MTH 995-001: Intro to CS and Big Data	Spring 2015
Lecture X — Mar 24, 2015	
Inst. Aditya Viswanathan	*Scribe: Sami Merhi*

1 Introduction

Consider an optimization problem in **standard** form:

$$\min_{x\in\mathbb{R}^n} f_0(x) \quad \text{subject to} \quad \begin{cases} f_i(x) \le 0, & i = 1,\dots,m \\ h_i(x) = 0, & i = 1,\dots,p. \end{cases} \tag{1}$$

We define the domain D of problem (1) as the intersection of the domains of all constraints. That is,

$$D = \left(\bigcap_{i=1}^{m} \mathrm{dom}\ f_i\right) \cap \left(\bigcap_{i=1}^{p} \mathrm{dom}\ h_i\right),$$

We assume that D is non-empty, and denote by p^* the **optimal** value of problem (1).

2 Duality

Definition 1. *The **Lagrangian** associated with* (1) *is the function* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ *defined by*

$$L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

with dom $L = D \times \mathbb{R}^m \times \mathbb{R}^p$*. Here*

$$\lambda \in \mathbb{R}^m,\ \nu \in \mathbb{R}^p$$

*are called **dual variables** (or Lagrange multiplier vectors).*

Definition 2. *(Lagrange dual function) The **dual function*** $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ *is the minimum value of the Lagrangian over all* x*; that is,*

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

*Note that the dual function is always **concave**, being the pointwise infimum of affine functions.*

Lemma 1. *The dual function provides a lower bound on the optimal value* p^* *for the optimization problem* (1)*; that is,*

$$g(\lambda,\nu) \le p^* \tag{2}$$

for all $\lambda \succeq 0$ and for all ν.

Remark. *By $\lambda \succeq 0$ we mean $\lambda_i \geq 0$ for all $i = 1, \ldots, m$.*

For a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \text{dom } g$, we say (λ, ν) are **dual feasible**.

Example 1. ***A simple linear program***

Recall the optimization problem

$$\min_{x \in \mathbb{R}^n} c^T x \quad \textit{subject to} \quad Ax = b,\ x \succeq 0.$$

*The **Lagrangian** associated with this problem is*

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x - \sum_{i=1}^{\infty} \lambda_i x_i + \nu^T (Ax - b) \\ &= -b^T \nu + \left(c + A^T \nu - \lambda\right)^T x. \end{aligned}$$

Here we have set the inequality constraints as $f_i(x) = -x_i$, for $i = 1, \ldots, m$.

*The **dual function** associated with this problem is*

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu, & \textit{if } A^T \nu - \lambda + c = 0, \\ -\infty, & \textit{otherwise.} \end{cases}$$

Definition 3. *(Lagrange dual problem) The **dual problem** associated with (1) is the optimization problem*

$$\textit{maximize } g(\lambda, \nu) \quad \textit{subject to} \quad \lambda \succeq 0. \tag{3}$$

*We say (λ^*, ν^*) are **dual optimal** is they are optimal for the above problem.*

2.1 Duality Gap

Definition 4. *Let d^* denote the optimal value of the dual problem (3). We call $p^* - d^*$ the **optimal duality gap**.*

In general we have $d^* \leq p^*$; this is called **weak duality**; when $d^* = p^*$, we have **strong duality**.

Slater's Condition

Slater's condition is a sufficient condition for strong duality to hold for a convex optimization problem: if the primal problem (1) is convex, and if x is in the **relative interior** of D ($x \in \text{relint} D$), that is

$$\begin{aligned} f_i(x) &< 0 \text{ for } i = 1, \ldots, m, \\ h_i(x) &= 0 \text{ for } i = 1, \ldots, p, \end{aligned}$$

then $p^* = d^*$. In this case we say "x is **strictly feasible**."

From (2), we see that (λ, ν) provides a **proof** or **certificate** that $p^* \geq g(\lambda, \nu)$.

Suppose now that $p^* = d^*$. Then if x^* minimizes $f_0(x)$, we have

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

by the equality and inequality constraints. This means that x^* minimizes $L(x, \lambda^*, \nu^*)$ over x, and that

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

Hence $\lambda_i^* f_i(x^*) = 0$ for $i = 1, \ldots, m$. This is called **complementary slackness**. In more detail,

$$\lambda_i^* > 0 \implies f_i(x^*) = 0 \quad \text{or} \quad f_i(x^*) < 0 \implies \lambda_i^* = 0.$$

2.2 KKT Conditions

Let $f_0, \ldots, f_m$ and $h_1, \ldots, h_m$ be differentiable functions. Let x^* and (λ^*, ν^*) be the primal-dual optimal points. We know x^* minimizes $L(x, \lambda^*, \nu^*)$ over x. Thus

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0.$$

This condition, along with the conditions

$$\begin{cases} f_i(x^*) \leq 0, & i = 1, \ldots, m \\ h_i(x^*) = 0, & i = 1, \ldots, p \\ \lambda_i^* \geq 0, & i = 1, \ldots, m \\ \lambda_i^* f_i(x^*) = 0, & i = 1, \ldots, m \end{cases}$$

are called the **KKT** (Karush-Kahn-Tucker) **Conditions**.

For **convex** optimization problems with **differentiable** objective and constraints satisfying **Slater's** condition, the **KKT** conditions are **necessary** and **sufficient** for optimality.

3 Extension to Generalized Inequalities

We now consider the optimization problem

$$\text{minimize } f_0(x) \quad \text{subject to} \quad \begin{cases} f_i(x) \preceq_{K_i} 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{cases}$$

where $K_i \subset \mathbb{R}^{K_i}$ are proper **cones**. Here $x \preceq_K y \iff y - x \in K$.

Definition 5. *A* ***cone*** *is a set invariant under multiplication by nonnegative scalars. That is, if* $x \in K$ *and* $\lambda \geq 0$ *then* $\lambda x \in K$.

Example 2. *Here are some examples of cones:*

1. ***Quadratic Cone:***
$$K_q = \{z \in \mathbb{R}^m | \, \|(z_2, \ldots, z_m)\|_2 \leq z_1\}.$$

2. ***Positive Orthant:***
$$K_+ = \{z \in \mathbb{R}^m | \, z_1 \geq 0, \, z_2 \geq 0, \ldots, \, z_m \geq 0\}.$$

3. ***Positive-semidefinite cone:***
$$K_{S_+} = \left\{X \in \mathbb{S}^{n \times n} | \, X \succeq 0\right\}.$$

Definition 6. *The* ***dual of a cone*** K *in a linear space* X *with topological dual space* X^* *is the set*
$$Dual(K) = \{\, z \in X^* | \, \langle y, x \rangle \geq 0 \, \forall x \in K\},$$
where $\langle y, x \rangle$ *is the duality pairing between* X *and* X^*.

In the case where $K \subset \mathbb{R}^n$*, the dual of* K *is*
$$Dual(K) = \left\{\, y \in \mathbb{R}^n | \, y^T x \geq 0 \, \forall x \in K\right\}.$$

Lemma 2. *The positive orthant cone* K_+ *in* $\mathbb{R}^m$ *is equal to its dual cone.*

Lemma 3. *The positive-semidefinite cone* K_{S_+} *in* $\mathbb{S}^{n \times n}$ *is equal to its dual cone.*

4 Homework

Prove Lemmata 1,2, and 3.