Lecture 10

Convex Optimization and Lagrangian Duality

April 29, 2005
Lecturer: Anna R. Karlin
Notes: Atri Rudra

In this lecture we will cover some basic stuff on Optimization. A very good book for this subject is *Convex Optimization* by Boyd and Vandenberghe. First part of this lecture would follow this book.

10.1 The Lagrangian

Consider the following general optimization problem

$$\begin{aligned} \text{minimize} \quad & f_0(\vec{x}) && (10.1)\\ \text{such that} \quad & f_i(\vec{x}) = 0 \quad i = 1..m && (10.2)\\ & h_j(\vec{x}) = 0 \quad j = 1..p && (10.3)\\ & \vec{x} \in \mathbb{R}^n && (10.4) \end{aligned}$$

Let p^* denote the (value) of the optimal solution to the above program. Further we say $\overline{x}$ is a *feasible* solution if it satisfies (10.2),(10.3) and (10.4). We now define the *Lagrangian*.

$$\mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu}) = f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x}) + \sum_{j=1}^{p} \nu_j h_j(\vec{x})$$

Finally the *Lagranage dual function* is given by

$$g(\vec{\lambda}, \vec{\nu}) = \inf_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu})$$

We now make a couple of simple observations.

Observation. When $\mathcal{L}(\cdot, \vec{\lambda}, \vec{\nu})$ is unbounded from below then the dual takes the value $-\infty$.

Observation. $g(\vec{\lambda}, \vec{\nu})$ is concave[1] as it is the infimum of a set of affine[2] functions.

If $\overline{x}$ is feasible solution of program (10.2)- (10.4), then we have the following

$$\begin{aligned} \mathcal{L}(\overline{x}, \vec{\lambda}, \vec{\nu}) &= f_0(\overline{x}) + \textstyle\sum_{i=1}^{m} \lambda_i f_i(\overline{x}) + \sum_{j=1}^{p} \nu_j h_j(\overline{x}) \\ &\leq f_0(\overline{x}) \qquad\qquad \text{for } \vec{\lambda} \geq 0 \end{aligned}$$

[1] A function $g(x)$ is concave is for any $0 \leq \alpha \leq 1$, $\alpha g(x) + (1-\alpha) g(y) \leq g(\alpha x + (1-\alpha) y)$.

[2] That is, linear in $\{\lambda_i\}$ and $\{\nu_j\}$.

Thus for $\vec{\lambda} \geq 0$, we have that

$$g(\vec{\lambda}, \vec{\nu}) = \min_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}, \vec{\nu}) \leq f_0(\vec{x^*}) = p^*$$

We now define the *Lagrange Dual Program*

$$\begin{array}{rl} \text{maximize} & g(\vec{\lambda}, \vec{\nu}) \\ \text{such that} & \vec{\lambda} \geq 0 \end{array}$$

Let d^* be the value of the optimal solution for the above program. We had just argued that $d^* \leq p^*$. This is the weak duality.

For almost all convex optimization problem, strong duality holds, that is, $d^* = p^*$.

10.1.1 Constraint Qualification

We now define Slater's condition which implies that the strong duality holds.

Definition 10.1 (Slater's Condition). $\exists \vec{x}$ such that $\vec{x}$ is in the interior of the feasible region, that is,

$$\exists \vec{x} \text{ such that } \forall i = 1..m;\ f_i(\vec{x}) < 0 \text{ and } \forall j = 1..p; h_j(\vec{x}) = 0$$

10.1.2 LP Duality

Let us consider the familiar example of LP duality. For that consider a typical LP:

$$\begin{array}{rll} \text{minimize} & \vec{c}^T \vec{x} & \\ \text{such that} & \vec{b} - A\vec{x} & \leq 0 \\ & -\vec{x} & \leq 0 \end{array}$$

The Lagrangian of the above LP is given by

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

where in the above, $\vec{\lambda} = \langle \vec{\lambda_1}, \vec{\lambda_2} \rangle$ and note that we do not need $\vec{\nu}$. The corresponding dual is as follows

$$g(\vec{\lambda}) = \vec{\lambda_1}^T \vec{b} + \inf_{\vec{x}} (\vec{c}^T - \vec{\lambda_1}^T A - \vec{\lambda_2}^T)\vec{x}$$

which is $\vec{\lambda_1}^T \vec{b}$ if $(\vec{c}^T - \vec{\lambda_1}^T A - \vec{\lambda_2}^T) = 0$ and $-\infty$ otherwise.

Thus, the lagrange dual program is the following

$$\begin{array}{rcl} \text{maximize} & \vec{\lambda_1}^T \vec{b} & \\ & (\vec{c}^T - \vec{\lambda_1}^T A - \vec{\lambda_2}^T) & = 0 \\ & \vec{\lambda_1}, \vec{\lambda_2} & \geq 0 \end{array}$$

Now rewriting the first condition we have $\vec{\lambda}1^T A + \vec{\lambda_2}^T = \vec{c}^T$, which along with the condition that $\vec{\lambda_2} \geq 0$ implies that $\vec{\lambda_1}^T A \leq \vec{c}^T$ which gives us the familiar form $A^T \vec{\lambda_1} \leq \vec{c}$.

10.1.3 More on Duality

By definition, the optimal value of the dual is give by

$$d^* = \text{sup}_{\vec{\lambda} \geq 0} \text{inf}_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda})$$

In what is to follow we will drop the dependence on $\vec{\nu}$ for simplicity (all the results presented also hold when $\vec{\nu}$ is present). We have the following claim.

Claim 1.

$$p^* = \text{inf}_{\vec{x}} \ \text{sup}_{\vec{\lambda} \geq 0} \mathcal{L}(\vec{c}, \vec{\lambda})$$

Proof. By definition, we have

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x})$$

Thus, we have

$$\text{sup}_{\vec{\lambda} \geq 0} \mathcal{L}(\vec{x}, \vec{\lambda}) = \begin{cases} f_0(\vec{x}) & \text{if } f_i(\vec{x}) \leq 0 \ \forall i \\ \infty & \text{otherwise} \end{cases}$$

By definition $p^* = \text{inf}_{\vec{x}} f_0(\vec{x})$ and the claim follows. □

Weak Duality states that

$$\text{sup}_{\vec{\lambda} \geq 0} \text{inf}_{\vec{x}} \mathcal{L}(\vec{x}, \vec{\lambda}) \leq \text{inf}_{\vec{x}} \text{sup}_{\vec{\lambda} \geq 0} \mathcal{L}(\vec{c}, \vec{\lambda})$$

For another interpretation consider a continuous zero-sum game where the first player chooses $\vec{x}$, the second player chooses $\vec{\lambda} \geq 0$ and first player pays the second player $\mathcal{L}(\vec{x}, \vec{\lambda})$. Weak duality says that it is better to go second. The famous min-max theorem is a special case of this where the strong duality holds.

10.1.4 Economic interpretation

Let $\vec{x}$ describe how the enterprise works (for example $\vec{x}$ might be the bandwidth allocation), $f_0(\vec{x})$ denote the cost to the enterprise while each $f_i(\vec{x})$ denotes constraints or limits on resources.

What the dual of the lagrangian says is that constraints can be violated if the enterprise is prepared to pay some cost. Let $\vec{\lambda}$ denote the per unit cost (or "price") of each resource. Now consider the term $\lambda_i f_i(\vec{x})$ which can be considered to be the extra payment that the firm has to make. If $f_i(\vec{x}) > 0$ then the firm has to pay more for the violation. On the other hand if $f_i(\vec{x}) < 0$ then the firm can rent out the resource (or recover the cost). Thus, the total cost to the firm is $f_0(\vec{x}) + \sum \lambda_i f_i(\vec{x})$. In other words, $g(\vec{\lambda})$ is the minimum cost to the firm at price $\vec{\lambda}$. d^* can be interpreted as the cost to the firm under the least favorable setting of set of prices. $\vec{\lambda}^*$ are called the *shadow prices*, that is, they are the "correct" prices for the resources so that there is no incentive to deviate).

10.1.5 Complimentary Slackness Conditions

Let $\vec{x}^*, \vec{\lambda}^*$ be the optimal solution to the primal program of (10.2)-(10.4) and say strong duality holds, that is, $p^* = d^*$. This implies that

$$\begin{aligned} f_0(\vec{x}^*) &= g(\vec{\lambda}^*) \\ &= \inf_{\vec{x}}(f_0(\vec{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\vec{x})) \\ &\leq f_0(\vec{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\vec{x}^*) \\ &= f_0(\vec{x}^*) \end{aligned}$$

As the left hand and right hand side of the inequality are the same, we have $\sum_{i=1}^{m} \lambda_i^* f_i(\vec{x}^*) = 0$. Further, as $\vec{x}^*, \vec{\lambda}^*$ is a feasible solution of the lagrangian, we have for all i; $\lambda_i^* \geq 0$ and $f_i(\vec{x}^*) \leq 0$. This implies that for any i, $\lambda_i^* f_i(\vec{x}^*) = 0$. Thus, we have

$$\begin{aligned} \lambda_i^* > 0 &\Rightarrow f_i(\vec{x}^*) = 0 \\ f_i(\vec{x}^*) < 0 &\Rightarrow \lambda_i^* = 0 \end{aligned}$$

Note that the above holds for any primal and dual solution where there is no gap.

KKT Conditions

Assume that the f_i's are differentiable and let $\vec{x}^*, \vec{\lambda}^*$ be the optimal primal solution with no gap, that is, $d^* = p^*$. Since $\vec{x}^*$ minimizes $f_0(\vec{x}) + \sum_i \lambda_i^* f_i(\vec{x})$ the gradient at $\vec{x}^*$ is zero. In other words, $\bigtriangledown f_0(\vec{x}^*) + \sum_i \lambda_i^* \bigtriangledown f_i(\vec{x}^*) = 0$.

Thus, the KKT conditions are–

$$\begin{aligned} \bigtriangledown f_0(\vec{x}^*) + \sum_i \lambda_i^* \bigtriangledown f_i(\vec{x}^*) = 0 & \\ f_i(\vec{x}^*) \leq 0 & \quad \forall i \\ \lambda_i^* \geq 0 & \quad \forall i \\ \lambda_i^* f_i(\vec{x}^*) = 0 & \quad \forall i \end{aligned}$$

The above conditions are necessary but not sufficient for general optimization programs but are sufficient for convex programs.

10.2 Applications of the Lagrangian

10.2.1 Kelly's Procedure

Recall Kelly's procedure which was introduced in the last lecture. Every agent i has a strictly monotone increasing and strictly convex utility function $u_i(\cdot)$. At each iteration, agent i asks for w_i amounts of bandwidth (from a maximum possible of B) and the link sends back a price of $\frac{\sum_j w_j}{B}$. For the next round, the agent i sends in a new demand $w_i' = \max_x[u_i(\frac{x}{p}) - x]$.

We will prove the following theorem.

Theorem 10.1. *The resulting NE from Kelly's procedure maximizes the social welfare, given by the following convex optimization program*

$$\begin{array}{rl} \textit{maximize} & \sum_i u_i(x_i) \\ \textit{subject to} & \sum_i x_i \le B \\ & x_i \ge 0 \qquad \forall i \end{array}$$

Proof. We will use lagrangian duality to prove the theorem– the basic idea being that the optimal condition of the dual of the convex program is same as the solution at which Kelly's procedure stabilizes. First we re-write the convex program in the more 'familiar' form and use lagrangian multipliers.

$$\begin{array}{rll} \text{minimize} & -\sum_i u_i(x_i) & \\ \text{subject to} & \sum_i x_i - B \le 0 & (\times\lambda) \\ & -x_i \le 0 & (\times\lambda_i) \end{array}$$

The Lagrangian is as follows

$$\mathcal{L}(\vec{x}, \lambda, \{\lambda_i\}) = -\sum_i u_i(x_i) + \lambda(\sum_i x_i - B) - \sum_i \lambda_i x_i$$

Thus, the optimal solution is given by

$$d^* = \text{sup}_{\lambda \ge 0, \lambda_i \ge 0} \text{inf}_{\vec{x}} \mathcal{L}(\vec{x}, \lambda, \{\lambda_i\})$$

Now note that Slater's conditions holds (take $x_i = \epsilon$ and let $\epsilon \to 0$ and note that $\vec{\epsilon}$ is a point in the interior) and thus, strong duality holds. Now applying the KKT conditions we have

- $\lambda^*(\sum_i x_i^* - B) = 0$. We then the have the following

$$\lambda^* > 0 \Rightarrow \sum_i x_i^* = B$$

- $-\lambda_i^* x_i^* = 0$ which implies the following–

$$x_i^* > 0 \Rightarrow \lambda_i^* = 0$$

- $-u_i'(x_i^*) + \lambda^* - \lambda_i^* = 0$ for all i which implies

$$u_i'(x_i^*) = \lambda^* \text{ if } x_i^* > 0$$

Thus, if we associate $p^* = \lambda^*$ and $x_i^* = \frac{w_i^*}{p^*}$ we are done. Recall that in the stabilizing point in Kelly's procedure, for all i, $u_i'(\frac{w_i^*}{p^*}) = p^*$ and all the bandwidth is allocated. □

From the proof above one can think of the Kelly's procedure as a way to solve the dual of the lagrangian using some sort of gradient descent.

10.2.2 The Johari-Tsitsiklis Algorithm

Recall from the last lecture that the Johari-Tsitsiklis algorithm chose $w'_i = \max_{w_i} u_i(\frac{w_i B}{w_i + \sum_{j \neq i} w_j} - w_i)$.

In this section, we will sketch the analysis which shows that the point where the Johari-Tsitsiklis algorithm stabilizes is at least $\frac{3}{4}$ as good as the social optimal.

First note that at the equilibrium the following holds–

$$u'_i(\frac{Bw_i}{w_i + \sum_{j \neq i} w_j})(\frac{B}{w_i + \sum_{j \neq i} w_j} - \frac{Bw_i}{(w_i + \sum_{j \neq i} w_j)^2}) = 1$$

If we set $x_i = \frac{w_i}{p} = \frac{w_i B}{\sum_j w_j}$ the above equation becomes

$$u'_i(x_i) = \frac{p}{1 - \frac{x_i}{B}}$$

For the analysis we will need two more functions which are based on the utility functions $u_i(\cdot)$'s. The first function $z_i(\cdot)$ is the the tangent line at x_i for $u_i(x_i)$, that is $z_i(x_i) = u_i(x_i)$ and the slope of the line $z_i(\cdot)$ is $u'_i(x_i)$. Further, $G_i(\cdot)$ is defined as the line parallel to $z_i(\cdot)$ such that it passes through the origin, that is, $G_i(x) = z_i(x) - c_i$ where $c_i = z_i(0)$. Note that as $u_i(\cdot)$ was monotone increasing and strictly convex, $c_i > 0$.

Let $\vec{x}$ denote the Nash Equilibrium of the procedure and $\vec{x}^*$ are the values of the social optimum. We are thus interested in the quantity

$$\frac{\sum u_i(x_i)}{\sum u_i(x_i^*)} = \frac{\text{Nash}(u_i\text{'s})}{\text{OPT}(u_i\text{'s})}$$

Now consider the case when the utility functions were the functions $z_i(\cdot)$. $\vec{x}$ still remains the NE and since $z_i(\cdot)$ is always strictly larger than $u_i(\cdot)$ (as $u_i(\cdot)$ is strictly convex), we have

$$\frac{\text{Nash}(u_i\text{'s})}{\text{OPT}(u_i\text{'s})} \geq \frac{\text{Nash}(z_i\text{'s})}{\text{OPT}(z_i\text{'s})}$$

As discussed in the previous paragraph, for any point x, $G_i(x) = Z_i(x) - c_i$ where $c_i > 0$, which gives us $\text{Nash}(G_i\text{'s}) - \text{Nash}(z_i\text{'s}) = \text{OPT}(G_i\text{'s}) - \text{OPT}(z_i\text{'s}) = \sum_i c_i > 0$. Finally the fact that $\text{OPT}(z_i\text{'s}) \geq \text{OPT}(G_i\text{'s})$ implies that

$$\frac{\text{Nash}(z_i\text{'s})}{\text{OPT}(z_i\text{'s})} \geq \frac{\text{Nash}(G_i\text{'s})}{\text{OPT}(G_i\text{'s})}$$

By the definition of $G_i(\cdot)$ we have

$$G_i(g_i) = \frac{p}{1 - \frac{x_i}{B}} g_i$$

Assume w.l.o.g assume that $u'_1(x_1) = \frac{p}{1 - \frac{x_1}{B}} \geq u'_i(x_i)$ for any $i \geq 2$. It is not to hard to see that the $\vec{x}$ is still the NE for the utilities $G_i(\cdot)$. Finally, as $G_i(\cdot)$ is a linear function, the optimal allocates all of B to agent 1 (as it has the largest slope). All this implies that

$$\frac{\text{Nash}(G_i\text{'s})}{\text{OPT}(G_i\text{'s})} = \frac{(\frac{p}{1 - \frac{x_1}{B}})x_1 + [\sum_{j \geq 2} \frac{p}{1 - \frac{x_j}{B}} x_j]}{(\frac{p}{1 - \frac{x_1}{B}})B}$$

One can show that the largest the latter ratio can be is $\frac{3}{4}$.

The crucial thing in the above analysis was to reduce the problem to the linear case. One can show that this ratio of $\frac{3}{4}$ is tight by the following example. Let $u_1(x) = x$ and $u_i(x) = \frac{x}{2}$ where $i = 2..n+1$ with $B = 1$. The optimal solution is to assign everything to agent 1 while in the Nash Equilibrium, $\frac{1}{2}$ is allocated to agent 1 and $\frac{1}{2n}$ bandwidth to the other n agent which gives a ratio of $\frac{3}{4}$.