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

Lecture 3: January 21

Lecturer: Nicole Rafidi (TA) Instructor: Ryan Tibshirani

Scribes: Allie Del Giorno, Chen-Hsuan Lin, Shashank Srivastava

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

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

3.1 Optimization terminology

3.1.1 Terminology

A **convex optimization program**, or convex problem, can either be written as a convex minimization (convex f):

$$\begin{aligned} \underset{x \in D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Or equivalently as a concave maximization (concave f):

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Note that the domain D is implicit. We will describe the terminology in terms of the first formalization.

$f(x)$: **Objective function, criterion** the convex function we are minimizing over.

D : **Optimization domain** the domain (set of x's) in which our solution can exist:

$$D = \text{dom}(f) \cap \textstyle\bigcap_i \text{dom}(g_i)$$

$g_i(x)$: **Inequality constraints** convex functions; the solution must satisfy these.

f^*: **Optimal value** the minimum of $f(x)$ over all feasible points x

X_{opt}: **Set of solutions** All x's that solve to a convex problem. Property: it's a convex set.

Solution, optimal point the x at which the function achieves the optimal value; i.e. - x such that $f(x) = f^*$; also called the minimizer.

Feasible points any x in the domain of f for which x satisfies the constraints.
i.e. - $x : x \in D, g_i(x) \leq 0 \forall i, Ax = b$

ϵ-suboptimal a feasible x for which the function is within ϵ of the optimal value,
i.e. - $f(x) \leq f^* + \epsilon$

Active if $g_i(x) = 0$ for some x, then we say the constraint g_i is active at x.

Unique solution The solution is unique if X_{opt} contains only one element.

Locally optimal a feasible point x is locally optimal if it looks like a solution in a smaller domain;

i.e. - there is some $R > 0$ s.t. $f(x) \leq f(y)$ for all feasible y such that $\|x - y\|_2 \leq R$

Globally optimal x is globally optimal if it is the solution over the full optimization domain.

3.2 Properties and first-order optimality

3.2.1 Simple theorems

Theorem 3.1 *The set of solutions X_{opt} is a convex set.*

Theorem 3.2 *If criterion f is strictly convex, then its solution is unique, i.e. X_{opt} contains only one element.*

Proof. By definition, f is strictly convex implies that $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. Suppose that f is optimal at both x and y, i.e. $f(x) = f(y) = f^\star$. Then we have

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) = tf^\star + (1-t)f^\star = f^\star \tag{3.1}$$

which implies that there exists a smaller value of f than $f^\star$. Thus we have proven by contradiction that the solution of f is unique if f is strictly convex.

Theorem 3.3 *Local optima are global optima for convex optimization problems.*

For a convex problem, a feasible point x is locally optimal if there is some $R > 0$ such that

$$f(x) \leq f(y) \text{ for all feasible } y \text{ such that } \|x - y\|_2 \leq R \tag{3.2}$$

.

3.2.2 First-order optimality condition

For a convex problem

$$\min f(x) \text{ subject to } x \in C \tag{3.3}$$

and differentiable f, a feasible point x is optimal if and only if

$$\bigtriangledown f(x)^T (y - x) \geq 0 \text{ for all } y \in C \tag{3.4}$$

This is called the first-order condition for optimality.

In other words: all feasible directions from x are aligned with increasing gradient $\bigtriangledown f(x)$.

Important special case: if $C = \mathbb{R}^n$ (unconstrained optimization), then optimality condition reduces to familiar $\bigtriangledown f(x) = 0$.

Example 1: Quadratic minimization. Consider minimizing the quadratic function

$$f(x) = \frac{1}{2}x^T Qx + b^T x + c \tag{3.5}$$

where $Q \succeq 0$. The first-order condition says that solution satisfies

$$\bigtriangledown f(x) = Qx + b = 0 \tag{3.6}$$

Cases:

- If $Q \succ 0$, then there is a unique solution $x = Q^{-1}b$.
- If Q is singular and $b \notin \text{col}(Q)$, then there is no solution (i.e. $\min_x f(x) = -\infty$).
- If Q is singular and $b \in \text{col}(Q)$, then there are infinitely many solutions

  $$x = Q^+ b + z, z \in \text{null}(Q) \tag{3.7}$$

  where $Q^+ = (Q^T Q)^{-1} Q^T$ is the pseudoinverse of Q.

Example 2: Equality-constrained minimization. Consider the equality-constrained convex problem:

$$\min f(x) \text{ subject to } Ax = b \tag{3.8}$$

with f differentiable. Let's prove Lagrange multiplier optimality condition

$$\bigtriangledown f(x) + A^T v = 0 \text{ for some } v \tag{3.9}$$

Acoording to first-order optimality, solution x satisfies $Ax = b$ and

$$\bigtriangledown f(x)^T (y - x) \geq 0 \text{ for all } y \text{ such that } Ay = b \tag{3.10}$$

Let $y = 2x - y'$. We have

$$\bigtriangledown f(x)^T (x - y') \geq 0 \text{ for all } y' \text{ s.t. } Ay' = b \tag{3.11}$$

Since we have the inequalities

$$\begin{cases} \bigtriangledown f(x)^T (y - x) \geq 0 \\ \bigtriangledown f(x)^T (y - x) \leq 0 \end{cases}$$

we have $\bigtriangledown f(x)^T (y - x) = 0$. This is equivalent to

$$\bigtriangledown f(x)^T v = 0 \text{ for all } v = y - x \in \text{ null } (A) \tag{3.12}$$

Result follows since $\text{null}(A)^\perp = \text{row}(A)$.

Example 3: Projection onto a convex set. Consider projection onto convex set C:

$$\min \|a - x\|_2^2 \text{ subject to } x \in C \tag{3.13}$$

First-order optimality condition says that the solution x satisfies

$$\bigtriangledown f(x)^T (y - x) = (x - a)^T (y - x) \geq 0 \text{ for all y } \in C \tag{3.14}$$

Equivalently, this says that $(a - x) \in \mathcal{N}_C(x)$ where recall $\mathcal{N}_C(x)$ is the normal cone to C at x.

3.2.3 Partial optimization

Reminder: $g(x) = \min_{y \in C} f(x, y)$ is convex in x, provided that f is convex in (x, y) and C is a convex set. Therefore, we can always partially optimize a convex problem and retain convexity.

For example, if we decompose $x = (x_1, x_2) \in \mathbb{R}^{n_1+n_2}$, then

$$\min_{x_1,x_2} f(x_1, x_2) \text{ s.t. } g_1(x_1) \le 0, g_2(x_2) \le 0 \iff \min_{x_1} \tilde{f}(x_1) \text{ s.t. } g_1(x_1) \le 0 \tag{3.15}$$

where $\tilde{f}(x_1) = \min \{f(x_1, x_2) : g_2(x_2) \le 0\}$. The right problem is convex if the left problem is.

Example: Hinge form of SVMs. Recall the SVM problem

$$\min_{\beta,\beta_0,\xi} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n} \xi_i \text{ subject to } \xi_i \ge 0, y_i(x_i^T\beta + \beta_0) \ge 1 - \xi_i, i = 1, \dots, n \tag{3.16}$$

Rewrite the constraints as $\xi_i \ge \max\{0, 1 - y_i(x_i^T\beta + \beta_0)\}$. Indeed we can argue that we have equality at the solution.

Therefore plugging in for optimal ξ gives the hinge form of SVMS:

$$\min_{\beta,\beta_0} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{n}[1 - y_i(x_i^T\beta + \beta_0)]_+ \tag{3.17}$$

where $a_+ = \max\{0, a\}$ is called the hinge function.

3.3 Transformations and change of variables

Sometimes a problem can be transformed to an equivalent problem which might be easier to solve. This can be achieved by transforming the objective function, or re-writing the constraints differently.

For example, if $h : \mathbb{R} \to \mathbb{R}$ is a *monotone increasing* mapping, then the problem:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to} \quad & x \in C \end{aligned}$$

can be transformed to the following equivalent problem:

$$\begin{aligned} \min_{x} \quad & h(f(x)) \\ \text{subject to} \quad & x \in C \end{aligned}$$

Since h is strictly increasing, $h(f(x))$ must reach its minimum value at exactly the same value x^*, which minimizes $f(x)$. However, in certain cases, minimizing $h(f(x))$ might be easier than directly minimizing $f(x)$. e.g. , the log-likelihood function in several statistical problem is often easier to optimize than directly optimizing the likelihood function. In many cases, such monotone transformations can reveal the 'hidden convexity' of a problem, which might not be cursorily obvious.

Similarly, if $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is a bijective mapping, with its image covering the feasible convex set C , then the variables in the optimization can be changed to yield the following equivalent optimization:

$$\begin{aligned} &\min_{x} f(x) \text{ subject to } x \in C \\ \Leftrightarrow &\min_{y} f(\phi(y)) \text{ subject to } \phi(y) \in C \end{aligned}$$

3.3.1 Eliminating equality constraints

Let us assume a general convex problem with (affine) equality constraints:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, \quad \ldots, m. \\ & Ax = b \end{array}$$

Suppose we have $Ax_0 = b$. Then, any solution x of $Ax = b$ corresponds to a solution $x - x_0$ for $Ax = 0$. Suppose the columns of a matrix M span the null-space of A. Then any solution for $Ax = b$ can be given by some My , where the vector y gives the weights of the basis solutions. Hence, any solution for $Ax = b$ can be given as $x := My + x_0$, where the columns of M span the null space of A.

Thus, the equality constraints can be eliminated in the above optimization to yield the following equivalent formulation:

$$\begin{array}{ll} \min\limits_{x} & f(My + x_0) \\ \text{subject to} & g_i(My + x_0) \leq 0, i = 1, \quad \ldots, m. \end{array}$$

The reformulated optimization may not be any easier to solve in many cases.

3.3.2 Introducing slack variables

In many scenarios, slack variables can be added to inequality constraints to transform them to equalities. These transformations are only possible for affine inequality constraints. Introducing a slack variable replaces an affine inequality with an equality and a non-negativity constraint.

In general, a given problem:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{subject to} & g_i(x) \leq 0, i = 1, \quad \ldots, m. \\ & Ax = b \end{array}$$

can be transformed by introducing slack variables s_i:

$$\begin{array}{lll} \min\limits_{x} & f(x) & \\ \text{subject to} & s_i \geq 0, & i = 1, \ldots, m. \\ & g_i(x) + s_i = 0, & i = 1, \ldots, m. \\ & Ax = b & \end{array}$$

The reformulated problem is convex if the transformed inequalities g_i are affine.

3.3.3 Relaxing non-affine equalities

Consider any optimization:

$$\min_x f(x) \text{ subject to } x \in C$$

Any such problem can be relaxed by enlarging its constraint set. .i.e., choosing $C' \subseteq C$:

$$\min_x f(x) \text{ subject to } x \in C'$$

Since the feasible set is larger, the optimal value of the new problem is at least as low as that of the original problem.

The most important scenario where this is useful is to relax non-affine equality constraints. i.e., replacing constraints of the form $h_i(x) = 0$ for convex but non-affine functions $h_i(x)$ by inequality constraints $h_i(x) \leq 0$. Relaxing such constraints can make a non-convex problem convex. In certain cases, it can be shown that the relaxed problem's solution is identical to that of the original problem. This can be done, for example, by arguing that the equality constraint must be active at the optimal solution. e.g., the Maximum Utility problem.