Lecture 3: September 4

Lecturer: Lecturer: Ryan Tibshirani *Scribes: Scribes: oneopane, ruogulin,2, rschmuck*

3.1 Optimization terminology

The following defines a convex optimization problem/program:

$$\begin{aligned} \underset{x\in D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0,\ i = 1,\dots,m \\ & Ax = b \end{aligned} \tag{3.1}$$

where f and g_i, $i = 1,\dots,m$ are all convex, and the optimization domain is $D = \text{dom}(f) \cap \bigcap_{i=1}^{m} \text{dom}(g_i)$. Some related terminology:

- f is called *criterion* or *objective* function
- g_i is called *inequality constraint* function
- If $x \in D$, $g_i(x) \le 0$, $i = 1,\dots,m$, and $Ax = b$ then x is called a *feasible* point
- The minimum of $f(x)$ over all feasible points $x \in D$ is called the *optimal value*, written f^*
- If x is feasible and $f(x) = f^*$, then x os called *optimal*, *solution* or *minimizer*
- If x is feasible and $f(x) \le f^* + \epsilon$, then x is called *ϵ-suboptimal*
- If x is feasible and $g_i(x) = 0$, then we say g_i is *active* at x
- Convex minimization can be reposed as concave maximization. (3.1) is equivalent to

$$\begin{aligned} \underset{x\in D}{\text{maximize}} \quad & -f(x) \\ \text{subject to} \quad & g_i(x) \le 0,\ i = 1,\dots,m \\ & Ax = b. \end{aligned}$$

  We call both convex optimization problems.

3.1.1 Solution Set

Let X_{opt} be the set of all solutions of a given convex problems, written

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

Lemma 3.1 *X_{opt} is a convex set*

Proof: Using definitions. If $x, y \in X_{opt}$, then for $0 \leq t \leq 1$,

- $g_i(tx + (1-t)y) \leq tg_i(x) + (1-t)g_i(y) \leq 0$
- $A(tx + (1-t)y) = tAx + (1-t)Ay = b$
- $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) = f^*$

It follows that $tx + (1-t)y)$ is also a solution. ■

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

3.1.2 Example: lasso

Given $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times p}$, consider the lasso problem:

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} \quad & \|y - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq s \end{aligned}$$

Is this convex? What is the criterion function? The inequality andequality constraints? Feasible set? Is the solution unique, when:

- $n \geq p$ and X has full column rank?
- $n \leq p$ (high-dimensional case)?

How do our answers change if we changed criterion to Huber loss:

$$\sum_{i=1}^{n} \rho(y_i - x_i^\intercal \beta), \quad \rho \begin{cases} \frac{1}{2}z^2, & |z| \leq \delta \\ \delta|z| - \frac{1}{2}\delta^2, & otherwise \end{cases}$$

3.1.3 Example: support vector machines

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, \ldots, x_n$, consider the support vector machine or SVM problem:

$$\begin{aligned} \min_{\beta, \beta_0, \xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=0}^{n} \xi_i \\ \text{subject to} \quad & \xi_i \geq 0,\ i = 1, \ldots, n \\ & y_i(x_i^\intercal \beta + \beta_0) \geq 1 - \xi_i,\ i = 1, \ldots, n \end{aligned}$$

Is this convex? What is the criterion, constraints, feasible set? Is the solution (β, β_0, ξ) unique? What if we changed the criterion to

$$\frac{1}{2}\|\beta\|_2^2 + \frac{1}{2}\beta_0^2 + C\sum_{i=0}^{n}\xi_i^{1.01}$$

For original criterion, what about β component, at the solution?

3.1.4 Local Minima are Global Minima

It turns out that for convex optimization problems, any local solution is also gloabally optimal. Formally, we are saying that whenever f is a convex function, if there exists an $R > 0$ such that $f(x) \leq f(y)$ whenever $||x - y||_2 \leq R$ then $f(x) \leq f(y)$ for all y.

3.1.5 Rewriting Constraints

There are multiple ways to write down an optimization problem. Previously we wrote them as

$$\begin{aligned} \underset{x \in D}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned} \tag{3.2}$$

however this is equivalent to writing

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

where $C = \{x : g_i(x) \leq 0, i = 1, \ldots, m, Ax = b\}$ is the feasible set. Another way of writing the same problem is

$$\min_x f(x) + I_C(x) \tag{3.4}$$

where I_C is the indicator of C.

3.1.6 Partial Optimization

We have previously seen that if a function $f(x, y)$ is convex in both arguments and if C is a convex set, then the function $g(x) = \min_{y \in C} f(x, y)$ is also convex in x. This allows us to partially optimize a convex problem and still retain convexity guarantees. For example

$$\begin{aligned} \underset{x_1, x_2}{\text{minimize}} \quad & f(x_1, x_2) \\ \text{subject to} \quad & g_1(x_1) \leq 0 \\ & g_2(x_2) \leq 0 \end{aligned} \tag{3.5}$$

is equivalent to

$$\begin{aligned} \underset{x_1}{\text{minimize}} \quad & \tilde{f}(x_1) \\ \text{subject to} \quad & g_1(x_1) \leq 0 \end{aligned} \tag{3.6}$$

where $\tilde{f}(x_1) = \min\{f(x_1, x_2) : g(x_2) \leq 0\}$.

3.1.7 Hierarchy of Convex Programs

It turns out that there are a bunch of interesting sub-classes of convex problems. Some of these include linear programs, qudaratic programs semidefinite programs and cone programs. These programs can be related as follows: Linear Programs $\subset$ Quadratic Programs $\subset$ Semidefinite Programs $\subset$ Conic Programs $\subset$ Convex Programs.

3.1.8 Linear Programs

A linear program is a special type of convex program. Any program that can be formulated as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Dx \leq d \\ & Ax = b \end{aligned} \tag{3.7}$$

is a linear program. Some methods for solving linear programs are the simplex algorithm and interior point methods.

3.1.9 Geometric Programming

A monomial is a function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ of the form:

$$f(x) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

for $\gamma > 0$, $a_1, a_2, \cdots, a_n \in \mathbb{R}$.

A posynomial is a sum of monomials,

$$f(x) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k1}} x_2^{a_{2k}} \cdots x_n^{a_{kn}}$$

.

A geometric program is of the form:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 1, i = 1, \ldots, m \\ & h_j(x) = 1, j = 1, \ldots, r \end{aligned} \tag{3.8}$$

where $f, g_i, i = 1, \ldots, m$ are posynomials and $h_j, j = 1, \ldots, r$ are monomials. This is non-convex.

Given $f(x) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$, let $y_i = \log x_i$ and rewrite this as:

$$\gamma (e^{y_1})^{a_1} (e^{y_2})^{a_2} \cdots (e^{y_n})^{a_n} = e^{a^T y + b}$$

for $b = \log \gamma$.

Also, a posynomial can be written as $\sum_{k=1}^{p} e^{a_k^T y + b_k}$. With this variable substitution, and after taking logs, a geometric program is equivalent to:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & \log\left(\sum_{k=1}^{p_0} e^{a_{0k}^T y + b_{0k}}\right) \\
\text{subject to} \quad & \log\left(\sum_{k=1}^{p_i} e^{a_{ik}^T y + b_{ik}}\right) \leq 0, i = 1, \ldots, m \\
& c_j^T y + d_j = 0, j = 1, \ldots, r
\end{aligned}
\tag{3.9}
$$

This is convex (recalling the convexity of softmax functions).

3.1.10 Eliminating Equality Constraints

Important special case of change of variables: eliminating equality constraints. Given the problem:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, i = 1, \ldots, m \\
& Ax = b
\end{aligned}
\tag{3.10}
$$

we can always express any feasible point as $x = My + x_0$, where $Ax_0 = b$ and col(M)=null(A). Hence the above is equivalent to:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(My + x_0) \\
\text{subject to} \quad & g_i(My + x_0) \leq 0, i = 1, \ldots, m
\end{aligned}
\tag{3.11}
$$

3.1.11 Introducing Slack Variables

Essentially opposite to eliminating equality contraints: introducing slack variables. Given the problem:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, i = 1, \ldots, m \\
& Ax = b
\end{aligned}
\tag{3.12}
$$

we can transform the inequality constraints via:

$$
\begin{aligned}
\underset{x,s}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & s_i \geq 0, i = 1, \ldots, m \\
& g_i(x) + s_i = 0, i = 1, \ldots, m \\
& Ax = b
\end{aligned}
\tag{3.13}
$$