**CPSC 665 : An Algorithmist's Toolkit** Lecture 10 : 11 Feb 2015

# Convex Programs & Lagrange Duality

*Lecturer: Sushant Sachdeva* *Scribe: Benjamin Peterson*

## 1. The General Framework and Some Examples

A general problem is minimizing a convex function $f$ over a convex domain $K \subset \mathbb{R}^n$.

We can define a convex domain $K \subset \mathbb{R}^n$ by giving a set of convex functions $\{f_i\} : \mathbb{R}^n \to \mathbb{R}$ and setting

$$K = \{x \in \mathbb{R} | f_i(x) \leq 0 \text{ for all } f_i\}.$$

That $K$ is convex follows from the fact that $\{x | f_i(x) \leq 0\}$ is convex for each $f_i$ and that the intersection of convex sets is convex. We can see that linear programming and semi-definite programming are special cases of convex optimization.

Another example of convex optimization is linear regression (aka least-squares regression). We are given data points $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}$ and think of the $b_i$ as a function $f$ of the $a_i$. Specifically, we suppose $f$ is linear and want to find the "best" approximation for the data. That is, we seek a $x \in \mathbb{R}^n$ such that $a_i \cdot x \approx b_i$. The amount of error is given by $\text{err}(i) = |a_i \cdot x - b_i|$. We try to minimize the $L^2$ norm of the error function. Explicitly, we have the minimization problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2,$$

where $A \in \mathcal{M}_n(\mathbb{R})$ is a matrix with the $a_i$ as rows. This is a convex optimization problem because the $L^2$ norm is a convex function (by the triangle inequality).

## 2. The Lagrange Dual to a Convex Problem

Consider the problem of minimizing a function $f_0(x)$ such that for $i \in [m]$, $f_i(x) \leq 0$ and for $j \in [k]$, $a_j \cdot x = b_j$. Call this program $P$ for primal.

For $i \in [m]$, $\lambda_i \geq 0$, $j \in [k]$, and $\mu_j \in \mathbb{R}$, we must have

$$f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{k} \mu_j (a_j \cdot x - b_j) \leq f_0(x)$$

for feasible $x \in \mathbb{R}^n$. The left hand side of that inequality is called the Lagrangian and denoted $L(x, \lambda, \mu)$.

To obtain a lower bound for $P$, we can compute the program

$$g(\lambda, \mu) = \min_{x \text{ feasible for } P} L(x, \lambda, u).$$

The function $g$ is called the Lagrange dual of $P$.

L is linear in $\lambda$ and $\mu$. Thus, $g$ is concave in $\lambda$ and $\mu$. The best lower bound is

$$\max_{\substack{\lambda \geq 0 \\ \mu \in \mathbb{R}}} g(\lambda, \mu).$$

This program is called the Lagrange dual program and denoted $D$. We'll say $(\lambda, \mu)$ is feasible if $g(\lambda, \mu) > -\infty$. Let

$$K' = \{(\lambda, \mu) | \lambda \geq 0, g(\lambda, \mu) > -\infty\}$$

denote the set of feasible $(\lambda, \mu)$ pairs. Note that $K'$ is convex.

We have the following duality theorem:

**Theorem 2.1** (Weak Duality). *Let be $x$ be feasible for $P$, $x^*$ optimal, $\lambda, \mu$ feasible for $D$, $\lambda^*, \mu^*$ optimal. Then*

$$g(\lambda, \mu) \le g(\lambda^*, \mu^*) \le f_0(x^*) \le f_0(x).$$

We can see how this applies to linear programs. If the linear program is minimizing $c \cdot x$ with constraints $-a_i \cdot x + b_i \le 0$, then

$$\begin{aligned} L(x, y) &= c \cdot x + \sum_i y_i(-a_i \cdot x + b_i) \\ &= c \cdot x + y \cdot (-Ax + b) \\ &= y \cdot b + (c - y^T A) \cdot X \end{aligned}$$

and

$$g(y) = \min_x L(x, y).$$

This gives us the familiar dual program $\max b \cdot y$ such that $y \ge 0$ and $A^T y = c$.

We have an equivalence of sets

$$\{y | y \ge 0, A^T y = c\} = \{y | \min_x L(x, y) > -\infty\}.$$

We can rewrite the problem as $\max_{y \ge 0} b \cdot y + g'(y)$ where $g'(y) = \min_{x \in \mathbb{R}^n} (c^T - y^t A)x$. The function $g'(y)$ is an indicator function for the feasibility region of the problem and allows us to remove the $A^T y = c$ constraint. We thus may rewrite the problem in the simpler form

$$\max_{y \ge 0} \min_{x \in \mathbb{R}^n} L(x, y).$$

Weak duality can now be restated as

$$\max_{y \ge 0} \min_{x \in \mathbb{R}^n} L(x, y) \le \min_{x \in \mathbb{R}^n} \max_{y \ge 0} L(x, y).$$

(It's homework to show that all this in fact holds for all $f(x, y) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.)

Now, we come to strong duality, where this machinery becomes much more useful.

**Definition 2.2.** *We say a program satisfies strong duality when the primal and the dual program are feasible and $p^* = d^*$.*

**Theorem 2.3** (Slater's Condition). *If a convex program is strictly feasible, strong duality holds.*

Strictly feasible means there is a feasible point that strictly satisfies the constraints i.e. the feasible region has a non-empty interior.

In fact, we can do slightly better than Slater's condition; it suffices to have an $x$ that strictly satisfies the non-affine constraints.

We also have a generalized version of complementary slackness.

**Lemma 2.4.** *If $\lambda^*$, $\mu^*$, and $x^*$ are feasible, optimal, and strong duality holds for $P$ and $D$ then $\lambda_i^* f_i(x_i^*) = 0$.*

*Proof.* Using strong duality, we have

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \mu^*) \\ &= \min_x f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \mu_j^*(a_j \cdot x - b_j) \\ &\le f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_j \mu_j^*(a_j \cdot x^* - b_j) \\ &= f_0(x^*). \end{aligned}$$

It is immediate that $\lambda_i^* f_i(x^*) = 0$. ∎