Lecture 11 — October 2

Lecturer: Sanghavi *Scribe: Ahmed Alkhateeb, Derya Malak and Mandar Kulkarni*

11.1 Recap

In the previous lecture, we studied duality as a tool to solve constrained optimization problems. We saw that the dual problem can be used to get a lower bound on the optimal value in general. We also saw that if Slater's conditions are satisfied then the dual problem gives exactly same value as the original primal. An example on matrix games was given in order to motivate the utility of re-writing the primal in it's dual form. In this class, we will give another example where duality helps us formulate simple optimization problems from complex looking primals. This example would be that of robust optimization. Following this we will look at some necessary conditions for arriving at optimal solutions.

11.2 Robust Optimization

11.2.1 Motivation

Optimization problems often contain uncertain parameters, due to either measurement/rounding errors, estimation errors or implementation errors. Till now, we have looked at convex optimization problems involving known constraints. In order to model a real-life optimization problem, however, we need to incorporate uncertainty in the parameters involved. In a linear robust optimization, the goal is to optimize $\boldsymbol{c}^\intercal \boldsymbol{x}$ with respect to constraints in the form $\boldsymbol{a}_i^\intercal \boldsymbol{x} \leq b_i$, $\forall \boldsymbol{a}_i \in \mathcal{U}_i$, where $\mathcal{U}_i$ is the uncertainty set of the parameter $\boldsymbol{a}_i$. There might be infinitely many constraints.

Standard LP

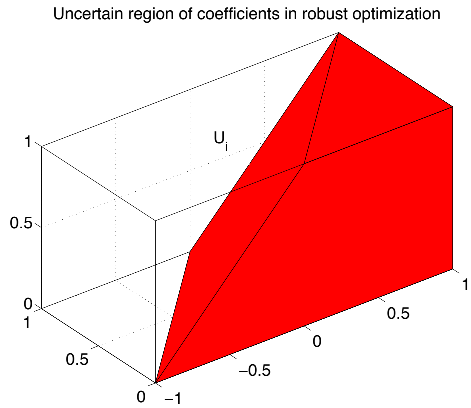
$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\intercal \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^\intercal \boldsymbol{x} \leq b_i \quad \forall i \end{aligned} \tag{11.1}$$

Robust LP

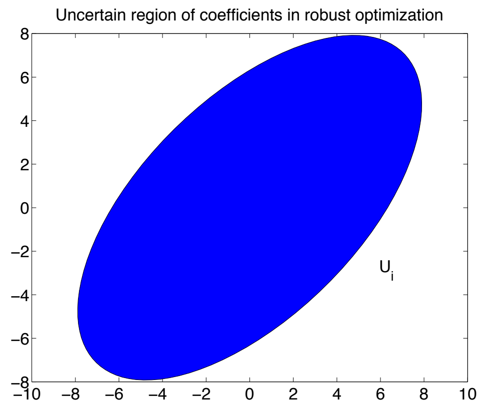
$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\intercal \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^\intercal \boldsymbol{x} \leq b_i, \quad \forall \boldsymbol{a}_i \in \mathcal{U}_i, \end{aligned} \tag{11.2}$$

where $\mathcal{U}_i$ is the uncertainty set of $\boldsymbol{a}_i$.

We study two examples where the uncertainty set is a polytope and an ellipse, and show that the robust optimization problem can be solved as a simple LP thanks to duality.



(a)



(b)

Figure 11.1. Uncertain region of coefficients $\mathcal{U}_i$ in robust optimization. (a) polytope, (b) ellipse.

11.2.2 Robust Optimization on Polytope

Approach 1

The uncertainty constraints in (11.2) can be written as a baby LP as follows:

$$\begin{array}{c} \boldsymbol{a}_i^\mathsf{T}\boldsymbol{x} \le b_i \\ \forall \boldsymbol{a}_i \in \mathcal{U}_i \end{array} \iff \begin{pmatrix} \max\limits_{\boldsymbol{a}_i} & \boldsymbol{a}_i^\mathsf{T}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{a}_i \in \mathcal{U}_i \end{pmatrix} \le b_i \tag{11.3}$$

Let's say that $\mathcal{U}_i = \{\boldsymbol{a}_i : \boldsymbol{D}_i\boldsymbol{a}_i \le \boldsymbol{d}_i\}$.

$$\begin{pmatrix} \max\limits_{\boldsymbol{a}_i} & a_i^\mathsf{T}\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{D}_i\boldsymbol{a}_i \le \boldsymbol{d}_i \end{pmatrix} \le b_i \tag{11.4}$$

Trying to incorporate (11.3) in the original optimization problem in (11.2), we may think of the problem as the following program

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{a}_i} \quad & \boldsymbol{c}^\mathsf{T}\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{a}_i^\mathsf{T}\boldsymbol{x} \le b_i \\ & \boldsymbol{D}_i\boldsymbol{a}_i \le \boldsymbol{d}_i \quad \forall i, \end{aligned} \tag{11.5}$$

which is no more a LP. Furthermore, the constraint $\boldsymbol{a}_i^\mathsf{T}\boldsymbol{x} \le b_i$ has to be satisfied $\forall i$, and in this formulation, there is no way to correctly incorporate this constraint.

Approach 2

Since Approach 1 is not tractable, we consider the dual of baby LP.

$$
\begin{array}{ll}
\text{Primal LP} & \text{Dual LP} \\
\begin{aligned}
\max_{\boldsymbol{a}_i} \quad & \boldsymbol{a}_i^\intercal \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{D}_i \boldsymbol{a}_i \leq \boldsymbol{d}_i \quad \forall i
\end{aligned}
&
\begin{aligned}
\min_{\boldsymbol{p}_i} \quad & \boldsymbol{p}_i^\intercal \boldsymbol{d}_i \\
\text{s.t.} \quad & \boldsymbol{D}_i^\intercal \boldsymbol{p}_i = \boldsymbol{x} \\
& \boldsymbol{p}_i \geq \boldsymbol{0} \quad \forall i
\end{aligned}
\end{array}
\tag{11.6}
$$

Since the primal problem is a linear optimization problem, Strong Duality holds, and the primal and dual LPs have equal optimal value.

Therefore, the constraint

$$
\underbrace{\begin{pmatrix} \max\limits_{\boldsymbol{a}_i} & \boldsymbol{a}_i^\intercal \boldsymbol{x} \\ \text{s.t.} & \boldsymbol{D}_i \boldsymbol{a}_i \leq \boldsymbol{d}_i \end{pmatrix}}_{\forall \boldsymbol{a}_i \text{ this should be less than } b_i.} \leq b_i \iff \underbrace{\begin{pmatrix} \min & \boldsymbol{p}_i^\intercal \boldsymbol{d}_i \\ \text{s.t.} & \boldsymbol{D}_i^\intercal \boldsymbol{p}_i = \boldsymbol{x} \\ & \boldsymbol{p}_i \geq \boldsymbol{0} \end{pmatrix}}_{\text{Only one good } \boldsymbol{p}_i \text{ is sufficient.}} \leq b_i
$$

Now, let's put the new form of constraint in original problem in (11.2).

$$
\begin{aligned}
\min_{\boldsymbol{x}, \boldsymbol{p}_i} \quad & \boldsymbol{c}^\intercal \boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{D}_i^\intercal \boldsymbol{p}_i = \boldsymbol{x} \\
& \boldsymbol{p}_i^\intercal \boldsymbol{d}_i \leq b_i \\
& \boldsymbol{p}_i \geq \boldsymbol{0}
\end{aligned}
\tag{11.7}
$$

which is a standard LP with additional variables.

11.2.3 Robust Optimization on Elliptic Uncertainty Region

The uncertainty set $\mathcal{U}_i$ is an ellipse. Then, $(\boldsymbol{a}_i + \hat{\boldsymbol{a}}_i)^\intercal \leq b_i \ \forall \ \hat{\boldsymbol{a}}_i \in \mathcal{U}_i$. $\mathcal{U}_i$ can be written as

$$\mathcal{U}_i = \{\hat{\boldsymbol{a}}_i \mid \hat{\boldsymbol{a}}_i^\intercal \boldsymbol{Q} \hat{\boldsymbol{a}}_i \leq 1\}.$$

The constraint $(\boldsymbol{a}_i + \hat{\boldsymbol{a}}_i)^\intercal \leq b_i$, $\hat{\boldsymbol{a}}_i \in \mathcal{U}_i$ is equivalent to

$$
\boldsymbol{a}_i^\intercal \boldsymbol{x} + \underbrace{\left\{ \begin{array}{ll} \max\limits_{\hat{\boldsymbol{a}}_i} & \hat{\boldsymbol{a}}_i^\intercal \boldsymbol{x} \\ \text{s.t.} & \hat{\boldsymbol{a}}_i^\intercal \boldsymbol{Q} \hat{\boldsymbol{a}}_i \leq 1 \end{array} \right\}}_{\boldsymbol{x}^\intercal \boldsymbol{Q}^{-1} x} \leq b_i,
\tag{11.8}
$$

which is a quadratic constraint.

11.2.4 Example: Robust beamforming

The problem of choosing beamforming weights to minimize the noise power subject to the net array gain and some interference constraints is a common example where robust optimization techniques are necessary. The optimal solution to this problem is very sensitive to the shape of the beam pattern of the array. That is, variations in the beam pattern also cause changes in the optimal beamforming weights. Modeling these variations and finding optimal beamforming vectors is an active area of research.

11.3 Optimality Conditions

So far, we made some comments on the optimal value. Now, we want to develop some conditions under which the optimal solution(s) exists. For that, we will first introduce the complementary slackness which leads to the Karush-Kuhn-Tucker (KKT) optimality conditions.

11.3.1 Complementary Slackness

Consider a general optimization problem

$$\begin{aligned} \min_{\boldsymbol{x}} \ & f_0(\boldsymbol{x}) \\ \text{s.t.} \ & f_i(\boldsymbol{x}) \leq 0 \ \forall i \\ & h_j(\boldsymbol{x}) = 0 \ \forall j \end{aligned} \tag{11.9}$$

This problem may or may not be convex. The Lagrangian function for solving the dual of this problem is given by

$$\mathcal{L}\left(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = f_0(\boldsymbol{x}) + \sum_i \lambda_i f_i(\boldsymbol{x}) + \sum_j \mu_j h_j(\boldsymbol{x}) \tag{11.10}$$

where $\boldsymbol{\lambda} \leq 0$ and there is no constraint on $\boldsymbol{\mu}$. Suppose that $\boldsymbol{x}^\star$ and $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ are the optimal solutions of the primal and dual problems, respectively. If strong duality holds, we get

$$f_0(\boldsymbol{x}^\star) = g(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star), \tag{11.11}$$

$$\stackrel{(a)}{=} \min_{\boldsymbol{x}} f_0(\boldsymbol{x}) + \sum_i \lambda_i^\star f_i(\boldsymbol{x}) + \sum_j \mu_j^\star h_j(\boldsymbol{x}), \tag{11.12}$$

$$\stackrel{(b)}{\leq} f_0(\boldsymbol{x}^\star) + \sum_i \lambda_i^\star f_i(\boldsymbol{x}^\star) + \sum_j \mu_j^\star h_j(\boldsymbol{x}^\star), \tag{11.13}$$

$$\stackrel{(c)}{=} f_0(\boldsymbol{x}^\star) + \sum_i \lambda_i^\star f_i(\boldsymbol{x}^\star), \tag{11.14}$$

where (a) follows from the definition of the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ at the dual optimal solution $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$, and (b) follows because $\boldsymbol{x}^\star$ is one feasible solution for the problem in the second line. Finally, we note that the equality conditions hold at $\boldsymbol{x}^\star$, i.e., $h_j(\boldsymbol{x}^\star) = 0$, which leads to (c).

Now, we know that $\lambda_i^\star \geq 0$ (from the feasibility conditions of the dual problem), and $f_i(\boldsymbol{x}^\star) \leq 0$ (from the feasibility conditions of the primal problem). Therefore, $\lambda_i^\star f_i(\boldsymbol{x}^\star) \leq 0$ $\forall i$. From this note and given the previous equations, we get

$$\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0 \quad \forall i. \tag{11.15}$$

The relation in (11.15) is called complementary slackness, and it implies that

$$f_i(\boldsymbol{x}^\star) < 0 \Rightarrow \lambda_i^\star = 0, \tag{11.16}$$

$$\lambda_i^\star > 0 \Rightarrow f_i(\boldsymbol{x}^\star) = 0. \tag{11.17}$$

Using complementary slackness, we can make some conditions on the optimal solutions as will be explained next.

11.3.2 KKT Conditions

Consider $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$, we can see from the equations (11.11)-(11.14) that $\boldsymbol{x}^\star$ minimizes $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$. This follows from the note that the inequality in equation (11.13) becomes an equality when the complementary slackness holds. Similarly, we can show that $(\boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ minimizes $\mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}, \boldsymbol{\mu})$. We, therefore, conclude that $\mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star)$ is a saddle point, which yields

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = \mathbf{0}, \tag{11.18}$$

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) \leq \mathbf{0}, \tag{11.19}$$

$$\nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star) = \mathbf{0}, \tag{11.20}$$

which result in the following conditions on the optimal solutions $\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\mu}^\star$ of the primal and dual problems

$$\nabla_{\boldsymbol{x}} f_0(\boldsymbol{x}^\star) + \textstyle\sum_i \lambda_i^\star \nabla_{\boldsymbol{x}} f_i(\boldsymbol{x}^\star) + \sum_j \mu_j^\star \nabla_{\boldsymbol{x}} h_j(\boldsymbol{x}^\star) = \mathbf{0}, \tag{11.21}$$

$$f_i(\boldsymbol{x}^\star) \leq \mathbf{0}, \tag{11.22}$$

$$h_j(\boldsymbol{x}^\star) = \mathbf{0}, \tag{11.23}$$

$$\boldsymbol{\lambda}^\star \geq \mathbf{0}, \tag{11.24}$$

$$\boldsymbol{\lambda}^\star f_i(\boldsymbol{x}^\star) = \mathbf{0}. \tag{11.25}$$

These conditions are called Karush-Kuhn-Tucker (KKT) conditions, and they are necessary conditions for the optimal solutions under the *strong duality* assumption. For convex problems, it can be shown that these conditions are also sufficient optimality conditions, which can be used to prove that a certain solution $\boldsymbol{x}$ is the optimal solution of the problem. The following example illustrates how KKT conditions can be used to analytically arrive at the solution of a simple convex problem.

11.3.3 Example: Power allocation using waterfilling

Consider a convex optimization problem of the type

$$\min_{\boldsymbol{p}} \quad \sum_{i=1}^{m} -\log(p_i + n_i) \tag{11.26}$$

$$\text{s.t.} \quad \mathbf{1}^T \boldsymbol{p} = 1 \tag{11.27}$$

$$\boldsymbol{p} \geq \mathbf{0}. \tag{11.28}$$

where $\boldsymbol{p} = [p_1, p_2, \ldots, p_m]^T$. Such a problem arises very often for power allocation in m parallel communication channels. Here p_i can be thought as the transmit power of each channel, which is limited by a total transmit power constraint (normalized to 1 in this problem) and n_i is the noise power over each channel. The minimization problem over $-\sum_{i=1}^{m} \log(p_i + n_i)$ is similar to maximizing the net achievable data rates using Shannon's capacity formula.

It is easy to see that this problem satisfies Slater's conditions and thus strong duality holds. We can thus solve this problem using the dual function

$$L(\boldsymbol{p}, \lambda, \boldsymbol{\mu}) = -\sum_{i=1}^{m} \log(p_i + n_i) + \lambda(\mathbf{1}^T\boldsymbol{p} - 1) - \boldsymbol{\mu}^T\boldsymbol{p}, \tag{11.29}$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}$. For this problem we obtain the following KKT conditions:

$$\boldsymbol{p}^\star \geq \mathbf{0}, \tag{11.30}$$

$$\mathbf{1}^T\boldsymbol{p}^\star = 1, \tag{11.31}$$

$$\boldsymbol{\mu}^\star \geq \mathbf{0}, \mathrm{the complementary slackness conditions} \tag{11.32}$$

$$p_i^* \mu_i^\star = 0,\ \forall i \in \{1, \ldots, m\}, \tag{11.33}$$

$$-\tfrac{1}{p_i^\star + n_i} + \lambda^\star - \mu_i^\star = 0\ \ \forall i \in \{1, \ldots, m\}. \tag{11.34}$$

Note that here $\mu^\star$ is a slack variable and can be removed from computations by restating the KKT conditions as $\boldsymbol{p}^\star \geq \mathbf{0}$, $\mathbf{1}^T\boldsymbol{p}^\star = 1$,

$$p_i^* \left(-\frac{1}{p_i^\star + n_i} + \lambda^\star \right) = 0\ \forall i \in \{1, \ldots, m\} \tag{11.35}$$

and

$$-\frac{1}{p_i^\star + n_i} + \lambda^\star \geq 0\ \ \forall i \in \{1, \ldots, m\}. \tag{11.36}$$

Now consider the following two cases for the value of λ^*:

Case 1: $0 \leq \lambda^\star < \frac{1}{n_i}$

In this case, we need that $p_i > 0\ \ \forall i \in \{1, \ldots, m\}$ in order to satisfy equation (11.36). Thus, from equation (11.35) we require

$$\lambda^\star = \frac{1}{p_i^\star + n_i}, \tag{11.37}$$

that is,

$$p_i^\star = \left(\frac{1}{\lambda^\star} - n_i \right). \tag{11.38}$$

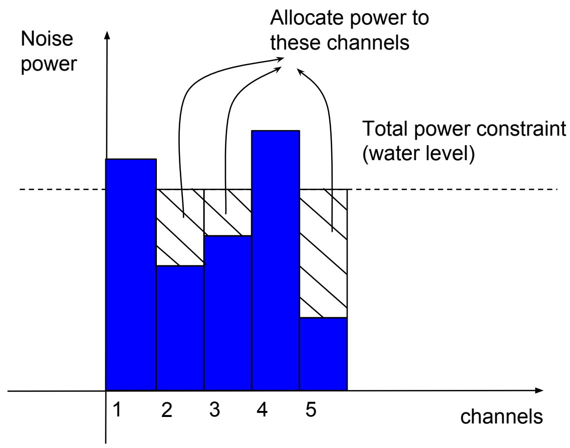


Figure 11.2. Waterfilling for optimal power allocation.

Case 2: $\lambda^\star \geq \frac{1}{n_i}$

In this case it is straightforward to see that in order to satisfy equation (11.35), we need $p_i^\star = 0$.

Thus, the solution to the convex optimization problem is

$$p_i^\star = \max\left\{0, \frac{1}{\lambda^\star} - n_i\right\} \tag{11.39}$$

if the value of $\lambda^\star$ is known. In order to find this, we can use the total power constraint

$$\sum_{i=1}^{m} \max\left\{0, \frac{1}{\lambda^\star} - n_i\right\} = 1. \tag{11.40}$$

A unique solution of $\lambda^\star$ would exist to the above problem, which is monotonically decreasing with $\lambda^\star$. Figure 11.2 illustrates the physical significance of the solution we have obtained. We see that for a fixed water level, which is the maximum transmit power (normalized to 1 in our case), allocating power to different channels for obtaining maximum net achievable data rate is like filling water up to some maximum level. Thus, we see that using KKT conditions it is possible to analytically arrive at solutions to convex optimization problems.