Chapter 14

Convex Optimization

This chapter briefly summarizes the essential elements of convex optimization. There are many good books on this topic that go into much greater depth. See the references at the end of the chapter for additional reading. Before reading this chapter, it may be helpful to revise the material on convex sets and functions in Chapter 7.

14.1 Convex Programs

A *convex program* is an optimization problem of the form

$$\begin{aligned} \min_{w\in\mathbb{R}^n} \quad & f(w) \\ \text{s.t.} \quad & f_i(w) \leq 0,\ i \in [1\!:\!k] \\ & Aw - b = \mathbf{0}, \end{aligned} \tag{14.1}$$

where $f, f_i \colon \mathbb{R}^n \to \mathbb{R}$, $i \in [1\!:\!k]$, are convex functions, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The function f is called the *objective function*, the functions f_i are called the *constraint inequalities*, and the rows of the affine vector function $Aw - b$ are called the *affine equality constraints*.

A point $w \in \mathbb{R}^n$ satisfying all of the constraints in (14.1) is said to be a *feasible point*, and the set of all feasible points is called the *feasible set*. The feasible set is the intersection of the 0-sublevel sets $L_0^{(i)} = \{w\colon f_i(w) \leq 0\}$, $i \in [1 : k]$, and the set $\mathcal{S} = \{w\colon Aw - b = \mathbf{0}\}$. The set $\mathcal{S}$ is either empty ($b \notin \mathcal{R}(A)$), or is an affine manifold of the form $w_p + \mathcal{N}(A)$ where $Aw_p = b$. Hence $\mathcal{S}$ is closed and convex. Since each of the functions f_i is convex, each sublevel set $L_0^{(i)}$ is closed and convex. So the feasible set is an intersection of closed convex sets and is hence closed and convex. Thus problem (14.1) seeks to minimize a convex function f over a closed, convex set. To be an interesting problem, we need the feasible set to be nonempty. In this case we say that the problem is *feasible*. Otherwise we say that it is *infeasible*.

14.1.1 Linear Programs

Let $F \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}^m$ and $h \in \mathbb{R}^n$. A *linear program* minimizes a linear function subject to linear inequality constraints:

$$\begin{aligned} \min_{w\in\mathbb{R}^n} \quad & h^T w \\ \text{s.t.} \quad & Fw \leq g. \end{aligned} \tag{14.2}$$

The objective function $f(w) = h^T w$ is linear and hence convex. Let $F_{i,:}$ denote the i-th row of F and g_i denote the i-th entry of g. Then there are m constraint inequalities $f_i(w) = F_{i,:}w - g_i \leq 0$. These are affine and hence convex. Thus a linear program is a convex program. To ensure feasibility, we need the existence of a $w \in \mathbb{R}^n$ such that $Fw \leq g$.

14.1.2 Quadratic Programs

Let $P \in \mathbb{R}^{n\times n}$ be a symmetric positive semidefinite matrix, $q \in \mathbb{R}^n$, $r \in \mathbb{R}$, $F \in \mathbb{R}^{k\times n}$, and $g \in \mathbb{R}^k$. A *quadratic program* minimizes a quadratic objective subject to affine inequality constraints:

$$\begin{aligned} \min_{w\in\mathbb{R}^n} \quad & 1/2\, w^T P w + q^T w + r \\ \text{s.t.} \quad & Fw \leq g. \end{aligned} \tag{14.3}$$

Since P is symmetric positive semidefinite, the objective function $f(w) = w^T P w + q^T w + r$ is convex. There are m affine constraint functions $f_i(w) = F_{i,:}w - g_i \leq 0$, $i \in [1\!:\!m]$. To ensure feasibility, we need at least one point $w \in \mathbb{R}^n$ to satisfy $Fw \leq g$.

14.2 The Lagrangian and the Dual Problem

Consider the general convex program (14.1). For clarity, we will refer to (14.1) as the *primal problem* and its objective and constraints as the *primal objective* and *primal constraints*, respectively.

Bring in *dual variables* $\lambda \in \mathbb{R}^k$, with $\lambda \geq \mathbf{0}$, and $\mu \in \mathbb{R}^m$, and form the *Lagrangian*:

$$L(w, \lambda, \mu) \triangleq f(w) + \sum_{i=1}^{k} \lambda_i f_i(w) + \mu^T(Aw - b).$$

By construction, for all $\lambda \geq \mathbf{0}$, $\mu \in \mathbb{R}^m$ and feasible w:

$$L(w, \lambda, \mu) \leq f(w) \qquad \text{and} \qquad \max_{\lambda \geq 0, \mu} L(w, \lambda, \mu) = f(w).$$

Notice that $L(w, \lambda, \mu)$ is a convex function of w. Moreover, if the objective $f(w)$ and constraint functions $f_i(w)$, $i \in [1\!:\!k]$, are differentiable w.r.t. w, then so is $L(w, \lambda, \mu)$.

We now set out to minimize the unconstrained convex function L with respect to w, without requiring that w is feasible. This gives rise to the *dual objective function* $g(\lambda, \mu)$ defined by

$$g(\lambda, \mu) = \min_{w\in\mathbb{R}^n} L(w, \lambda, \mu).$$

The domain of g is the set of all $(\lambda, \mu) \in \mathbb{R}^k \times \mathbb{R}^m$ satisfying $\lambda \geq \mathbf{0}$ and $g(\lambda, \mu) > -\infty$. Such points are said to be *dual feasible*. By construction, for all dual feasible (λ, μ) and feasible w,

$$g(\lambda, \mu) = \min_{v\in\mathbb{R}^n} L(v, \lambda, \mu) \leq L(w, \lambda, \mu) \leq \max_{\lambda \geq 0, \mu} L(w, \lambda, \mu) = f(w). \tag{14.4}$$

So for all dual feasible (λ, μ) and all feasible w, the dual objective lower bounds the primal objective:

$$g(\lambda, \mu) \leq f(w). \tag{14.5}$$

© Peter J. Ramadge, 2015, 2016, 2017, 2018. Please do not distribute without permission.

We now use $g(\lambda, \mu)$ to define the *dual problem*:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^k, \mu \in \mathbb{R}^m} \quad & g(\lambda, \mu) \\ \text{s.t.} \quad & (\lambda, \mu) \text{ is dual feasible.} \end{aligned} \tag{14.6}$$

The requirement of dual feasibility sometimes imposes implicit constraints on λ and μ that go beyond the obvious requirement that $\lambda \geq \mathbf{0}$.

Example 14.2.1 (The Dual of a Quadratic Program). Consider the quadratic program introduced in §14.1.2 with the stronger assumption that P is symmetric positive definite. Bring in a dual variable $\lambda \in \mathbb{R}^m$ with $\lambda \geq \mathbf{0}$ and form the Lagrangian

$$L(w, \lambda) = {}^1\!/_2\, w^T P w + q^T w + r + \lambda^T (Fw - c). \tag{14.7}$$

The corresponding dual objective function is $g(\lambda) = \min_{w \in \mathbb{R}^n} L(w, \lambda)$. Setting the derivative of L with respect to w equal to zero yields

$$w^T P h + q^T h + \lambda^T F h = (w^T P + q^T + \lambda^T F) h = 0.$$

Hence the value of w minimizing $L(w, \lambda)$ is

$$w = -P^{-1} q - P^{-1} F^T \lambda. \tag{14.8}$$

Substituting this expression into $L(w, \lambda)$ and simplifying gives the dual objective

$$g(\lambda) = -{}^1\!/_2\, \lambda^T (F P^{-1} F^T) \lambda - (q^T P^{-1} F^T + c^T) \lambda + (r - {}^1\!/_2\, q^T P^{-1} q).$$

Thus the dual problem is

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m} \quad & - {}^1\!/_2\, \lambda^T (F P^{-1} F^T) \lambda - (q^T P^{-1} F^T + c^T) \lambda + (r - {}^1\!/_2\, q^T P^{-1} q) \\ \text{s.t.} \quad & \lambda \geq \mathbf{0}. \end{aligned} \tag{14.9}$$

Using some simple algebra, one can write the dual problem in the standard form (14.3) for a quadratic program. In this sense, it is also a quadratic program.

14.3 Weak Duality, Strong Duality and Slater's Condition

Let $w^\star$ denote a solution of the primal problem and $(\lambda^\star, \mu^\star)$ denote a solution of the dual problem. The point $w^\star$ must be feasible, and $(\lambda^\star, \mu^\star)$ must be dual feasible. Hence by (14.5) the following *weak duality* condition always holds:

$$g(\lambda^\star, \mu^\star) \leq f(w^\star). \tag{14.10}$$

However, in general, there need not be equality in this expression. If for a particular convex program $g(\lambda^\star, \mu^\star) = f(w^\star)$, then we say that *strong duality* holds.

14.3.1 Slater's Condition

There are a number of sufficient conditions, known as *constraint qualifications*, each ensuring that strong duality holds. One of the simplest of these is known as *Slater's condition*. We describe this below. The important point is that if Slater's condition is satisfied, then strong duality holds.

For the convex program (14.1), Slater's condition requires that there is feasible point w satisfying $f_i(w) < 0$, $i \in [1:k]$. In the special case when the inequality constraints in (14.1) are affine constraints of the form $f_i(w) = a_i^T w - b_i \leq 0$, Slater's condition is even simpler: it only requires that the primal problem (14.1) has a feasible point.

© Peter J. Ramadge, 2015, 2016, 2017, 2018. Please do not distribute without permission.

14.4 Complementary Slackness

Assume that strong duality holds. Then $g(\lambda^\star, \mu^\star) = f(w^\star)$ and by (14.4) it follows that $g(\lambda^\star, \mu^\star) = L(w^\star, \lambda^\star, \mu^\star) = f(w^\star)$. Hence

$$\begin{aligned} f(w^\star) &= L(w^\star, \lambda^\star, \mu^\star) \\ &= f(w^\star) + \sum_{i=1}^{k} \lambda_i^\star f_i(w^\star) + \mu^{\star T}(Aw^\star - b) \\ &= f(w^\star) + \sum_{i=1}^{k} \lambda_i^\star f_i(w^\star). \end{aligned}$$

Thus $\sum_{i=1}^k \lambda_i^\star f_i(w^\star) = 0$. Since the terms in this sum are nonpositive, each term must be zero:

$$\lambda_i^\star f_i(w^\star) = 0, \quad i \in [1\!:\!k]. \tag{14.11}$$

The equations (14.11) are called *complementary slackness* conditions. If $f_i(w^\star) = 0$, we say that the constraint is *active*, otherwise it is *inactive* or *slack*. The dual variable $\lambda_i^\star$ must satisfy $\lambda_i^\star \geq 0$. Hence when $\lambda_i^\star = 0$, its constraint is active, and when $\lambda_i^\star > 0$, it is slack. So (14.11) says that at an optimal solution: if a primal constraint is slack ($f_i(w^\star) < 0$), then the corresponding dual variable constraint must be active ($\lambda_i^\star = 0$), and if the dual variable constraint is slack ($\lambda_i^\star > 0$), then the corresponding primal constraint must be active ($f_i(w^\star) = 0$). Hence the term complementary slackness.

14.5 The KKT Conditions

Assume that strong duality holds and that the functions $f(w)$ and $f_i(w)$, $i \in [1 : k]$, are continuously differentiable. For feasible w, the inequality

$$g(\lambda^\star, \mu^\star) \leq L(w, \lambda^\star, \mu^\star) \leq f(w),$$

and strong duality imply that $w^\star$ minimizes $L(w, \lambda^\star, \mu^\star)$ over feasible w. Since $L(w, \lambda^\star, \mu^\star)$ is differentiable w.r.t. w, it follows that

$$\nabla f(w^\star) + \sum_{i=1}^{k} \lambda_i^\star \nabla f_i(w^\star) + A^T \mu^\star = 0.$$

Hence the following conditions, known as the *KKT conditions*[1], are necessarily satisfied at $w^\star$, $\lambda^\star$, $\mu^\star$:

$$\begin{aligned} \nabla f(w^\star) + \sum_{i=1}^{k} \lambda_i^\star \nabla f_i(w^\star) + A^T \mu^\star = 0 &\qquad \nabla_w L(w^\star, \lambda^\star, \mu^\star) = 0 \\ \text{for } i \in [1\!:\!k], \ \ f_i(w^\star) \leq 0 &\qquad \text{primal constraint} \\ Aw^\star - b = \mathbf{0} &\qquad \text{primal constraint} \\ \lambda^\star \geq \mathbf{0} &\qquad \text{dual constraints (There can be more)} \\ \text{for } i \in [1\!:\!k], \ \ \lambda_i^\star f_i(w^\star) = 0 &\qquad \text{complementary slackness} \end{aligned}$$

For general optimization problems that's as much as we can say. However, for convex programs satisfying the above assumptions one can say more.

[1]Named for those who first published the result: William Karush (1939), and Harold W. Kuhn and Albert W. Tucker (1951).

© Peter J. Ramadge, 2015, 2016, 2017, 2018. Please do not distribute without permission.