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

Dual: Inequality form LP

$$\begin{aligned} \max \quad & -b^T \nu \\ \text{subj. to} \quad & A^T \nu + c \succeq 0 \end{aligned}$$

- Lagrangian function

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

- Dual function

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

- Dual problem

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{subj. to} \quad & \lambda \succeq 0 \end{aligned}$$

# Dual of inequality form LP

Primal: Inequality form LP

$$\begin{aligned} \min \quad & c^T x \\ \text{subj. to} \quad & Ax \preceq b \end{aligned}$$

Dual: Standard form LP

$$\begin{aligned} \max \quad & -b^T \lambda \\ \text{subj. to} \quad & A^T \lambda + c = 0 \\ & \lambda \succeq 0 \end{aligned}$$

- Lagrangian function

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

- Dual function

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

- Dual problem

$$\begin{aligned} \max \quad & g(\lambda) \\ \text{subj. to} \quad & \lambda \succeq 0 \end{aligned}$$

# Weak duality

Recall $f^*$ is primal optimal value, $d^*$ is dual optimal value, no convexity assumptions.

**Result 1:** For any primal feasible $x$ (that is, satisfying $f_i(x) \leq 0$, $i = 1, \dots, m$, and $h_i(x) = 0$, $i = 1, \dots, p$) and any dual feasible $\lambda$ (that is, satisfying $\lambda \succeq 0$), we have that

$$f_0(x) \geq g(\lambda, \nu)$$

which implies the following inequalities (for any primal feasible $x$ and dual feasible $\lambda$)

$$\begin{aligned} f^* &\geq g(\lambda, \nu) \\ f_0(x) &\geq d^* \\ f^* &\geq d^* \quad \text{(weak duality)} \end{aligned}$$

Why? Well, primal and dual feasibility imply

$$\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \leq 0 \Longrightarrow L(x, \lambda, \nu) \leq f_0(x)$$

By definition of the dual function, for any feasible $x$ and $\lambda, \nu$, it holds that

$$g(\lambda, \nu) = \inf_{z \in X} L(z, \lambda, \nu) \leq L(x, \lambda, \nu) \leq f_0(x)$$

# Strong duality

- If $f^* = d^*$, we say strong duality holds
- *Optimality gap* is equal to $f^* - d^*$
- For nonconvex problems, $f^* > d^*$ is typical
- For convex problems, additional requirements called *constraint qualifications* guarantee strong duality:
  Convexity + Constraint qualification $\Longrightarrow$ Strong Duality
- We will give three such constraint qualifications later in this module

# Strong duality: Convex cost-linear constraints

Primal optimization problem with variable $x \in \mathbb{R}^n$

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & g_i^T x - h_i \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x - b_i = 0, \quad i = 1, \ldots, p \\ & x \in X \end{aligned}$$

**Result 2a:** Suppose that $X$ is **polyhedral**, and that $f_0$ is convex with domain $\mathrm{dom}(f) = \mathbb{R}^n$. Then, strong duality holds.

# Strong duality: Convex cost-convex constraints—no equality constraints

Primal optimization problem with variable $x \in \mathbb{R}^n$

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & x \in X \end{aligned}$$

**Result 2b:** Suppose that $f_0, f_1, \ldots, f_m$ are convex. Suppose also that there is a point $\tilde{x} \in X$ such that $f_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$. (The last condition is called **Slater constraint qualification**).Then, strong duality holds.

# Strong duality: General convex problem

Primal optimization problem with variable $x \in \mathbb{R}^n$

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x - b_i = 0, \quad i = 1, \ldots, p \\ & x \in X \end{aligned}$$

**Result 2c:** Suppose that $f_0, f_1, \ldots, f_m$ are convex. Suppose also that there is a point $\tilde{x} \in \operatorname{int}(X)$ such that $f_i(\tilde{x}) < 0$ for $i = 1, \ldots, m$ and $a_i^T \tilde{x} - b_i = 0$ for $i = 1, \ldots, p$. Then, strong duality holds.

# What's next?

Primal optimization problem with variable $x \in \mathbb{R}^n$ (no convexity assumptions)

$$\begin{aligned} f^* = \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \\ & x \in X \end{aligned}$$

We will use duality to obtain conditions for optimality of a point $x$

# Karush-Kuhn-Tucker (KKT) optimality conditions (general form)

1. Primal feasibility: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, \ldots, p$, and $x \in X$
2. Dual feasibility: $\lambda \succeq 0$
3. Lagrangian optimality: $x$ minimizes the Lagrangian function $L(x, \lambda, \nu)$

$$\min_{x \in X} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

4. Complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$

Condition 4 combined with Condition 2 can be equivalently written as

$$\lambda_i > 0 \Longrightarrow f_i(x) = 0$$

Condition 4 combined with Condition 1 can be equivalently written as

$$f_i(x) < 0 \Longrightarrow \lambda_i = 0$$

## Necessary conditions

**Result 3: Suppose strong duality holds**. Then, the KKT conditions are necessary for optimality: If $x^*$ and $(\lambda^*, \nu^*)$ are primal and dual optimal, respectively, then they must satisfy the KKT conditions.

- If $X = \mathbb{R}^n$ and all functions $f_i$ and $h_i$ are continuously differentiable, the Lagrangian optimality (condition 3) can be significantly simplified:

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

- Difficulty in using the result: Check if strong duality holds
- Strong duality holds for convex problems under constraint qualifications
- Result 3 is not very useful for nonconvex problems
  - Strong duality does not generally hold for nonconvex problems
  - Except very special problems such as nonconvex QCQP with a single inequality constraint (p. 229 in textbook).

# KKT conditions under differentiability and $X = \mathbb{R}^n$

Assume the $f_i$ and $h_i$ are continuously differentiable, and $X = \mathbb{R}^n$

1. Primal feasibility: $f_i(x) \leq 0$, $i = 1, \ldots, m$, and $h_i(x) = 0$, $i = 1, \ldots, p$
2. Dual feasibility: $\lambda \succeq 0$
3. Lagrangian optimality: $\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$
4. Complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$

- Summary: for any opt. problem with differentiable objective and constraint functions for which strong duality holds, any pair of primal and dual optimal points must satisfy the KKT conditions
- When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal
- 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

# Example: Convex Quadratic Programs

- Consider the following problem:

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

- This problem, after deriving the KKT conditions

$$Ax^* = b, \qquad Px^* + q + A^T \nu^* = 0$$

can be written as

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

- Solving this set of $m+n$ equations in the $m+n$ variables $x^*$ and $v^*$ gives the optimal primal and dual variables

# Example: Convex Quadratic Programs

- Consider the following problem:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{subject to} & x \geq 0, \ \ \mathbf{1}^T x = 1 \end{array}$$

- This problem, after deriving the KKT conditions

$$x \geq 0, \quad \mathbf{1}^T x^* = 1, \quad \lambda^* \geq 0, \quad \lambda_i^* x_i^* = 0, \quad -\frac{1}{x_i^* + \alpha_i} - \lambda_i^* + \nu = 0$$

- The above system of equations has a non-trivial solution, but here's where you can start from..

# KKT Conditions — Example

Find the minimizer of the following optimization problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) = (x_1-1)^2 + x_2 - 2 \\ \text{subject to} \quad & f(x) = x_1 + x_2 - 2 \leq 0 \\ & h(x) = x_2 - x_1 - 1 = 0 \end{aligned}$$

- First, find the Lagrangian function:

$$\mathcal{L}(x,\lambda,\nu) = (x_1-1)^2 + x_2 - 2 + \lambda(x_2 - x_1 - 1) + \nu(x_1 + x_2 - 2)$$

- Second, find the conditions of optimality (from previous slide):
  1. $\nabla_x \mathcal{L}(x^*,\lambda^*,\nu^*) = \begin{bmatrix} 2x_1^* - 2 - \lambda^* + \nu^* & 1 + \lambda^* + \nu^* \end{bmatrix}^\top = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$
  2. $\lambda^*(x_1^* + x_2^* - 2) = 0$
  3. $\lambda^* \geq 0$
  4. $x_1^* + x_2^* - 2 \leq 0$
  5. $x_2^* - x_1^* - 1 = 0$
  6. $\nabla_x^2 \mathcal{L}(x^*,\lambda^*,\nu^*) = \nabla_x^2 f_0(x^*) + \lambda^* \nabla_x^2 f(x^*) + \nu^* \nabla_x^2 h(x^*) \succeq 0$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} + \lambda^* \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} + \nu^* \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$$

# Example — Cont'd

- To solve the system equations for the optimal $x^*, \lambda^*, \nu^*$, we first try $\lambda^* > 0$.
- Given that, we solve the following set of equations:
  1. $2x_1^* - 2 - \lambda^* + \nu^* = 0$
  2. $1 + \lambda^* + \nu^* = 0$
  3. $x_1^* + x_2^* - 2 = 0$
  4. $x_2^* - x_1^* - 1 = 0$

  $\Rightarrow$ $x_1^* = 0.5, x_2^* = 1.5, \lambda^* = -1, \nu^* = 0$
- But this solution contradicts the assumption that $\lambda^* > 0$
- **Alternative:** assume $\lambda^* = 0 \Rightarrow x_1^* = 0.5, x_2^* = 1.5, \nu^* = -1, \lambda^* = 0$
- This solution satisfies $f(x^*) \leq 0$ constraint, hence it's a candidate for being a minimizer
- We now verify the second order conditions: $L(x^*, \lambda^*, \nu^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$
- Thus, $x^* = \begin{bmatrix} 0.5 & 1.5 \end{bmatrix}^\top$ is a global minimizer (since the problem is convex)

# SDP: Lagrange multipliers for LMIs

Inequality form SDP with $A_i, B \in \mathbb{S}^k$

$$\begin{aligned} f^* = \text{minimize} \quad & c^T x \\ \text{subj. to} \quad & x_1 A_1 + \ldots + x_n A_n \preceq B \end{aligned}$$

- Lagrange multiplier for LMI is a *matrix* $Z \in \mathbb{S}_+^k$
- Inner product ($\sum_i \lambda_i f_i(x)$ previously in the lecture) is the trace

$$\begin{aligned} L(x, Z) &= c^T x + \mathrm{tr}((x_1 A_1 + \ldots + x_n A_n - B) Z) \\ &= \sum_{i=1}^{n} x_i [c_i + \mathrm{tr}(A_i Z)] - \mathrm{tr}(BZ) \end{aligned}$$

Notice that dual feasibility means $Z \succeq 0$.

# Dual of SDP

Dual function

$$g(Z) = \min_{x \in \mathbb{R}^n} L(x, Z) = \begin{cases} -\mathrm{tr}(BZ), & \text{if } c_i + \mathrm{tr}(A_i Z) = 0,\ i = 1, \ldots, n \\ -\infty, & \text{otherwise} \end{cases}$$

Dual problem (with explicit constraints)

$$\begin{aligned} d^* = \text{maximize} \quad & -\mathrm{tr}(BZ) \\ \text{subject to} \quad & c_i + \mathrm{tr}(A_i Z) = 0,\ i = 1, \ldots, n \\ & Z \succeq 0 \end{aligned}$$

- Dual of inequality-form SDP is standard-form SDP
- Dual of inequality-form LP is standard-form LP

# Weak duality

Analogous to weak duality for the general optimization problem in Result 1, it holds for any *primal feasible* $x$ and *dual feasible* $Z$ that

$$\begin{aligned} f^* &\geq g(Z) \\ c^T x &\geq d^* \\ f^* &\geq d^* \quad \text{(weak duality)} \end{aligned}$$

# Strong duality

Again we have that

$$\text{Convexity} + \text{Constraint qualification} \Rightarrow \text{Strong duality } (d^* = f^*)$$

Standard constraint qualification is Slater's condition: There exists $x \in \mathbb{R}^n$ such that

$$x_1 A_1 + \ldots + x_n A_n \prec B.$$

- Notice a small discrepancy with LP
- Result 2a claims that we always have strong duality when the constraints are linear
- But in SDP, we need a little extra as constraint qualification (strict feasibility of the LMI)

# KKT conditions

The KKT conditions are necessary (under strong duality, e.g., by Slater's condition) and sufficient for optimality.
(Define $A(x) = \sum_{i=1}^{n} x_i A_i - B$.)

1. Primal feasibility: $x_1^* A_1 + \ldots + x_n^* A_n \preceq B$
2. Dual feasibility: $Z^* \succeq 0$
3. Lagrangian optimality: $c_i + \mathrm{tr}(A_i Z^*) = 0,\ i = 1, \ldots, n$
4. Complementary slackness: $\mathrm{tr}(A(x^*) Z^*) = 0$

Under primal and dual feasibility, the complementary slackness is equivalent to $A(x^*)Z^* = 0$.
The latter is due to the following result:

$$Z \succeq 0,\ G \preceq 0,\ \mathrm{tr}(GZ) = 0 \Rightarrow GZ = 0.$$

# Necessary optimality conditions: Summary so far

We have seen two flavors of necessary optimality conditions:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{subj. to} \quad & x \in C \end{aligned}$$

$$\begin{aligned} \min \quad & f_0(x) \\ \text{subj. to} \quad & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \\ & x \in X \end{aligned}$$

If $C$ is a **convex set** and $x^*$ is **local minimum**, then $x^*$ satisfies the condition

$$\nabla f_0(x^*)^T(x - x^*) \geq 0\ \forall x \in C.$$

If **strong duality holds** and $x^*$ is **(global) minimum**, then there exist $\lambda^*, \nu^*$ that jointly satisfy the KKT conditions of Result 3.

Next, we give additional necessary optimality conditions for the second problem above.

# Necessary optimality conditions: Outlook

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

- Result 3 requires strong duality
  - This is not very useful when the problem is nonconvex because 1) very few nonconvex problems have strong duality and 2) strong duality is hard to check a priori
- Result 3 is geared towards convex problems. What about nonconvex problems?
- Necessary optimality conditions for nonconvex problems have the following form:

$$\left.\begin{array}{c} f_i, h_i \text{ continuously differentiable} \\ x^* \text{ is a } \textbf{local} \text{ minimum} \\ \text{a constraint qualification holds} \end{array}\right\} \Rightarrow \quad \begin{array}{c} \text{there exist } \lambda^*, \nu^* \text{ such that} \\ x^*, \lambda^*, \nu^* \text{ satisfy KKT cond.} \end{array}$$

# On constraint qualifications (CQs)

- CQ is a condition that the constraint functions $f_i, h_i$ must satisfy
- The most common CQ: *Linear independence constraint qualification (LICQ)*
- Set of inequality constraints that are **active** (that is, satisfied as equality) at $x^*$:

$$\mathcal{A}(x^*) = \{i = 1, \dots, m \,|\, f_i(x^*) = 0\}$$

- **LICQ:** The gradients of the active inequality constraints $\{\nabla f_i(x^*)\}_{i \in \mathcal{A}(x^*)}$ and equality constraints $\{\nabla h_i(x^*)\}_{i=1,\dots,p}$ are *linearly independent*
- Note that to check if LICQ holds, you must have a candidate local minimum $x^*$
- Next, we formally give the necessary optimality conditions and various CQs
- For proofs and additional material, see the book by D. Bertsekas, *Nonlinear Programming*, or any nonlinear programming textbook

# Necessary optimality conditions with LICQ

Suppose $x^*$ is a local minimum for the following problem:

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

Suppose further that $f_i$ $(i = 0, 1, \ldots, m)$ and $h_i$ $(i = 1, \ldots, p)$ are continuously differentiable and that LICQ holds.
Then there exist $\lambda^* \in \mathbb{R}^m, \nu^* \in \mathbb{R}^p$ such that the following hold:

1. Primal feasibility: $f_i(x^*) \leq 0,\ i = 1, \ldots, m;\ h_i(x^*) = 0,\ i = 1, \ldots, p$
2. Dual feasibility: $\lambda^* \succeq 0$
3. Lagrangian optimality: $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$
4. Complementary slackness: $\lambda_i^* f_i(x^*) = 0,\ i = 1, \ldots, m$

# Necessary optimality conditions with Slater CQ

Suppose that the LICQ in the previous result is replaced by the following conditions:

Functions $f_i$ $(i = 1, \ldots, m)$ are convex, and $h_i$ $(i = 1, \ldots, p)$ are linear. In addition, there exists feasible $\tilde{x}$ such that

$$f_i(\tilde{x}) < 0 \quad \forall i \in \mathcal{A}(x^*).$$

Then, the conclusions of the previous result hold.

## Summary and the two cents

- So why should you care?
- Honestly most of this discussion is theoretical
- Point is: if you have a nonconvex problems, deriving KKT conditions and checking CQ is important
- Why? Because solvers for nonconvex problems can take so long to converge
- But for convex problems, probably no one cares...interior point algorithms and common solvers take care of the rest
- Meaning of dual variables in applications
- Why did we derive the duals? Because that's how interior point methods work actually

# Questions And Suggestions?


**Thank You!**

Please visit

https://lab.vanderbilt.edu/taha/

**IFF** you want to know more ☺