| | |
|---|---|
| **EE 381V: Large Scale Optimization** | **Fall 2012** |

# Lecture 12 — October 9

*Lecturer: Caramanis & Sanghavi* *Scribe: Xinyang Yi & Yicong Wang*

## 12.1 Last time

In the previous lecture, in the first place, we talked about duality for general non-convex optimization. And we know that dual functions are always convex. The dual of the dual problem is also convex. In applications, the dual of the dual can be used as the convex relaxation of the primal (and we will see this explicitly in the next lecture). Then, we covered the concepts of weak duality and strong duality. We proved that weak duality always holds and strong duality always holds for linear programming. This time we will continue the topics about duality, strong duality and related applications.

## 12.2 Quadratically constrained quadratic program

The optimization problem are called quadratically constrained quadratic program(QCQP) if objective funciton are both quadratic. In particular, we have:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2}x^T P_0 x + p_0^T x + r_0 \\ \text{subject to} \quad & \frac{1}{2}x^T P_i x + p_i^T x + r_i \leq 0,\ i = 1, \ldots, m. \end{aligned} \tag{12.1}$$

Note that $P_0$, $P_i$ may not be positive semidefinite, which means the optimizaton problem doesn't have to be convex. In this case, we can use the dual of the dual to get the convex relaxation of the primal. For now, we make the assumption that $P_0, P_i \succeq 0$, the QCQP problem is convex. The we derive the dual problem. Recall what we did in last time, we get the dual function($\lambda$):

$$\begin{aligned} q(\lambda) &= \inf_x \left( \frac{1}{2}x^T P_0 x + p_0^T x + r_0 + \sum_i \lambda_i \left( \frac{1}{2}x^T P_i x + p_i^T x + r_i \right) \right) \\ &= \inf_x \left( \frac{1}{2}x^T \left[ P_0 + \sum_i \lambda_i P_i \right] x + \left[ p_0 + \sum_i \lambda_i p_i \right]^T x + \left[ r_0 + \sum_i \lambda_i r_i \right] \right). \end{aligned}$$

To simplify expression, we let $P(\lambda) := P_0 + \sum_i \lambda_i P_i$, $p(\lambda) := p_0 + \sum_i \lambda_i p_i$,$r(\lambda) := r_0 + \sum_i \lambda_i r_i$, then we can get:

$$q(\lambda) = \inf_x \left( \frac{1}{2}x^T P(\lambda) x + p(\lambda)^T x + r(\lambda) \right). \tag{12.2}$$

Since $\lambda_i$ is non-negative, $P(\lambda) \succ 0$. So $x^* = -P(\lambda)^{-1}p(\lambda)$, and

$$q(\lambda) = \frac{1}{2}p(\lambda)^T P(\lambda)^{-1}p(\lambda) + r(\lambda). \tag{12.3}$$

The dual of the primal problem (12.1) is:

$$\begin{aligned} \max_{\lambda} \quad & q(\lambda) \\ \text{s.t.} \quad & \lambda \geq 0. \end{aligned} \tag{12.4}$$

**Excercise 1:** Check that $q(\lambda)$ is concave, which means that the dual problem is convex.

**Proof:** Actually this is true for a general optimization problem, convex or not. Let $L(x, \lambda)$denote the lagrangian dual function. Suppose $q(\lambda_1) = \inf_x L(x, \lambda_1)$, $q(\lambda_2) = \inf_x L(x, \lambda_2)$. For $0 \leq \theta \leq 1$, we have:

$$\begin{aligned} L(x, \theta\lambda_1 + (1-\theta)\lambda_2) &= \theta L(x, \lambda_1) + (1-\theta)L(x, \lambda_2) \\ &\geq \theta q(\lambda_1) + (1-\theta)q(\lambda_2) \end{aligned}$$

By taking the infimum of the left side, we get $q(\theta\lambda_1 + (1-\theta)\lambda_2) \geq \theta q(\lambda_1) + (1-\theta)q(\lambda_2)$. Thus $q(\lambda)$ is concave. Another simple way to see this is that $q(\lambda)$ is the infimum of linear functions, and hence concave. □

## 12.3 Strong duality—Slater's condition

We already know that weak duality $p^* \geq d^*$ always holds. The natural question is that when strong duality $p^* = q^*$holds. However, it's difficult to give the necessary condition of strong duality. On the other hand, there are sets of sufficient conditions where strong duality holds. These sets of condition are generally called *constraint qualification*. Among them, Slater's condition is one of the most important and commonly used since it is able to cover a large set of optimization problems, and in many cases, it can be easy to check that it holds.

**Theorem 12.1.** *Consider the following optimzition problem:*

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & g_i(x) \leq 0, \ i = 1, \ldots, m. \\ & A_1 x \leq b_1 \\ & A_2 x = b_2. \end{aligned} \tag{12.5}$$

*Suppose that $f$ and $g_i$ are convex (nonlinear) functions. If there exists $\bar{x} \in$ **relint** $dom(f)$ such that $g_i(\bar{x}) < 0$, then strong duality holds.*

**Proof:** For the detailed proof of Slater's condition, please refer to *Convex analysis and optimization* by D. Bertsekas. The basic idea is that for a convex set, if there exists a point which lies outside the set, we could find a hyperplane to seperate support the set. Actually, when $f, g$ are convex, $C = \textbf{epigraph}(f(x) - f^*, g(x))$ is also convex. Note that $(0,0) \notin \textbf{relint}\, C$, so when slater's condition are satisfied, we could find hyperplane $(1, \lambda), \lambda \geq 0$ such that $f(x) - f^* + \lambda g(x) \geq 0$. □

The insight that we gained from the proof of Slater's condition is that strong duality typically doesn't hold for nonconvex problems. It's easy to give such examples.
**Excercise 2:** Construct a non-convex optimization problem which shows strong duality doesn't hold (Hint: LP + integral constraints).
Next, we give two examples where Slater's condition is not satisfied and strong duality doesn't hold in the same time.
**Example 1:** Considering the following optimization problem:

$$\begin{aligned} \min_{x_1,x_2} \quad & x_1 + x_2 \\ \text{s.t.} \quad & (x_1+1)^2 + x_2^2 - 1 = 0, \\ & (x_1-2)^2 + x_2^2 - 4 = 0 \end{aligned} \tag{12.6}$$

It's obvious that the feasible set is $\mathfrak{X} = (0,0)$. Figure 12.1 shows that the constraints are two circles with only one intersect point. Thus, the optimal solution $x^* = (0,0)$. We know that for convex optimization, the optimal solution should satisfy $0 \in \nabla f(x^*) + N_{\mathfrak{X}}(x^*)$, where $N_{\mathfrak{X}}(x^*)$ is the norm cone at $x^*$. Remember that this is always true. However, let us consider another expression $0 \in \nabla f(x^*) + \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*)$. In this case,$\nabla f(x^*) = (1,1)^T$, $\nabla g_1(x^*) = (2,0)^T$, $\nabla g_2(x^*) = (-4,0)^T$. We can't find such $\lambda_1, \lambda_2$ that:

$$0 \in \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -4 \\ 0 \end{pmatrix}.$$

This tells us the latter expression doesn't always holds which means the norm cone cannot always be written as the linear combination of constraints' gradient. For this example, it's easy to check that Slater's condition doesn't hold and there is no guarantee of strong duality.
**Example 2:** Let's consider this problem:

$$\begin{aligned} \min_{x} \quad & e^{x_2} \\ \text{s.t.} \quad & \|x\|_2 - x_1 \leq 0 \end{aligned} \tag{12.7}$$

From the constraint we know the feasible region $X = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\}$. The constraint is always active so Slater's condition is not satisfied. Next we show that strong duality also doesn't hold. The dual:

$$q(\lambda) = \inf_x \left( e^{x_2} + \lambda(\|x\|_2 - x_1) \right). \tag{12.8}$$

**Excercise 3:** Check that for any value of $\lambda \geq 0$, $q(\lambda) = 0$, which implies that $d^* = 0 < p^* = 1$.

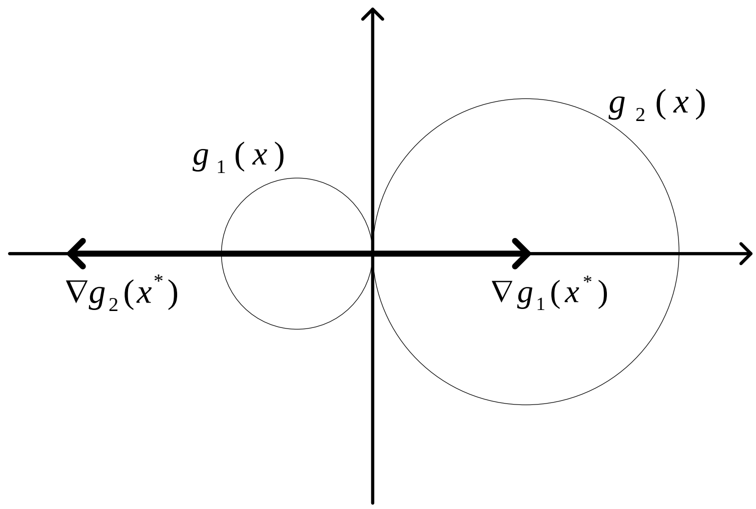


**Figure 12.1.** Illustration of Example 1

**Proof:** It's obvious that $e^{x_2} + \lambda(\|x\|_2 - x_1) \geq 0$. Let $x_1 = x_2^4$. We have:

$$e^{x_2} + \lambda(\|x\|_2 - x_1) = e^{x_2} + \lambda x_2^4(\sqrt{1 + \frac{1}{x_2^6}} - 1).$$

When $x_2 \to -\infty$, $e^{x_2} \to 0$, $x_2^4\sqrt{1 + \frac{1}{x_2^6}} - 1 \sim \frac{1}{2x_2^2} \to 0$. Thus $q(\lambda) = 0$ for any $\lambda \geq 0$. □

The previous two examples show that strong duality doesn't hold when Slater's condition is not satisfied. But it's worth to note that Slater's condition is just sufficient, not neccesary. It's possible that strong duality holds when Slater's condition is not satisfied.

## 12.4 Complementary Slackness

Let us consider the optimization problem:

$$\begin{aligned} \max_x \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0 \end{aligned} \tag{12.9}$$

$F$ and $G$ are convex functions and the same assumption holds in the following unless we state that they are non-convex. Take the dual function:

$$L(\lambda, x) = f(x) + \sum_i \lambda_i g_i(x). \tag{12.10}$$

We could think of $\lambda_i$ as penalties, that is, we will get punished if $g_i(x) > 0$. We need to guarantee $g_i(x) \leq 0$ if we want to minimize the price we pay. Suppose *Strong Duality* holds for the above optimization problem, e.g. Slater's Condition is satisfied, and $x^*$ is a solution to the problem with finite optimal value, then we have:

$$\begin{aligned} f(x^*) = q(x^*) &\triangleq \inf_x (f(x) + \sum_i \lambda_i g_i(x)) \\ &\leq f(x^*) + \sum_i {\lambda_i}^* g_i(x^*) \\ &\leq f(x^*). \end{aligned}$$

The first inequality holds since ${\lambda_i}^* \geq 0$, $g_i(x^*) \geq 0$ . Since $f(x^*) = g(x^*)$ and $g(x^*) \leq f(x^*)$, we can see that

$$\sum_i {\lambda_i}^* g_i(x^*) = 0,$$

$$\lambda_i^* g_i(x^*) = 0 \text{ for } \forall i. \tag{12.11}$$

At least one of the two terms is zero for every constraint and this is *Complementary Slackness*. Let's see two application examples of complementary slackness.

### 12.4.1 Power Allocation Problem: Waterfilling

Consider the problem of allocating a certain amount of power to different frequencies with the object to maximize the total data rate. We have a Power Budget, $\bar{P}$, and we allocate $P_i$ to frequency band $i$, while the noise at frequency band $i$ is $n_i$ . According to Shannon's formula, the data transmission rate at band $i$ with transmission power $P_i$ and noise $n_i$ is:

$$\begin{aligned} Rate(P_i) &= \log(1 + SNR) \\ &= \log(1 + \tfrac{P_i}{n_i}) \\ &= \log(n_i + P_i) - \log(n_i) \end{aligned} \tag{12.12}$$

We can write formulate the Optimization problem as:

$$\begin{aligned} \max \quad & \sum_i Rate(P_i) \\ s.t. \quad & \sum_i P_i \leq \bar{P} \\ & P_i \geq 0 \end{aligned} \tag{12.13}$$

Figure 12.2 shows how data rate changes with $P_i$. Clearly, when $n_i$ are all equal, then we equally allocate power to each channel in order to maximize the total rate. Now let's consider the general case in which $n_i$ may be different. Rewrite the optimization problem:

$$\begin{aligned} \min \quad & -\sum_i \log(n_i + P_i) \\ s.t. \quad & \sum_i P_i \leq \bar{P} \\ & P_i \geq 0 \end{aligned} \tag{12.14}$$

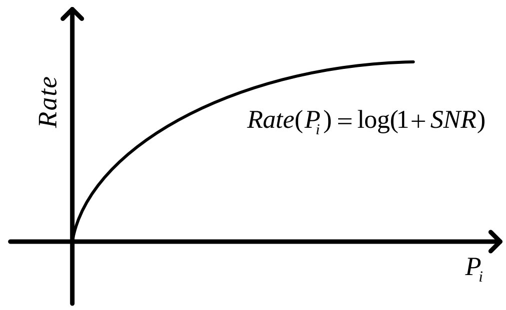


**Figure 12.2.** Relation bewteen channel rate and power

We can prove that the above optimization problem satisfies Slater's Condition, thus *Strong Duality* holds. The dual function of the problem is

$$q(\lambda, \mu_i) = \inf_{P_i}(-\sum_i \log(n_i + P_i) + \lambda(\bar{P} - \sum_i P_i) + \sum_i \mu_i P_i),$$

with $\lambda \leq 0, \mu_i \leq 0$. Function $q$ here is convex. Take gradient of $q$ with respect to $P_i$, we have

$$-\frac{1}{n_i P_i} - \lambda + \mu_i = 0$$

$$\mu_i \leq 0 \Leftrightarrow -\frac{1}{n_i + P_i} - \lambda = -\mu i \geq 0$$

According to *Complementary Slackness*, $\mu_i \cdot P_i = 0$, then $P_i(\lambda + \frac{1}{n_i+P_i}) = 0$. Now we consider two cases, $0 \leq -\lambda < \frac{1}{n_i}$, and $-\lambda > \frac{1}{n_i}$. If $0 \leq -\lambda < \frac{1}{n_i}$, we have

$$\begin{aligned} & -\lambda - \tfrac{1}{n_i+P_i} = \mu_i \geq 0 \\ \Rightarrow \quad & P_i > 0 \\ \Rightarrow \quad & \lambda + \tfrac{1}{n_i+P_i} = 0 \\ \Rightarrow \quad & P_i = -\tfrac{1}{\lambda} - n_i \end{aligned}$$

If $-\lambda > \frac{1}{n_i}$, then

$$\lambda + \frac{1}{n_i + P_i} \neq 0 \Rightarrow P_i = 0.$$

In conclusion, if we know $\lambda$, then $P_i$ could be expressed as

$$P_i = \begin{cases} -\frac{1}{\lambda} - n_i & \text{for} \quad -\lambda \leq \frac{1}{n_i} \\ 0 & \text{for} \quad -\lambda \geq \frac{1}{n_i} \end{cases} \tag{12.15}$$

Explanation of the allocation scheme of $P_i$: Assume we know $\lambda$, which is the penalty of $\sum_i P_i \geq \bar{P}$. If the noise in one channel, $n_i > -\frac{1}{\lambda}$ , then $p_i = 0$, which means we will not allocate energy to this channel; If $n_i \leq -\frac{1}{\lambda}$, then we'll allocate $P_i = -\frac{1}{\lambda} - n_i$ to this channel. The shaded part in figure 12.3 shows how much power is allocated to each channel.

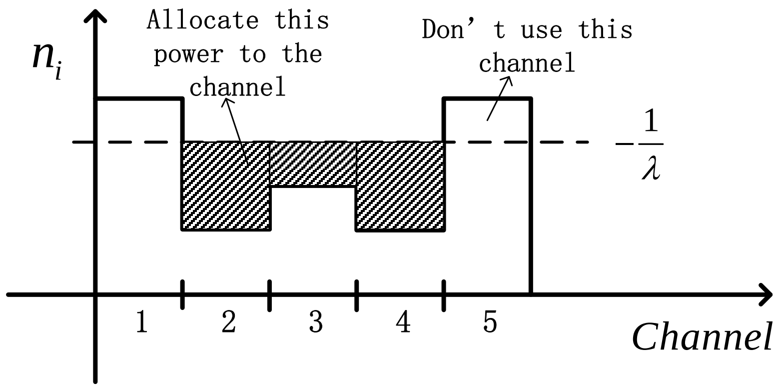


**Figure 12.3.** Illustration of waterfilling strategy

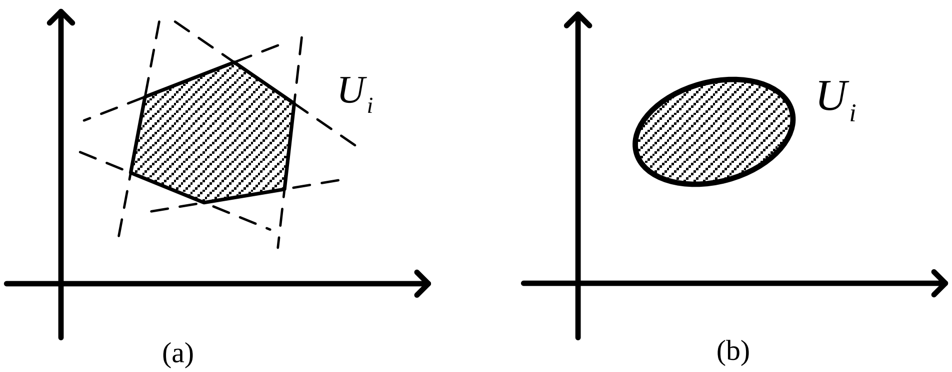


**Figure 12.4.** Uncertain region of coefficients in robust optimization.(a) $\mathcal{U}$ region in case 1;(b) $\mathcal{U}$ in case 2

## 12.4.2 Robust Optimization Problem

Consider the optimization problem as follows:

$$\begin{array}{ll} \min & c^T x \\ s.t. & x \in \mathfrak{X} \end{array} \tag{12.16}$$

In most applications, $a_i$ is only approximately known. For example, Stochastic Programming deals with the problems such as $P(a_i^T x \leq b_i) \geq 1 - \varepsilon$. Robust Optimization is another approach, that its goal is to optimize $c^T x$ with constraints $a_i^T x \leq b_i$ , $\forall a_i \in \mathcal{U}_i$. The optimization problem should justify every value in a set, which is shown in figure 12.4. Thus there are infinitely many constraints. Here we study two cases of different $\mathcal{U}$.

**Case 1:** $\mathcal{U}_i$ is a polytope, as shown in figure 12.4($a$), $\mathcal{U}_i = \{a_i : D_i a_i \leq d_i\}$. Then $x$ satisfies $a_i^T x \leq b_i$ , $\forall a_i \in U_i$ is equivalent to $\begin{bmatrix} \max : a_i^T x \\ s.t : a_i \in U_i \end{bmatrix} \leq b_i$ By Strong Duality, we could see that the following two optimization problems are equivalent:

$$\begin{bmatrix} \max & {a_i}^T x \\ s.t. & D_i a_i \leq d_i \end{bmatrix} = \begin{bmatrix} \min & {P_i}^T d_i & & \\ s.t. & {P_i}^T D & = & x \\ & P_i & \geq & 0 \end{bmatrix} \tag{12.17}$$

After changing the constraints, we now have a new optimization problem as the original one,

$$\begin{array}{lrcl} \min & c^T x & & \\ s.t. & P_i^T d_i & \leq & b_i \\ & P_i^T D & = & x \\ & P_i & \geq & 0 \\ & x & \in & \mathfrak{X} \end{array} , \tag{12.18}$$

which is a usual Linear Programming problem that we already know how to solve.
**Case 2:** $\mathcal{U}_i$ is an ellipse, then the constraints are $(a_i + \hat{a}_i)^T x \leq b_i \quad \forall \hat{a}_i \in \mathcal{U}_i$. $\mathcal{U}_i$ can be written as

$$\begin{array}{rcl} U_i & = & \{\hat{a}_i : \hat{a}_i{}^T Q_i \hat{a}_i + q_i^T \hat{a}_i - d_i \leq 0\} \\ & = & \{\hat{a}_i : \hat{a}_i{}^T Q_i \hat{a}_i \leq d_i\} \\ & = & \{\hat{a}_i : \|\hat{a}_i\|_{Q_i}^2 \leq d_i\} \end{array} \tag{12.19}$$

The previous constraint of $x$,

$$(a_i + \hat{a}_i)^T x \leq b_i \ , \hat{a}_i \in \mathcal{U}_i \ ,$$

is now equivalent to

$$a_i^T x + \left[ \begin{array}{ll} \max & \hat{a}_i{}^T x \\ s.t. & \|\hat{a}_i\|_{Q_i}^2 \leq d_i \end{array} \right] = a_i^T x + d_i \|x\|_{\mathfrak{X}} \leq b_i$$

Here $\|\cdot\|_{\mathfrak{X}}$ is the Dual Norm. Now we can rewrite the optimization problem as:

$$\begin{array}{ll} \max & c^T x \\ s.t. & a_i^T x + d_i \|x\|_{\mathfrak{X}} \leq b_i \end{array}$$

Conclusion: For any convex set $\mathcal{U}_i$ for which we can solve,

$$\left[ \begin{array}{ll} \max & a^T x \\ s.t. & a \in \mathcal{U} \end{array} \right] \tag{12.20}$$

produces a tractable Robust Optimization problem.