AM 221: Advanced Optimization	Spring 2016
Prof. Yaron Singer	*Lecture 14 — March 21st*

1 Overview

In the previous lectures we presented the concept of duality for convex programs. In this lecture we discuss duality gaps, complementary slackness, and the Karush-Kuhn-Tucker conditions that provide us necessary and sufficient conditions for strong duality of convex programs.

2 Recap

In this lecture we will continue using the same concepts as we did in lecture 12.

- We will consider the **primal constrained optimization problem** (*primal* for short):

$$\begin{aligned} &\min\ f(\mathbf{x}) \\ &\text{s.t. } g_i(\mathbf{x}) \leq 0 && \forall i \in [m] \\ &\quad\ \ h_j(\mathbf{x}) = 0 && \forall j \in [p] \end{aligned}$$

- The **Lagrangian** associated with the primal optimization problem is:

$$L(\mathbf{x}, \lambda, \nu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x})$$

  The variables $\lambda_1, \ldots, \lambda_m$ and $\nu_1, \ldots, \nu_p$ are called the **Lagrangian multipliers**.

- The **Lagrangian dual function** is:

$$F(\lambda, \nu) = \inf_{\mathbf{x}} L(\mathbf{x}, \lambda, \nu)$$

- The **Dual optimization problem** (*dual* for short) is:

$$\begin{aligned} &\max\ F(\lambda, \nu) \\ &\text{s.t. } \lambda \geq 0 \end{aligned}$$

- For every optimization problem **weak duality** holds, i.e. if $\alpha^\star$ is the value of the optimal solution for the primal problem, and $\beta^\star$ be the value of the optimal solution for the dual problem, then: $\beta^\star \leq \alpha^\star$.

- When $\alpha^\star = \beta^\star$ we say that **strong duality** holds. Strong duality does not always hold, but there are some conditions that do guarantee this property. In lecture 12 we saw that Slater's conditions guarantee strong duality.

3 Duality Gap

In some cases, computing the optimal solution for the dual problem is in fact easier than computing the optimal solution to the primal problem. Using $\alpha^\star$ to denote the value of the optimal solution to the primal problem and $\beta^\star$ to denote the value of the optimal solution to dual problem, we know from weak duality that $\alpha^\star \geq \beta^\star$. Thus, any feasible solution to the dual problem is a lower bound on the optimal solution. Furthermore, a point that is dual feasible enables us to bound how far away a solution is from optimal. If $\mathbf{x}$ is primal feasible and (λ, ν) is dual feasible then:

$$f(\mathbf{x}) - \alpha^\star \leq f(\mathbf{x}) - F(\lambda, \nu)$$

And thus if $f(\mathbf{x}) - F(\lambda, \nu) \leq \epsilon$, this implies that the solution $\mathbf{x}$ is at most ϵ away from optimal. We refer to this gap as the *duality gap.*

Definition. *For an optimization problem with objective $f : \mathbb{R}^n \to \mathbb{R}$ and dual $F : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, let $\mathbf{x}$ be primal feasible and (λ, ν) be dual feasible. The* ***duality gap*** *is defined as:*

$$f(\mathbf{x}) - F(\lambda, \nu)$$

If the duality gap is zero then we have strong duality. The useful property here is that if we have an algorithm that produces a series of primal and dual points, $\{\mathbf{x}^{(i)}, (\lambda^{(i)}, \nu^{(i)})\}_{i=1}^{t}$ and we know that for some $k \in [t]$ the duality gap is smaller than some $\epsilon > 0$, we can use $\mathbf{x}^{(k)}$ as our solution and be guaranteed that we're at most ϵ away from the optimal solution.

4 Complementary Slackness

The following claim will be useful for proving necessary and sufficient conditions for strong duality in the next section.

Claim 1. *Let $\mathbf{x}^\star \in \mathbb{R}^n$ be primal optimal and $(\lambda^\star, \nu^\star) \in \mathbb{R}^m \times \mathbb{R}^p$ be dual optimal, and suppose that strong duality holds. Then:*

- $\mathbf{x}^\star \in \operatorname{argmin}_{\mathbf{x}} L(\mathbf{x}, \lambda^\star, \nu^\star)$;
- $\lambda_i^\star g_i(\mathbf{x}^\star) = 0, \ \forall i \in [m]$.

Proof.

$$\begin{aligned}
f(\mathbf{x}^\star) &= F(\lambda^\star, \nu^\star) && (1)\\
&= \inf_{\mathbf{x}} \left(f(\mathbf{x}) + \sum_{i \in [m]} \lambda_i^\star g_i(\mathbf{x}) + \sum_{i \in [p]} \nu_i^\star h_i(\mathbf{x}) \right) && (2)\\
&\leq f(\mathbf{x}^\star) + \sum_{i \in [m]} \lambda_i^\star g_i(\mathbf{x}^\star) + \sum_{i \in [p]} \nu_i^\star h_i(\mathbf{x}^\star) && (3)\\
&\leq f(\mathbf{x}^\star) && (4)
\end{aligned}$$

Equality (1) holds due to strong duality, and (3) is simply by definition of the Lagrangian dual function. By minimality of the infimum, we have inequality (3), and inequality (4) is due to the fact that $\mathbf{x}^\star$ is primal feasible, thus $h_i(\mathbf{x}^\star) = 0$ for all $i \in [p]$, and since $\lambda^\star$ is dual feasible we know that $\lambda_i^\star \geq 0$ for all $i \in [m]$.

Since the left-hand side and the right hand side of the chain of inequalities are equal, we have that the inequalities (3) and (4) are actually an equalities. This proves the two points of our claim. □

The second point in the above claim is known as *complementary slackness*. For any optimal $\mathbf{x}^\star$ and dual optimal $(\lambda^\star, \nu^\star)$ when strong duality holds complementary slackness is the condition which states that:

- $\lambda_i^\star > 0 \implies g_i(\mathbf{x}^\star) = 0$;
- $g_i(\mathbf{x}^\star) < 0 \implies \lambda_i^\star = 0$.

5 KKT Conditions for Strong Duality

We now state the Karush-Kuhn-Tucker (KKT) conditions. These are conditions on the properties of primal and dual feasible points that are on the one hand necessary whenever strong duality holds, and other other hand these conditions guarantee strong duality for convex optimization problems.

Definition. *Given a primal optimization problem, we say that the points* $\mathbf{x}^\star \in \mathbb{R}^n$ *and* $(\lambda^\star, \nu^\star) \in \mathbb{R}^m \times \mathbb{R}^p$ *respect the* ***Karush-Kuhn-Tucker (KKT) conditions*** *if:*

$$g_i(\mathbf{x}^\star) \leq 0, \ \forall i \in [m] \tag{5}$$

$$h_i(\mathbf{x}^\star) = 0, \ \forall i \in [p] \tag{6}$$

$$\lambda_i^\star \geq 0, \ \forall i \in [m] \tag{7}$$

$$\lambda_i^\star g_i(\mathbf{x}^\star) = 0, \ i \in [m] \tag{8}$$

$$\nabla f(\mathbf{x}^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla g_i(\mathbf{x}^\star) + \sum_{i=1}^{p} \nu^\star \nabla h_i(\mathbf{x}^\star) = 0 \tag{9}$$

Theorem 2. *For any optimization problem, if strong duality holds then any primal optimal solution* $\mathbf{x}^\star \in \mathbb{R}^n$ *and dual optimal solution* $(\lambda^\star, \nu^\star) \in \mathbb{R}^m \times \mathbb{R}^p$ *respect the KKT conditions. Conversely, if* f *and* g_i *are convex for all* $i \in [m]$ *and* h_i *are* affine *for all* $i \in [p]$ *then the KKT conditions are sufficient for strong duality.*

Proof. We will show that the KKT conditions are both necessary and sufficient.

Necessary conditions: Assume that strong duality holds and that $\mathbf{x}^\star, (\lambda^\star, \nu^\star)$ are primal and dual optimal, respectively. Since the $\mathbf{x}^\star$ is primal feasible it must be that conditions (5) and (6) hold. Since $(\lambda^\star, \nu^\star)$ is dual feasible it must be that Condition (7) holds. By Claim 1 we know that Condition (8) must hold. By Claim 1 we know that $\mathbf{x}^\star \in \operatorname{argmin}_{\mathbf{x}} L(\mathbf{x}, \lambda^\star, \nu^\star)$. The gradient must therefore vanish at $\mathbf{x}^\star$ and thus we get Condition 9.

Sufficient conditions: Conditions (5) and (6) indicate that problem is primal feasible. Condition (7) together with the fact that f and g_i are convex, for all $i \in [m]$ and h_i are affine for all $i \in [p]$, implies that

$$L(\mathbf{x}, \lambda^\star, \nu^\star) = f(\mathbf{x}) + \sum_{i\in[m]} \lambda^\star g_i(\mathbf{x}) + \sum_{i\in[p]} \nu^\star h_i(\mathbf{x})$$

is a convex function. Since the function is convex, the fact that by Condition (9) the gradient vanishes at $\mathbf{x}^\star$ implies that $\mathbf{x}^\star$ is a global minimum. Thus:

$$\begin{aligned} F(\lambda^\star, \nu^\star) &= L(\mathbf{x}^\star, \lambda^\star, \nu^\star) \\ &= f(\mathbf{x}^\star) + \sum_{i\in[m]} \lambda_i^\star g_i(\mathbf{x}) + \sum_{i\in[p]} \nu_i^\star h_i(\mathbf{x}) \\ &= f(\mathbf{x}^\star) \end{aligned}$$

where the last line is due to the fact that both $h_i(\mathbf{x}^\star) = 0$ for all $i \in [p]$ and $\lambda_i^\star g_i(\mathbf{x}) = 0$ for all $i \in [m]$. Since by weak duality we know that for any λ, ν we have that $F(\lambda, \nu) \leq \inf_{\mathbf{x}} f(\mathbf{x})$, the above equality indicates that we have strong duality. $\square$

6 Discussion and Further Reading

This lecture is based on Chapter 5 from [1]. For more examples, applications, and interpretations of duality see Chapter 5 in [1].

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, New York, NY, USA, 2004.