Mathematical methods in communication	January 24th, 2012

Lecture 11

Lecturer: Haim Permuter — *Scribe: Moti Teitel*

A. Lagrange Duality

The primary and dual functions are defined as following:

Primary:

$$\begin{aligned} \text{minimize } & f_o(x) \\ \text{subject to } & f_i(x) \leq 0 \;\; i = 1, 2, \ldots, m \\ & h_j(x) = 0 \;\; p = 1, 2, \ldots, p \end{aligned} \tag{1}$$

Dual:

$$\begin{aligned} \text{maximize } & g(\lambda, \upsilon) \\ \text{subject to } & \lambda \geq 0 \end{aligned} \tag{2}$$

The **Lagrangian** is defined as following:

$$L(x, \lambda, \upsilon) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x) \tag{3}$$

The **Lagrangian dual function** is defined as following:

$$g(\lambda, \upsilon) = \inf_{x \in D} L(x, \lambda, \upsilon) = \inf_{x \in D} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x) \right] \tag{4}$$

B. Properties

1) $L(x, \lambda, \upsilon)$ is convex in x, linear in λ and in υ .
2) $g(\lambda, \upsilon)$ is concave in λ, υ as an infimum over linear functions of λ, υ, **regardless of the convexity of the primary problem!**
3) for all $\lambda \geq 0$, we have $g(\lambda, \upsilon) \leq p^*$, where p^* is the optimal value of the prime problem.

Proof: Let x^* be the argument that achieves the optimal value of p^*, namely $f_0(x^*) = p^*$, and $f_i(x^*) \le 0, h_i(x^*) = 0$. Since $\lambda \ge 0$ and $f_i(x^*) \le 0$ we have

$$\sum_{i=1}^{m} \lambda_i f_i(x^*) \le 0. \tag{5}$$

Since $h_i(x^*) = 0$ we have

$$\sum_{i=1}^{p} \upsilon_i h_i(x^*). \tag{6}$$

So by Eq. (5) and (6) we have

$$\sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{i=1}^{p} \upsilon_i h_i(x^*) \le 0. \tag{7}$$

Then,

$$\begin{aligned} g(\lambda, \upsilon) &= \inf_{x \in D} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \upsilon_i h_i(x) \right] \qquad (8) \\ &\le f_0(x^*) + \sum_{i=1}^{m} \lambda_i f_i(x^*) + \sum_{i=1}^{p} \upsilon_i h_i(x^*) \overset{(a)}{\le} f_0(x^*) \overset{(b)}{=} p^*. \end{aligned}$$

where (a) follows from (5), and (b) is by definition.

4) Since the above is true for all $\lambda \ge 0$, we have

$$\max_{\lambda \ge 0} g(\lambda, \upsilon) \le p^* \tag{9}$$

5) For convex primary problems, under the condition that there exists some $x \in \mathrm{dom} f$ such that $f_i(x) < 0$ for all i (strong inequality), then:

$$\max_{\lambda \ge 0} g(\lambda, \upsilon) = p^* \tag{10}$$

■

6) Any solution of the primary problem (not necessarily the minimal value) is an upper bound for the solution. Any solution of the dual problem is a lower bound.

C. Examples

1) **Primary**:

$$\text{minimize } \mathbf{x}^T\mathbf{x} \tag{11}$$
$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

So we have

$$\begin{aligned} L(\mathbf{x}, \lambda, \upsilon) &= \mathbf{x}^T\mathbf{x} + \upsilon^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ \frac{\partial}{\partial \mathbf{x}} L(x, \lambda, \upsilon) = 0 \Rightarrow \mathbf{x}_{min} &= -\frac{1}{2}\mathbf{A}\upsilon \\ g(\lambda, \upsilon) = L(\mathbf{x}_{min}, \lambda, \upsilon) &= -\frac{1}{4}\upsilon^T\mathbf{A}\mathbf{A}^T\upsilon - \upsilon^T\mathbf{b} \end{aligned} \tag{12}$$

and the **dual** problem is:

$$\max_{\upsilon} \left[-\frac{1}{4}\upsilon^T\mathbf{A}\mathbf{A}^T\upsilon - \upsilon^T\mathbf{b} \right] \tag{13}$$

Since the primary problem is convex (and there are no inequality constraints), we have that $\max_{\upsilon} g(\upsilon) = \min_{\mathbf{x}} f_0(x)$ with the constraints satisfied.

2) **Primary**:

$$\text{minimize } \mathbf{c}^T\mathbf{x} \tag{14}$$
$$\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$
$$-\mathbf{x} \leq 0$$

So we have

$$\begin{aligned} L(\mathbf{x}, \lambda, \upsilon) &= \mathbf{c}^T\mathbf{x} - \lambda^T\mathbf{x} + \upsilon^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ g(\lambda, \upsilon) = \inf_{\mathbf{x}} L(\mathbf{x}_{min}, \lambda, \upsilon) &= \begin{cases} -\upsilon^T\mathbf{b}, & \mathbf{c}^T - \lambda^T + \upsilon^T\mathbf{A} = 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned} \tag{15}$$

So,

$$g(\lambda, \upsilon) = \max_{\upsilon} \left[-\upsilon^T\mathbf{b} \right] \tag{16}$$

$$\text{subject to} \begin{cases} \mathbf{c}^T - \lambda^T + \upsilon^T \mathbf{A} = 0 \\ \lambda \geq 0 \end{cases}$$

Since λ is not a part of the maximization, we can write

$$g(\lambda, \upsilon) = \max_{\upsilon} \left[-\upsilon^T \mathbf{b} \right] \qquad (17)$$
$$\text{subject to } \mathbf{c}^T + \upsilon^T \mathbf{A} = 0$$

3) **Primary**:

$$\text{minimize } f_0(x) \qquad (18)$$
$$\text{subject to } \begin{cases} \mathbf{Ax} \leq \mathbf{b} \\ \mathbf{Cx} = \mathbf{d} \end{cases}$$

So we have

$$L(\mathbf{x}, \lambda, \upsilon) = f_0(x) + \lambda^T (\mathbf{Ax} - \mathbf{b}) + \upsilon^T (\mathbf{Cx} - \mathbf{d}) \qquad (19)$$
$$g(\lambda, \upsilon) = \inf_{\mathbf{x} \in \mathrm{dom} f_0} L(\mathbf{x}_{min}, \lambda, \upsilon)$$

We simplify $g(\lambda, \upsilon)$:

$$\begin{aligned} g(\lambda, \upsilon) &= \inf_{\mathbf{x} \in \mathrm{dom} f_0} \left[f_0(x) + \lambda^T (\mathbf{Ax} - \mathbf{b}) + \upsilon^T (\mathbf{Cx} - \mathbf{d}) \right] \qquad (20) \\ &= -\sup_{x \in \mathrm{dom} f_0} \left[-\left(\lambda^T \mathbf{A} + \upsilon^T \mathbf{C} \right) \mathbf{x} - f_0(\mathbf{x}) + \lambda^T \mathbf{b} + \upsilon^T \mathbf{d} \right] \\ &= -f^* \left(-\lambda^T \mathbf{A} + \upsilon^T \mathbf{c} \right) - \lambda^T \mathbf{b} - \upsilon^T \mathbf{d} \end{aligned}$$

where $f^*(y) = \max_x [yx - f(x)]$ is the conjugate function.

4) Entropy maximization **Primary**:

$$\text{minimize } \sum_{i=1}^{n} x_i \log x_i \qquad (21)$$
$$\text{subject to } \begin{cases} \mathbf{Ax} \leq \mathbf{b} \\ \mathbf{1}^T \mathbf{x} = 1 \end{cases}$$

So we have

$$L(\mathbf{x},\lambda,\upsilon)=\sum_{i=1}^{n}x_i\log x_i+\lambda^T(\mathbf{A}\mathbf{x}-\mathbf{b})+\upsilon\left(\mathbf{1}^T\mathbf{x}-1\right) \tag{22}$$

Taking the derivative in order to find the minmimum, we have

$$\frac{\partial}{\partial x_i}L(\mathbf{x},\lambda,\upsilon)=\log x_i+1+\lambda^T\mathbf{a}_i+\upsilon=0 \tag{23}$$

so that the minimal value is obtained for

$$x_i=e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)} \tag{24}$$

. substituting this into (22) we have

$$\begin{aligned} g(\lambda,\upsilon)=L(\mathbf{x}_{min},\lambda,\upsilon)&=\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}\log e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)} \\ &+\lambda^T\left(\sum_{i=1}^{n}\mathbf{a}_ie^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}-\mathbf{b}\right) \\ &+\upsilon\left(\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}-1\right) \\ &=-\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}-\lambda^T\mathbf{b}-\upsilon \end{aligned} \tag{25}$$

So the dual problem is

$$\begin{aligned} &\max_{\lambda,\upsilon}\left[-\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}-\lambda^T\mathbf{b}-\upsilon\right] \\ &\text{subject to } \lambda\geq 0 \end{aligned} \tag{26}$$

We can find the optimal υ analytically:

$$\begin{aligned} \frac{\partial}{\partial\upsilon}g(\lambda,\upsilon)&=\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}-1=0 \\ \Rightarrow \upsilon_{opt}&=\log\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i\right)} \end{aligned} \tag{27}$$

Substituting this into the maximization problem, and recalling that $\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i-\upsilon\right)}=1$ from (27), the dual problem can be written as

$$\max_{\lambda}\left[-1-\lambda^T\mathbf{b}-\upsilon_{opt}\right]=\max_{\lambda}\left[-1-\lambda^T\mathbf{b}-\log\sum_{i=1}^{n}e^{\left(-1-\lambda^T\mathbf{a}_i\right)}\right] \tag{28}$$

So finally the optimization problem can be written as

$$\max_{\lambda} \left[-\lambda^T \mathbf{b} - \log \sum_{i=1}^{n} e^{\left(-1-\lambda^T \mathbf{a}_i\right)} \right] \tag{29}$$
$$\text{subject to } \lambda \geq 0$$

I. KKT CONDITIONS AND DUAL FUNCTIONS

Let (λ^*, υ^*) be the argument that achieves the maximum value of the dual problem, denoted by d^*. Let x^* be the argument that achieves the maximum of the primary problem, denoted by p^*. Assuming that strong duality holds ($p^* = d^*$), we may write

$$f_0(x^*) = g(\lambda^*, \upsilon^*) = \inf_{x \in \text{dom} f} \left[f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \upsilon_i^* h_i(x) \right] \tag{30}$$
$$\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \upsilon_i^* h_i(x^*) \overset{(a)}{\leq} f_o(x^*),$$

where (a) follows from (7), so that all the inequalities are equalities (as we started with $f_0(x^*)$ and ended with $f_0(x^*)$). Is so, we have that

$$f_0(x^*) = f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \upsilon_i^* h_i(x^*). \tag{31}$$

Theorem 1 KKT Conditions: Let x^* and (λ^*, υ^*) be any primal and dual optimal points which satisfy $p^* = d^*$. Since x^* minimizes $L(\lambda^*, \upsilon^*, x)$ over x, we must have

$$0 = \nabla_x L(\lambda, \upsilon, x) = \nabla_x f_0(x^*) + \sum_i \lambda_i^* \nabla_x f_i(x^*) + \sum_i \upsilon_i^* \nabla_x h_i(x^*) \tag{32}$$

Thus necessary conditions for x^* and (λ^*, υ^*) to be primal and dual optimal points are

$$f_i(x^*) \leq 0 \quad \forall i \tag{33}$$
$$h_i(x^*) = 0 \quad \forall i \tag{34}$$
$$\lambda_i^* \geq 0 \quad \forall i \tag{35}$$
$$\lambda_i^* f_i(x^*) = 0 \quad \forall i \tag{36}$$
$$\nabla_x f_0(x^*) + \sum_i \lambda_i^* \nabla_x f_i(x^*) + \sum_i \upsilon_i^* \nabla_x h_i(x^*) = 0 \tag{37}$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

If the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal.