10-725/36-725: Convex Optimization **Fall 2015**

Lecture 16: Duality revisited

Lecturer: Javier Pena *Scribes: Susu Xu, Xidong Pi, Yongjin Cho*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

16.1 Last time: barrier method

16.1.1 Barrier method

The main idea of barrier method is to approximate the indicator function $I_C(x)$ restricted to a set C with barrier problem. Then central idea of barrier method is to extend Newton's method to a problem with equality constraints and inequality constraints. The inequality constraints are difficult constraints to deal with. The barrier problem is easier to solve compare to the original problem.
The original problem:

$$\begin{aligned} min_x \quad & f(x) + I_C(x) \\ s.t. \quad & Ax = b \end{aligned}$$

The barrier problem:

$$\begin{aligned} min_x \quad & f(x) + \frac{1}{t}\phi(x) \\ s.t. \quad & Ax = b \end{aligned}$$

where $t > 0$ and ϕ is a barrier function for C.
The basic case is

$$C = \{x : h_i(x) \leq 0, i = 1, \cdots, m\}$$

and there is a canonical barrier function, which is a logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} log(-h_i(x))$$

Note: this is for convex problem, so the objective function f and inequality constraints are convex and smooth functions. The KKT conditions hold for both orginal and barrier problem. From KKT conditions, we get the solution to the barrier problem which satisfies

$$f(x^\star(t)) - f^\star \leq m/t$$

The optimality gap is bounded by m/t. Big t here means there are more weight on the objective function, and it will give a good approximation to the original problem.

16.1.2 Strict feasibility

An important detail here is that the barrier method essentially by construction we enforce the constraints throughout the algorithm. When doing Newton's method and line search, we need to make sure the step length is small enough that the constraints are not violated. In principle, to get started, we need a strictly feasible point. The start point should satisfy the inequalities strictly as well as the quality constraints. That is, find an initial x such that

$$\begin{aligned} h_i(x) &\leq s \\ Ax &= b \end{aligned}$$

In simple cases, it is very straightforward to start with a point. In other cases, we can set up phase I problem. If there are solutions that s is negative, then that automatically gives an initial strictly physical point. Otherwise the original problem is infeasible.

16.2 Lagrangian duality revisited

For the primal problem:

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & h_i(x) \leq 0, i = 1, \cdots, m \\ & l_j(x) = 0, j = 1, \cdots, r \end{aligned}$$

This can be any problem. Associated with Lagrangian function, we have

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + \sum_{j=1}^{r} v_j l_j(x)$$

We can rewrite the primal problem as

$$p := \min_x \max_{u,v} L(x, u, v)$$

The dual problem is

$$d := \max_{u,v} \min_x L(x, u, v)$$

16.2.1 Weak and strong duality

The key fact is that $p \geq d$. That is, the optimal primal is an upper bound of the optimal dual.
For strong duality, if we assume $f, h_i(x)$ are convex, and some inequality constraints and all equality constraints are affine. If there exist a point that satisfies the non-affine inequality constraints strictly, and satisfies the other constraints, then the strong duality holds. For $h_1, \cdots, h_p$ are convex with domain D and $h_{p+1}, \cdots, h_m, l_1, \cdots, l_r$ are affine,

$$\begin{aligned} if\ \exists \hat{x} \in relint(D),\ h_i(\hat{x}) < 0,\ i = 1, \cdots, p; \\ h_i(\hat{x}) \leq 0,\ i = p+1, \cdots, m; \\ l_j(\hat{x}) = 0,\ j = 1, \cdots, r \end{aligned}$$

the strong duality holds.

16.2.2 Example: linear programming

Primal and dual problems:

$$\begin{aligned} \min_{x} \quad & c^T x \\ s.t \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} \max_{y,s} \quad & b^T y \\ s.t. \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

16.2.3 Example: convex quadratic programming

The convex quadratic programming problem is similar. The primal and dual problems are:

$$\begin{aligned} \min_{x} \quad & \frac{1}{2} x^T Q x + c^T x \\ s.t \quad & Ax = b \\ & x \geq 0 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} \max_{u,y,s} \quad & b^T y - \frac{1}{2} u^T Q u \\ s.t. \quad & A^T y + s - c = Qu \\ & s \geq 0 \end{aligned}$$

where Q symmetric and positive semi-definite.

The main difference between linear and quadratic programming problem is in quadratic programming problem the objective function includes a quadratic term. And this form of dual highlights a fact that the dual of a convex quadratic programming problem is another convex quadratic programming problem.

16.2.4 Example: barrier problem for linear programming

One more example that connected to what we learned last time is applying the barrier method to linear programming.

Primal problem:

$$\begin{aligned} \min_{x} \quad & c^T x - \tau \sum_{i=1}^{n} \log x_i \\ s.t \quad & Ax = b \end{aligned}$$

where $\tau > 0$ and it corresponds to $1/t$ we used previously.

Dual problem:

$$\begin{aligned} \max_{y,s} \quad & b^T y + \tau \sum_{i=1}^{n} \log s_i + n(\tau - \tau \log \tau) \\ s.t. \quad & A^T y + s = c \end{aligned}$$

Proof: (Derivation of the dual)

Using the notation more conventional in linear programming (dual variable y), the Lagrangian of the primal barrier problem is:

$$L(x, y) = c^T x - \tau \sum_{i=1}^{n} \log x_i + y^T (b - Ax)$$

Then the dual problem should be:

$$\max_y \min_x \quad (c - A^T y)^T x - \tau \sum_{i=1}^n \log x_i + b^T y$$

Firstly we look at the parts with x:

$$\min_x \quad (c - A^T y)^T x - \tau \sum_{i=1}^n \log x_i$$

Let $s = c - A^T y$, then the above problem decouples to:

$$\min_x \quad \sum_{i=1}^n \{s_i - \tau \log x_i\}$$

and then we can solve it for each component separately. The minimizer is $x_i^\star = \tau / s_i$. Hence the dual problem is:

$$\begin{aligned} \max_x \quad & b^T y + n\tau - \tau \sum_{i=1}^n \log \frac{\tau}{s_i} \\ s.t. \quad & A^T y + s = c \end{aligned}$$

which is equivalent to the dual problem we want to prove above after rearranging of terms. ■

One thing to be highlighted here is if we look at both linear programming problem and the barrier problem for linear programming, we'll find the dual of the barrier for the primal problem turns to be some barrier of the dual problem (of standard-form linear programming), if we ignore the $n(\tau - \tau \log \tau)$ term which is just a constant.

16.3 Optimality conditions

Then we look at the optimality conditions of convex problem, that is:

$$\begin{aligned} \min_x \quad & f(x) \\ s.t. \quad & Ax = b \\ & h_i(x) <= 0, \; i = 1, 2, ..., m \end{aligned}$$

Assume $f, h_1, ..., h_m$ convex and differentiable. Also assume that strong duality holds.

Then $x^\star$ and $(u^\star, v^\star)$ are respectively primal and dual optimal solution if and only if $(x^\star, u^\star, v^\star)$ solves the KKT conditions:

$$\begin{aligned} \nabla f(x) + A^T v + \nabla h(x) u = 0 & \text{ (Stationarity)} \\ Ax = b & \text{ (Primal feasible)} \\ U h(x) = 0 & \text{ (Complementarity)} \\ u, -h(x) \geq 0 & \text{ (Primal and dual feasible)} \end{aligned}$$

Here $U = diag(u), \; \nabla h(x) = [\nabla h_1(x), ..., \nabla h_m(s)]$.

16.3.1 Central path equations

Then we look at the optimality conditions for the barrier problem, which are also called central path equations.

Barrier problem:

$$\begin{aligned} \min_x \quad & f(x) + \tau\phi(x) \\ s.t. \quad & Ax = b \end{aligned}$$

where:

$$\phi(x) = -\sum_{i=1}^{m} \log(-h_i(x))$$

Optimality conditions for barrier problem (and its dual):

$$\begin{aligned} \nabla f(x) + A^T v + \nabla h(x) u = 0 \quad & \text{(Stationarity)} \\ Ax = b \quad & \text{(Primal feasible)} \\ Uh(x) = -\tau\mathbf{1} \quad & \text{(Complementarity)} \\ u, -h(x) > 0 \quad & \text{(Primal and dual feasible)} \end{aligned}$$

The 1st and 2nd conditions are identical to the case without barrier. The 3rd condition is not identical, but becomes $u_i h_i(x) = -\tau$. And the 4th condition becomes strictly positive because this is the barrier problem.

At the end of day, we could solve the original optimization problem by equivalently solve the optimality conditions. This is the same for the barrier problem. We'll see more of that in next class.

16.3.2 Special case: linear programming

Primal and dual problems:

$$\begin{aligned} \min_x \quad & c^T x \\ s.t \quad & Ax = b \\ & x \geq 0 \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \max_{y,s} \quad & b^T y \\ s.t. \quad & A^T y + s = c \\ & s \geq 0 \end{aligned}$$

Optimality conditions for both:

$$\begin{aligned} A^T y + s &= c \\ Ax &= b \\ XS\mathbf{1} &= 0 \\ x, s &\geq 0 \end{aligned}$$

Here $X = Diag(x), S = Diag(s)$.

Note in linear programming, we just need feasibility for strong duality to hold, because all inequalities in linear programming are affine. Does the same thing happen with quadratic programming? Does the strong duality automatically holds? The answer: they are essentially the same thing.

16.3.3 Algorithms for linear programming

So if the linear programming problem feasible, then its optimal solution is perfectly characterized by the optimal conditions. Two main classes of algorithms for solving linear programming problem are:

- Simplex: maintain first three conditions and aim for fourth one.
- Interior-point methods: maintain fourth condition (and maybe first and second) and aim for third one.

The simplex algorithm is invented in 1940's and is one of the most influential methods. It is still a very competitive algorithm. And it can be summarized in one sentence as above: it maintains first 3 conditions and $x \geq 0$, and look for $s \geq 0$, and it solves the problem once it finds qualified s. Interior-point method, eg. the barrier method applied to linear programming, it satisfies the 4th condition and violate the 3rd condition, and only satisfies the 3rd condition when it converges. For different implementations of interior-point method, it may or may not satisfy the first two conditions.

16.3.4 Duality gap for barrier problem of linear programming

If $x^\star(\tau), u^\star(\tau), v^\star(\tau)$ solve KKT conditions for barrier problem (Central path equations), the duality gap equals to:

$$f(x^\star(\tau)) - \min_x L(x, u^\star(\tau), v^\star(\tau))$$

From stationarity we know $x^\star(\tau)$ is exactly the minimizer of $L(x, u^\star(\tau), v^\star(\tau))$, hence:

$$\begin{aligned}
& f(x^\star(\tau)) - \min_x L(x, u^\star(\tau), v^\star(\tau)) \\
= \; & f(x^\star(\tau)) - L(x^\star(\tau), u^\star(\tau), v^\star(\tau)) \\
= \; & f(x^\star(\tau)) - [f(x^\star(\tau)) + u^\star(\tau)^T h(x^\star(\tau)) + v^\star(\tau)^T (b - Au^\star(\tau))] \\
= \; & -u^\star(\tau)^T h(x^\star(\tau)) \\
= \; & m\tau
\end{aligned}$$

Since $x^\star(\tau)$ is feasible, $b - Au^\star(\tau) = 0$. And since the complementarity of central path equations, $u^\star(\tau)^T h(x^\star(\tau)) = -m\tau$. The result $m\tau$ is exactly the same of the sub-optimality gap m/t we learned last time.

16.3.5 Central Path for linear Programming

Primal problem :

$$\begin{aligned}
\min_x \quad & c^T x - \tau \sum_{i=1}^n log(x_i) \\
s.t. \quad & Ax = b
\end{aligned}$$

Dual problem :

$$\begin{aligned}
\max_{y,s} \quad & b^T y + \tau \sum_{i=1}^n log(s_i) \\
s.t. \quad & A^T y + s = c
\end{aligned}$$

Optimality conditions for both:

$$\begin{aligned} A^T y + s &= c \\ Ax &= b \\ XS1 &= \tau 1 \\ x, s &> 0 \end{aligned}$$

Compelementarity of the original optimality condition tweaked for this case ($XS1 = \tau 1$).

16.4 Fenchel duality

Fenchel duality is useful in that we can express dual problems explicitly in terms of the conjugates of the primal problem.

Consider the primal problem :

$$\min_x \quad f(x) + g(Ax)$$

Rewrite it as

$$\begin{aligned} \min_x \quad & f(x) + g(z) \\ s.t. \quad & Ax = z \end{aligned}$$

Lagrangian is

$$L(x, z, v) = f(x) + g(z) + v^T(z - Ax)$$

Then Dual problem is

$$\max_v \min_{x,z} \left(v^t z + g(z) - (A^T v)^T x + f(x) \right)$$
$$\text{recall that :} \ f^*(s) := \max_x (s^T x - f(x))$$

Thus, dual problem is

$$\max_v \quad - f^*(A^T v) - g^*(-v)$$

This special type of duality is called Fenchel duality.

Nice Fact : if f, g are convex and closed then the dual of the dual is the primal, because $f^{**} = f$ and $g^{**} = g$.

16.4.1 Example: conic programming

Primal problem (in standard form)

$$\begin{aligned} \min_x \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \in K \end{aligned}$$

where K is a closed convex cone.

Dual problem

$$\begin{aligned} \max_{y,s} \quad & b^T y \\ s.t. \quad & A^T y + s = c \\ & s \in K^* \end{aligned}$$

Two ways to get dual problem :

- Derive dual using Lagrangian
- Fenchel duality

  re-define primal problem:

$$\min_x \quad f(x) + g(Ax)$$

  where

$$\begin{aligned} f(x) &= c^T x + I_K(x) \\ g(z) &= I_{\{b\}}(z) \end{aligned}$$

  Observe $f(x) + g(Ax)$ is:

$$\begin{cases} c^T x & \text{if we have a feasible point} \\ +\infty & \text{otherwise} \end{cases}$$

  Recall : if K is closed convex cone, then

$$\begin{aligned} I_k^*(s) &= \max_x s^x - I_k(x) \\ &= \max_{x \in K} s^T x \\ &= \begin{cases} 0 & = s \in -K^* \\ +\infty & = s \notin -K^* \end{cases} \\ &= I_{-K^*}(s) \end{aligned}$$

Strong duality holds if one of the problem is strictly feasible. In this case, strict feasibility implies that there is a solution that is at the (relative) interior of the cone.

If both primal and dual are strictly feasible, then strong duality holds and both primal and dual optima are attained and are the same.

16.4.2 Example: semidefinite programming

Special case of the conic programming when the space of the problem is the space of symmetric matrices and K in conic programming is cone of positive semidefinite matrices.

Primal

$$\begin{aligned} \min_{x} \quad & C \cdot X \\ s.t. \quad & A_i \cdot X = b_i, i = 1, \ldots, m \\ & X \succeq 0 \end{aligned}$$

Dual

$$\begin{aligned} \max_{y} \quad & b^T y \\ s.t. \quad & \sum_{i=1}^{m} y_i A_i + S = C \\ & S \succeq 0 \end{aligned}$$

Recall trace inner product in $\mathbb{S}^n$

$$X \cdot S = trace(XS)$$

16.4.3 Strong duality does not always hold

Here are the cases in which the strong duality fails to hold in semidefinite programming.

Examples:

$$\begin{aligned} min \quad & 2x_{12} \\ s.t. \quad & \begin{bmatrix} 0 & x_{12} \\ x_{12} & x_{22} \end{bmatrix} \succeq 0 \end{aligned}$$

Translating this example into standard from of SDP, we have

$$\begin{aligned} C &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ A &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \\ b &= 0 \end{aligned}$$

Then dual is

$$\begin{aligned} \max \quad & 0 \cdot y \\ s.t. \quad & \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} y + S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, S \succeq 0 \\ \iff & \begin{bmatrix} -b & 1 \\ 1 & 0 \end{bmatrix} \succeq 0 \end{aligned}$$

$\begin{bmatrix} -y & 1 \\ 1 & 0 \end{bmatrix} \succeq 0$, but it has negative determinant which is $0 * y - 1 = 1$, thus we have a contradiction.

Second example:

$$\begin{aligned} \min \quad & x_{11} \\ s.t. \quad & \begin{bmatrix} x_{11} & 1 \\ 1 & x_{22} \end{bmatrix} \succeq 0 \end{aligned}$$

This is the case in which the strong duality holds, but the primal is not attained.

Third example :

$$\begin{aligned} \min \quad & a x_{22} \\ s.t. \quad & \begin{bmatrix} 0 & x_{12} & 1 - x_{22} \\ x_{12} & x_{22} & x_{23} \\ 1 - x_{22} & x_{23} & x_{33} \end{bmatrix} \succeq 0, \ \text{for } a > 0 \end{aligned}$$

This example is the case in which the primal and dual optimals are attained, but there is a non-zero duality gap. Thus, strong duality doesn't hold.