# Optimization - Duality Theory

Qamrul Hasan Ansari
*Aligarh Muslim University, India*
*KFUPM, Dhahran, Saudi Arabia*
*E-mail: qhansari@gmail.com*

October 20, 2020

This talk is based on the following books:

- W. Forst and D. Hoffmann, *Optimization – Theory and Practice*, Springer Science & Business Media (2010)
- M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithm*, Third Edition, Wiley-Interscience (2006)

**Helped by:**
**Pradeep, Jauny, Ramsurat, Gourav, Debdas**

# Table of contents

# PART - III
# DUALITY THEORY

# Duality Theory

The concept of duality provides an alternative formulation of the problem that leads to an indirect but sometimes more efficient solution method.

The original problem is referred as primal problem and the transformed problem as the dual.

# Duality Theory

The possible incentives for solving the dual problem in place of primal are the following:

1. The dual is a convex optimization problem. The primal need not be convex.
2. The dual may have smaller dimension and/or simpler constraints than the primal.
3. Under certain circumstance (if there is no duality gap), from the solution of the dual, we can find a solution for primal.
4. Even if there is a duality gap, the dual value is a lower bound of the optimal primal value. This lower bound may be useful in the context of discrete optimization and branch-and-bound techniques.
5. Duality theory gives a potential means of termination criteria for optimization algorithms.

♣ **Perspective 1** (*Min-max perspective*). Consider the following (primal) problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0,\ i = 1, 2, \ldots, m \\ & h_j(x) = 0,\ j = 1, 2, \ldots, \ell, \\ & x \in X. \end{cases} \tag{1}$$

No convexity or continuity assumption on the objective and constraint functions is considered.

Define a new function, called Lagrangian, $L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^\ell \to \mathbb{R}$ by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x). \tag{2}$$

1. weighted sum of objective and constraint functions
2. $\lambda_i$ is Lagrange multiplier associated with $g_i(x) \leq 0$
3. $\mu_j$ is Lagrange multiplier associated with $h_j(x) = 0$

Since

$$\max_{\lambda \geq 0,\, \mu} L(x, \lambda, \mu) = \begin{cases} f(x), & \text{if } x \text{ is feasible} \\ +\infty, & \text{otherwise,} \end{cases}$$

the primal problem can be restated as

$$\min_{x \in X} \max_{\lambda \geq 0,\, \mu} L(x, \lambda, \mu). \tag{3}$$

### Kneser min-max theorem

Let $X \subseteq \mathbb{R}^p$ and $Y \subseteq \mathbb{R}^q$ be a nonempty compact convex sets and $F : X \times Y \to \mathbb{R}$ be a function such that

- $F(x, \cdot) : Y \to \mathbb{R}$ is convex and lower semicontinuous for fixed $x$;
- $F(\cdot, y) : X \to \mathbb{R}$ is concave for fixed $y$.

Then the following *max-min inequality* holds:

$$\min_{x \in X} \ \max_{y \in Y} F(x, y) \geq \max_{y \in Y} \min_{x \in X} F(x, y).$$

Thus, if (3) is difficult to solve, one may attempt to solve the following problem:

$$\max_{\lambda \geq 0, \mu} \min_{x \in X} L(x, \lambda, \mu). \tag{4}$$

We will see below the following facts about the duality:

1. Even if (4) does not always solve (3), a solution to (4) gives a lower bound of the problem (4) due to the max-min inequality.
2. No matter if the objective and constraint functions of the primal problem are discontinuous, (4) is a convex optimization problem and hence it is easier to solve than (3).

♣ **Perspective 2** (*Penalty perspective*). Consider the following primal problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \ldots, m \\ & x \in X. \end{cases} \tag{5}$$

Note that the problem (1) can be written as (5) since any equality constraint can be presented by inequality constraints:

$$h_j(x) = 0 \quad \Leftrightarrow \quad h_j(x) \leq 0 \; \& \; -h_j(x) \leq 0.$$

♣ **Perspective 2** (*Penalty perspective*). Consider the following primal problem:

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \ldots, m \\ & x \in X. \end{cases} \tag{5}$$

Note that the problem (1) can be written as (5) since any equality constraint can be presented by inequality constraints:

$$h_j(x) = 0 \quad \Leftrightarrow \quad h_j(x) \leq 0 \ \& \ -h_j(x) \leq 0.$$

Associated to the constraint function $g_i$, we impose a penalty $\lambda_i \geq 0$ for the violation of $g_i(x) \leq 0$: if $g_i(x_0) > 0$, then we penalize the objective function (that we want to minimize) by an amount $\lambda_i g_i(x_0)$. Accordingly, we attempt to recast the primal problem (5) through

$$\min_{x \in X} \left( f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right). \qquad \text{(Penalty problem)}$$

Question arises what could be the best choices for the values of $\lambda_i$'s? As we wish to solve the primal problem, best choices for $\lambda_i$ are those possible values which give maximum value of the penalty problem because the maximum will refrain us to move out of the feasible region. Thus, to solve the primal problem, we formulate the following problem:

$$\max_{\lambda \geq 0} \min_{x \in X} \left( f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \right).$$

### Remark

Ideally, the equivalent unconstrained formulation is

$$\min_{x \in X} (f(x) + I_X(x)),$$

where $I_X$ is the indicator function of the feasible region $X$.

In the penalty perspective, we are approximating $I_X(x)$ by a linear penalty function.

For further detail on this topic, see Chapter 5 of the following book:

S. Boyd and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press.

♣ **Perspective 3**. (*Two person zero-sum game perspective*). Through game theoretic perspective, one can discover that for every maximization problem, there is a hidden minimization problem.

Consider a two person zero-sum game where the strategy set of the Player 1 and Player 2 are $X$ and $Y$, respectively, and the pay-off function is $F(x, y)$: if Player 1 chooses the strategy $x \in X$ and Player 2 chooses $y \in Y$, then Player 2 has to pay $F(x, y)$ amount to Player 1.

Note that $F(x, y)$ can take both positive and negative values:
a positive value of $F(x, y)$ signifies a gain of Payer 1 and a negative value is loss for Player 1.

We assume that the players are equally intelligent and both players have full knowledge of the other player's action, once it is made.

Given this circumstance, what strategies are best for the players?

Let us see the problem from the view point of Player 1. If he chooses the strategy $x' \in X$, then obviously Player 2 will choose the strategy $y' \in Y$ so that $F(x', y')$ is the minimum most value of $F(x', y)$ in $Y$. Thus, for a chosen $x' \in X$, Player 1 will get the amount $\min_{y \in Y} F(x', y)$. Thus, Player 1 will like to choose that $x \in X$ which gives maximum most value of $\min_{y \in Y} F(x, y)$. Hence, to Player 1, the problem is the following minimization problem:

$$\max_{x \in X} \left( \min_{y \in Y} F(x, y) \right).$$

From the view point of Player 2, we note that if he chooses the strategy $y'' \in Y$, then Player 1 will go for $x'' \in X$ such that $F(x'', y'')$ is the maximum most value of $F(x, y'')$ in $X$. Thus, for $y'' \in Y$, Player 2 will lose an amount $\max_{x \in X} F(x, y'')$. Player 2 obviously will like to choose that $y \in Y$ which gives minimum most loss, i.e., he solves the following maximization problem:

$$\min_{y \in Y} \left( \max_{x \in X} F(x, y) \right).$$

As Player 1 and Player 2 are equally intelligent, you can take the position of any of the players. Depending on whose position you are taking, the problem is a minimization problem or a maximization problem. So, to solve the game problem, you can solve either of maximization or the minimization problem.

For a general primal problem (1), the pay-off function is the Lagrangian function (6). The variable $x$ is your strategy and (dual) variable vector $(\lambda, \mu)$ is opponent's strategy.

There is another perspective: Fenchel conjugate, but that is little advanced.

# Primal problem

As Lagrange duality for nonlinear programming problems gives the general flavour of the results those are obtained in other kinds of duality, we consider to discuss Lagrange duality in this preliminary course on optimization.

Consider the following nonlinear programming problem, which we call the primal problem:

$$
(\mathrm{P}) \qquad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \ldots, m \\ & h_j(x) = 0, j = 1, 2, \ldots, \ell, \\ & x \in X(\neq \emptyset) \subseteq \mathbb{R}^n. \end{cases}
$$

# Lagrange duality

For the problem (P), define the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^\ell \to \mathbb{R}$ by

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x). \tag{6}$$

The Lagrangian dual problem of (P) is stated below:

$$\text{(D)} \qquad \begin{cases} \text{maximize} & \theta(\lambda, \mu) \\ \text{subject to} & \lambda \geq 0, \end{cases}$$

where

$$\theta(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu),$$

$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^m$, $\mu = (\mu_1, \mu_2, \dots, \mu_\ell) \in \mathbb{R}^\ell$, and by $\lambda \geq 0$, we mean $\lambda_i \geq 0$ for all $i = 1, 2, \dots, m$.

- The vector $(\lambda, \mu)$ is called the dual variable or multiplier associated with the problem.
- For each $i = 1, 2, \ldots, m$, we refer to $\lambda_i$ as the dual variable or multiplier associated with the inequality constraint $g_i(x) \leq 0$.
- For each $j = 1, 2, \ldots, \ell$, we refer to $\mu_j$ as the dual variable associated with the equality constraint $h_j(x) = 0$.

The word "primal" is used just to contrast the term "dual".

### Lower bound property

If $p^*$ is the optimal value of the primal problem (P) and $\lambda \geq 0$, then $\theta(\lambda, \mu) \leq p^*$.

### Proof

If $\bar{x}$ is feasible and $\lambda \geq 0$, then

$$f(\bar{x}) \geq L(\bar{x}, \lambda, \mu) \geq \inf_{x \in X} L(x, \lambda, \mu) = \theta(\lambda, \mu)$$

minimizing over all feasible $\bar{x}$ gives $p^* \geq \theta(\lambda, \mu)$.

### Remark

Note that the problem (P) can also be written as

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g_2(x) \geq 0, g_3(x) \geq 0, \ldots, g_m(x) \geq 0, \\ & h_1(x) = 0, h_3(x) = 0, h_4(x) = 0, \ldots, h_l(x) = 0, \\ & x \in X', \end{cases}$$

where $X' = \{x \in X : g_1(x) \geq 0, h_2(x) = 0\}$.

As there are many ways of defining $X$ versus $g$ and $h$, there are many possible Lagrangian duals for the optimization problem (P).

There could be two different Lagrange dual of the same problem.

Find example!

# Geometric interpretation of duality

We now discuss briefly the geometric interpretation of the dual problem. For the sake of simplicity, we consider that the feasible region is given by only one inequality constraint. Then, the primal problem is

$$\begin{cases} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0, \;\; x \in X. \end{cases}$$

In the $(y, z)$-plane, the set

$$G = \{(y, z) : y = g(x), \;\; z = f(x) \text{ for some } x \in X\}$$

is denoted by $G$ in Figure 1. Thus, $G$ is the image of $X$ under the $(g, f)$ map. The primal problem asks us to find a point in $G$ with $y \leq 0$ that has a minimum ordinate. Obviously, the point is $(\bar{y}, \bar{z})$ in Figure 1.

Now suppose that $\lambda = u \geq 0$ is given.

To determine $\theta(u)$, we need to minimize $f(x) + ug(x)$ over all $x \in X$.

Letting $y = g(x)$ and $z = f(x)$ for $x \in X$, minimization of $f(x) + ug(x)$ over $X$ is equivalent to minimization of $z + uy$ over points in $G$.

Note that $z + uy = \alpha$ is an equation of a straight line with slope $-u$ and intercept $\alpha$ on the $z$-axis.

To minimize $z + uy$ over $G$, we need to drag the line $z + uy = \alpha$ parallel to itself as far down as possible until it supports $G$ from below, i.e., the set $G$ lies above the line and touches it.

At the optimum situation of this dragging phenomenon, the intercept on the $z$-axis gives $\theta(u)$, as shown in Figure 1.

The dual problem is therefore equivalent to finding slope $(-\lambda)$ of the supporting hyperplane such that its intercept on the $z$-axis is maximal. In Figure 1, such a hyperplane has slope $-\bar{u}$ and supports the set $G$ at the point $(\bar{y}, \bar{z})$.

Thus, the optimal dual solution is $\bar{u}$, and the optimal dual objective value is $\bar{z}$.

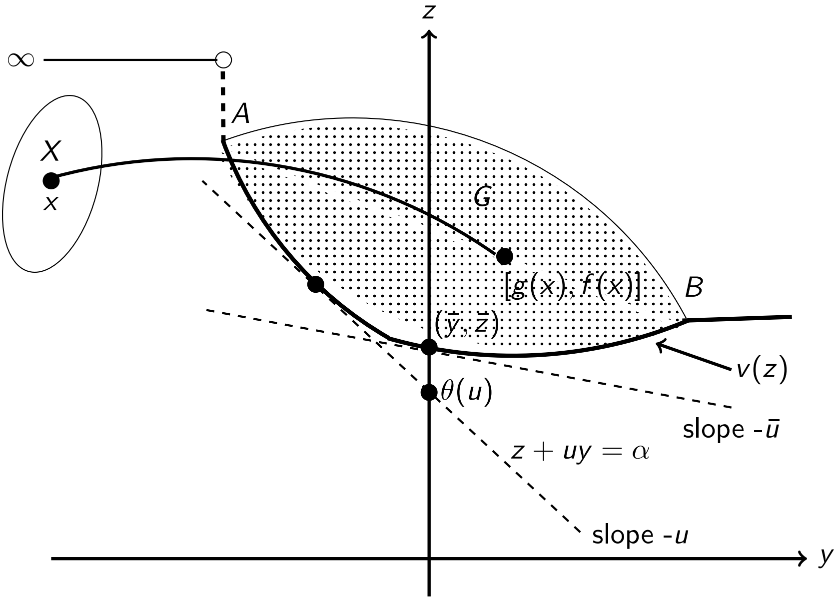



Figure: Geometric interpretation of duality

# Dual function is concave

### Theorem

*Consider the primal problem (P). Let $X$ be a nonempty compact set in $\mathbb{R}^n$, and the functions $f$, $g_i$ and $h_j$, for all $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, \ell$, are continuous on $X$. Then, the objective function*

$$\theta(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu)$$

*of the dual problem (D) is concave over $\mathbb{R}^{m+\ell}$.*

### Remark: The dual problem is a convex optimization problem

From above theorem, we see that the dual objective function $-\theta$ is convex (even if the initial problem is not convex). Hence, the dual problem (D)

$$\max_{\lambda \geq 0, \mu} \theta(\lambda, \mu) \equiv - \min_{\lambda \geq 0, \mu} -\theta(\lambda, \mu)$$

# Dual function is concave

### Theorem

*Consider the primal problem (P). Let $X$ be a nonempty compact set in $\mathbb{R}^n$, and the functions $f$, $g_i$ and $h_j$, for all $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, \ell$, are continuous on $X$. Then, the objective function*

$$\theta(\lambda, \mu) = \inf_{x \in X} L(x, \lambda, \mu)$$

*of the dual problem (D) is concave over $\mathbb{R}^{m+\ell}$.*

### Remark: The dual problem is a convex optimization problem

From above theorem, we see that the dual objective function $-\theta$ is convex (even if the initial problem is not convex). Hence, the dual problem (D)

$$\max_{\lambda \geq 0,\, \mu} \theta(\lambda, \mu) \equiv - \min_{\lambda \geq 0,\, \mu} -\theta(\lambda, \mu)$$

Even though dual problem is convex and it might seem attractive to solve the dual problem, we can have the following issues on solving the dual problem:

1. Explicit expression of the dual objective is difficult to find. This is due to the fact that $\theta$ can be evaluated at a point only after solving a minimization subproblem. Hence, often finding an explicit expression for $\theta$ in terms of $(\lambda, \mu)$ is not available.
2. The dual objective function may not be differentiable even if primal is a smooth problem.
3. The dual problem may not have a solution, even if the primal problem has one; conversely, the primal problem may not have a solution, even if the dual problem has one.

We now provides example for each such cases.

**Example (Dual does not have an explicit expression)**

Let $X = \mathbb{R}^n$, and $A$, $b$, $C$ and $d$ be matrices of order $m \times n$, $m \times 1$, $p \times n$ and $p \times 1$, respectively. Consider the problem

$$\begin{cases} \text{minimize} & \|Ax - b\|_1 \\ \text{subject to} & Cx = d, \ x \in \mathbb{R}^n. \end{cases}$$

The dual function is

$$\theta(\lambda) = \inf_{x \in \mathbb{R}^n} \left( \|Ax - b\|_1 + \langle \lambda, Cx - d \rangle \right), \ \lambda \in \mathbb{R}^p.$$

Although it is possible to evaluate $\theta(\lambda)$ for each given $\lambda$ by solving a linear programming problem, there is no explicit expression for $\theta(\lambda)$.

## Example (Smooth dual problem)

Consider $X = \mathbb{R}^2_+$ and the primal problem

$$\begin{cases} \text{minimize} & f(x) := x_1^2 + x_2^2 \\ \text{subject to} & g_1(x) = 6 - x_1 - x_2 \leq 0 \\ & x \in X. \end{cases}$$

The Lagrangian function for this problem is

$$\begin{aligned} L(x, \lambda) &= x_1^2 + x_2^2 + \lambda(-x_1 - x_2 + 6) \quad (\lambda \geq 0,\, x \in X) \\ &= (x_1 - \lambda/2)^2 + (x_2 - \lambda/2)^2 + 6\lambda - \frac{\lambda^2}{2}. \end{aligned}$$

For a given $\lambda \geq 0$, we get the minimum of $L(x, \lambda)$ in $X$ for $x_1 = x_2 = \frac{\lambda}{2}$ with value $6\lambda - \frac{\lambda^2}{2}$. Hence, the dual objective is $\theta(\lambda) = 6\lambda - \frac{\lambda^2}{2}$, which is differentiable.

## Example (Nonsmooth dual problem)

Consider $X = \mathbb{R}^2$ and the problem

$$\begin{cases} \text{minimize} & -x_1 - x_2 \\ \text{subject to} & 2x_1 + 4x_2 - 3 \leq 0, \\ & 0 \leq x_1 \leq 1, \\ & 0 \leq x_2 \leq 2. \end{cases}$$

The Lagrangian function is

$$L(x_1, x_2, \lambda) = -x_1 - x_2 + \lambda(2x_1 + 4x_2 - 3)\ (\lambda \geq 0, x \in X).$$

The dual objective function is

$$\begin{aligned} \theta(\lambda) &= -3\lambda - \inf_{0 \leq x_1 \leq 2} (-1 + 2\lambda)x_1 - \inf_{0 \leq x_1 \leq 2} (-1 + 4\lambda)x_2 \\ &= \begin{cases} -3 + 5\lambda & \text{if } 0 \leq \lambda \leq \frac{1}{4} \\ -2 + \lambda & \text{if } \frac{1}{4} \leq \lambda \leq \frac{1}{2} \\ -3\lambda & \text{if } \frac{1}{2} \leq \lambda, \end{cases} \end{aligned}$$

Example (Dual does not have optimum solution but primal has one)

Consider $X = \mathbb{R}$ and the primal as

$$\begin{cases} \text{minimize} & f(x) := x + 2010 \\ \text{subject to} & g_1(x) := \frac{1}{2}x^2 \leq 0, \\ & x \in X. \end{cases}$$

The Lagrangian is

$$L(x, \lambda) := f(x) + \lambda g(x) = x + 2010 + \frac{\lambda}{2}x^2 \ (\lambda \geq 0,\ x \in \mathbb{R}).$$

Therefore, the dual problem is

$$\max_{\lambda \geq 0} \ \theta(\lambda) := 2010 - \frac{1}{2\lambda},$$

which does not have a solution. However, note that primal has an optimum solution, namely $x^* = 0$.

Example (Primal does not have a solution but dual has one)

Consider $X = \mathbb{R}$, and the primal problem

$$\begin{cases} \text{minimize} & f(x) := \exp(-x) \\ \text{subject to} & g_1(x) := -x \leq 0, \\ & x \in X. \end{cases}$$

We have $\inf\{\exp(-x) \ : \ x \geq 0\} = 0$, but there exists no $x \geq 0$ with $f(x) = 0$. Hence, primal does not have a solution.
The Lagrangian is

$$L(x, \lambda) := f(x) + \lambda g(x) = \exp(-x) - \lambda x \ (\lambda \geq 0, x \in \mathbb{R}).$$

The only value of $\lambda$ in $\mathbb{R}_+$ for which $\inf_{x \in X} L(x, \lambda) > +\infty$ is 0. So we have $\max\{\theta(\lambda) : \lambda \geq 0\} = 0 = \theta(0)$. Hence, dual problem has a solution.

Next, we explore the relationship between the primal and dual optimal if the exist.

### Theorem (Weak Duality Theorem)

*Let $x_0$ is feasible to the primal problem (P), i.e., $x_0 \in X$, $g(x_0) \leq 0$, and $h(x_0) = 0$. Also, let $(\lambda', \mu')$ be feasible to the dual problem (D), i.e., $\lambda' \geq 0$. Then, we have*

$$f(x_0) \geq \theta(\lambda', \mu').$$

### Proof

By the definition of $\theta$, and since $x \in X$, we have

$$\begin{aligned}
\theta(\lambda', \mu') &= \inf_{x \in X} \{f(x) + \langle \lambda, g(x) \rangle + \langle \mu, h(x) \rangle : x \in X\} \\
&\leq f(x_0) + \langle \underbrace{\lambda'}_{\geq 0}, \underbrace{g(x)}_{\leq 0} \rangle + \langle \mu, \underbrace{h(x_0)}_{=0} \rangle \\
&\leq f(x_0).
\end{aligned}$$

### Remark

Theorem 7 immediately implies that if both of primal and dual has optimum solutions, and
$d^* = \max_{\lambda \geq 0,\ \mu} \theta(\lambda, \mu)$ and
$p^* = \inf\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$, then

$$d^* \leq p^*.$$

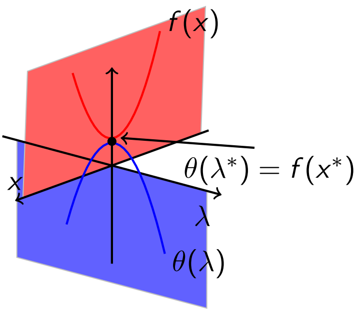
### Duality gap

The difference $p^* - d^*$ is called the *duality gap*. If this duality gap is zero, that is, $p^* = d^*$, then we say that *strong duality* holds.

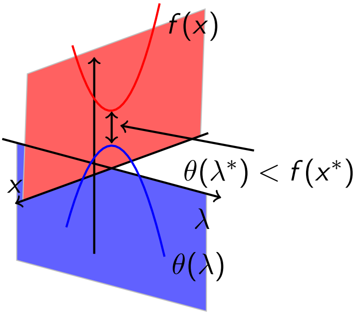
We will see later that if the function $f$ and $g$ are convex (on the convex set $X$) and a certain constraint qualification holds, then one has strong duality.

In nonconvex cases, however, a duality gap $p^* - d^* > 0$ has to be expected.

# Duality gap




*Strong Duality*




*Weak Duality*

# Example with duality gap

Consider $X = [0,1]$ and the primal problem

$$\begin{cases} \text{minimize} & f(x) := -x^2 \\ \text{subject to} & g(x) := 2x - 1 \leq 0, \\ & x \in X. \end{cases}$$

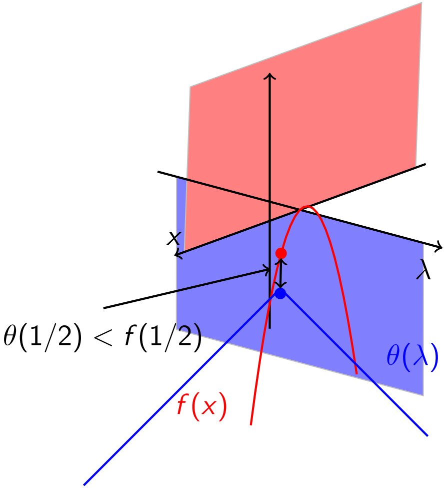
In this problem, primal optimum is $p^* = f(\frac{1}{2}) = -\frac{1}{4}$.

The Lagrangian is $L(x,\lambda) = -x^2 + \lambda(2x-1)$ $(x \in [0,1], \lambda \geq 0)$.
Therefore,

$$\theta(\lambda) = \begin{cases} -\lambda, & \text{if } \lambda \geq 1/2 \\ \lambda - 1, & \text{if } \lambda < 1/2, \end{cases}$$

and hence, $d^* = \max(D) = \varphi(1/2) = -1/2$.

# Example with duality gap




*Weak Duality*

# No duality gap

### Example

Consider $X = \mathbb{R}^2_+$ and the primal problem

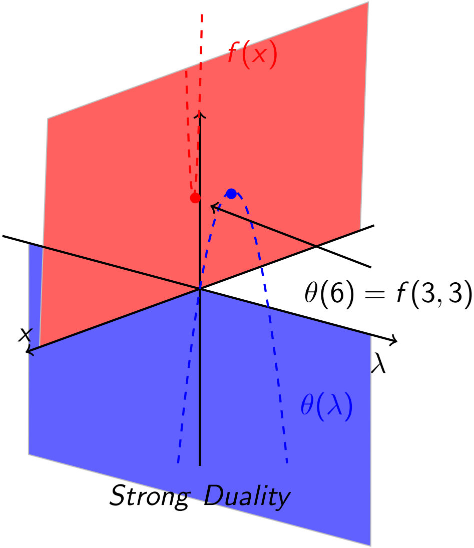
$$\begin{cases} \min & f(x) := x_1^2 + x_2^2 \\ \text{subject to} & g_1(x) = 6 - x_1 - x_2 \leq 0 \\ & x \in X. \end{cases}$$

Note that $g_1(x) \leq 0$ implies $6 \leq x_1 + x_2$. The equality $6 = x_1 + x_2$ gives $f(x) = x_1^2 + (6 - x_1^2) = 2((x_1 - 3)^2 + 9)$.

The minimum is attained at $x^* = (3, 3)$ with $f(x^*) = 18$

Thus, $\theta(\lambda) = 6\lambda - \lambda^2/2$, which describes a parabola whose maximum arises at $\lambda = 6$ with value $\varphi(\lambda) = 18$.

# No duality gap



# Equal primal and dual objective values implies optimal

## Corollary

*If $f(x^*) = \theta(\lambda^*, \mu^*)$ for some primal and dual feasible points $x^*$ and $(\lambda^*, \mu^*)$, respectively, then $x^*$ is a minimizer to the primal problem (P) and $\lambda^*$ is a maximizer to the dual problem (D).*

## Proof

From Remark 34, we note that

$$\max_{\lambda \geq 0, \mu} \theta(\lambda, \mu) \leq \min\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}.$$

As $x^*$ and $(\lambda^*, \mu^*)$ are primal and dual feasible points, respectively, we have

$$\theta(\lambda^*, \mu^*) \leq \max_{\lambda \geq 0, \mu} \theta(\lambda, \mu) \leq \min\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} \leq f(x^*).$$

Since $f(x^*) = \theta(\lambda^*, \mu^*)$, we obtain
$f(x^*) = \min\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$ and $\theta(\lambda^*, \mu^*) = \max_{\lambda \geq 0, \mu} \theta(\lambda, \mu)$.

# Remark

Suppose that we need to solve $\min_{x \in C} f(x)$.

How do we measure the goodness of an approximate solution $\hat{x}$?

Ideally, we compute $f(\hat{x}) - f^*$ to see the goodness of $\hat{x}$.

However, we do not know $f^*$!

Surprisingly, duality can give us a way to check the closeness of $\hat{x}$ to an optima through the following way.

Suppose that we have a primal feasible point $\hat{x}$ and a dual-feasible vector $(\hat{\lambda}, \hat{\nu})$ such that $f(\hat{x}) - \theta(\hat{\lambda}, \hat{\nu}) \leq \epsilon$. From weak duality, we have

$$\theta(\hat{\lambda}, \hat{\nu}) \leq \theta^* \leq f^* \leq f(\hat{x}),$$

which implies that $\hat{x}$ and $(\hat{\lambda}, \hat{\nu})$ must be $\epsilon$-optimal (i.e., their objective functions differ by no more than $\epsilon$ from the objective functions of the true optima $x^*$ and $(\lambda^*, \nu^*)$, respectively).

### Theorem (Saddle points and strong duality)

*Let $x^*$ be a point in $C$ and $\lambda^* \in \mathbb{R}^m_+$. Then the following statements are equivalent:*

1. *$(x^*, \lambda^*)$ is a saddle point of the Lagrange function $L$.*
2. *$x^*$ is a minimizer to problem $(P)$ and $\lambda^*$ is a maximizer to problem $(D)$ with*
$$f(x^*) = L(x^*, \lambda^*) = \theta(\lambda^*).$$

*In other words: A saddle point of the Lagrangian $L$ exists if and only if the problem $(P)$ and $(D)$ have the same value and admit optimizers.*

If the SLATER constraint qualification holds and the original problem is convex, then we have strong duality, that is, $p^* = d^*$.

**Theorem (Convex optimization problem and SQ implies strong duality)**

*Suppose that the Slater's constraint qualification:*

$$\exists \hat{x} \in \mathscr{F} \text{ such that } g_i(\hat{x}) < 0 \text{ for all } i \in \mathcal{I}_1$$

*holds for the convex problem* $(P)$*. Then, we have strong duality, and the value of the dual problem is attained if* $p^* > -\infty$.

# Other kinds of duality

There are several other kinds of duality in the literature. We list four most commonly used duality.

- *Dorn duality:* It is exclusively for quadratic programming problem.
- *Wolfe duality:* Extension of duality theory for linear programming to convex programming problems.

# Other kinds of duality

There are several other kinds of duality in the literature. We list four most commonly used duality.

- *Dorn duality:* It is exclusively for quadratic programming problem.
- *Wolfe duality:* Extension of duality theory for linear programming to convex programming problems.
- *Mond-Weir duality:* Convexity is relaxed by pseudo-convexity.

# Other kinds of duality

There are several other kinds of duality in the literature. We list four most commonly used duality.

- *Dorn duality:* It is exclusively for quadratic programming problem.
- *Wolfe duality:* Extension of duality theory for linear programming to convex programming problems.
- *Mond-Weir duality:* Convexity is relaxed by pseudo-convexity.
- *Fenchel Duality:* Needs good acknowledge of Convex Analysis

# Other kinds of duality

There are several other kinds of duality in the literature. We list four most commonly used duality.

- *Dorn duality:* It is exclusively for quadratic programming problem.
- *Wolfe duality:* Extension of duality theory for linear programming to convex programming problems.
- *Mond-Weir duality:* Convexity is relaxed by pseudo-convexity.
- *Fenchel Duality:* Needs good acknowledge of Convex Analysis

**Remark**

The Lagrange duality approach is more general than the Wolfe Duality and the Mond-Weir duality because in order to hold weak and strong duality, Lagrange duality does not assume apriori the existence of a minimizer of the primal.

Possibly for this reason Lagrange duality received more attention than any other kinds of duality. However, main drawback of the Lagrange duality is that an explicit expression of the dual objective function is often very difficult to identify.

# Dorn duality

The first extension of duality for linear programming to quadratic programming was done by Dorn [1] in 1960. He considered the following (primal) problem:

$$\begin{cases} \min & \frac{1}{2}\langle x, Qx\rangle + \langle p, x\rangle \\ \text{subject to} & Ax \geq b \\ & x \geq 0, x \in \mathbb{R}^n, \end{cases}$$

where $Q$ is an $n \times n$ symmetric positive semidefinite matrix, $p$ is an $n \times 1$ vector, $A$ is an $m \times n$ matrix, and $b$ is an $m \times 1$ vector.

W. S. Dorn (1960). Duality in quadratic programming. *Quarterly of Applied Mathematics*, 18(2), 155–162.

# Dorn duality

Dorn duality for this problem is the following maximization problem:

$$\begin{cases} \max & -\frac{1}{2}\langle x, Qx\rangle - \langle b, \lambda\rangle \\ \text{subject to} & Hx + \langle A, \lambda\rangle + p = 0 \\ & \lambda \geq 0, \lambda \in \mathbb{R}^m. \end{cases}$$

The prime result of Dorn duality is that if either of primal or dual is feasible, then strong duality holds without a constraint qualification. In addition, an optimal solution to a self-dual convex quadratic programming problem always lies at an extreme point of the constraint set, which is a very important computational property similar to the linear programming problems.

However, note that the dual problem contains the primal variable $x$ which makes the number of variables in the dual more than the primal problem.

# Wolfe duality

Extension of duality theory for linear programming to smooth convex programming problems was done by Wolfe [1].

Let $f$ and $g_i$, $i = 1, 2, \ldots, m$ be convex and continuously differentiable functions in $\mathbb{R}^n$. Consider the following (primal) problem

$$\begin{cases} \min & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \ldots, m \\ & x \in \mathbb{R}^n. \end{cases}$$

P. Wolfe (1961). A duality theorem for nonlinear programming, *Quarterly of Applied Mathematics*, 19, 239–244.

# Wolfe duality

The Wolfe dual of the above problem is

$$
\text{(WD)} \qquad \begin{cases} \max & f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \\ \text{subject to} & \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ & \lambda_i \geq 0, i = 1, 2, \ldots, m. \end{cases}
$$

As the equality constraint of (WD) is nonlinear, in general, the Wolfe dual problem may be a nonconvex optimization problem.

However, weak duality holds by (WD) and under Kuhn-Tucker constraint qualification (closure of the attainable directions is equal to the linearizing cone) (WD) enjoys strong duality.

# Wolfe duality

The Wolfe dual of the above problem is

$$
\text{(WD)} \qquad \begin{cases} \max & f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) \\ \text{subject to} & \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ & \lambda_i \geq 0, i = 1, 2, \ldots, m. \end{cases}
$$

As the equality constraint of (WD) is nonlinear, in general, the Wolfe dual problem may be a nonconvex optimization problem.

However, weak duality holds by (WD) and under Kuhn-Tucker constraint qualification (closure of the attainable directions is equal to the linearizing cone) (WD) enjoys strong duality.

# Remark on Wolfe duality

As Wolfe himself noted, (WD) may be difficult to deal with computationally since its constraint set involves nonlinear equality constraints and need not even be convex.

Another difficulty is that (WD) is more prone than Lagrange duality of a duality gap since it has an extra constraint than just $\lambda \geq 0$.

Nonetheless, Wolfe's dual is preferred than Lagrange duality for smooth convex programming since (WD) often easily finds a suitable (closed) form of the dual problem.

# Mond-Weir duality

In order to relax the convexity requirements in Wolfe duality, Mond and Weir [1] proposed a duality theory for pseudo-convex functions.

Consider the following (primal) problem:

$$\begin{cases} \min & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, 2, \ldots, m \\ & x \in \mathbb{R}^n. \end{cases}$$

B. Mond (2009). Mond–Weir duality. Chapter 8, In: Optimization, C. Pearce and E. Hunt (Editors), pp. 157–165. Springer, New York, NY, 2009.

# Mond-Weir duality

The Mond-Weir dual of the above problem is

$$\text{(MWD)} \qquad \begin{cases} \max & f(x) \\ \text{subject to} & \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ & \lambda_i g_i(x) \geq 0, i = 1, 2, \ldots, m \\ & \lambda_i \geq 0, i = 1, 2, \ldots, m. \end{cases}$$

If $f$ is pseudo-convex and $\lambda^\top g$ is quasi-convex, then weak duality of (MWD) holds. Under Slater's CQ, strong duality holds for (MWD) provided primal problem has an optimum solution. The main advantage of Mond-Weir type dual is that the objective function is same as in the primal problem. Thus, one does not need any effort to find explicit expression of the dual objective function. However, as discussed in Wolfe duality, (MWD) is in general nonconvex even for a convex primal.

# Mond-Weir duality

The Mond-Weir dual of the above problem is

$$
\text{(MWD)} \qquad \begin{cases} \max & f(x) \\ \text{subject to} & \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ & \lambda_i g_i(x) \geq 0, i = 1, 2, \ldots, m \\ & \lambda_i \geq 0, i = 1, 2, \ldots, m. \end{cases}
$$

If $f$ is pseudo-convex and $\lambda^\top g$ is quasi-convex, then weak duality of (MWD) holds. Under Slater's CQ, strong duality holds for (MWD) provided primal problem has an optimum solution. The main advantage of Mond-Weir type dual is that the objective function is same as in the primal problem. Thus, one does not need any effort to find explicit expression of the dual objective function. However, as discussed in Wolfe duality, (MWD) is in general nonconvex even for a convex primal.

# Questions?

&

# Questions?

&

# Thanks!

# Questions?

&

# Thanks!