Math 612: Single Cell Analysis **2019W Term 1**

Lecture 10: October 8

Lecturer: Geoffrey Schiebinger *Scribe: Sophie Boerlage*

10.1 Convex Optimization

10.1.1 The Optimization Problem

We want to minimize the **objective function**, $f_0(x)$, over the optimization variable, x, subject to constraints.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0 \quad i = 1, ..., m \\ & h_i(x) = 0 \quad i = 1, ..., n \end{aligned} \tag{10.1}$$

The solution is a point x^*, the **optimizer**, which satifies the constraints and has the lowest objective value.

$$f_0(x^*) \leq f_o(x)$$

for all other feasible x (i.e. that satisfy the constraints).

The problem can be very hard, and may find local minima. However, there are many well-developed tools for optimizing convex functions. For this group of functions there exist provable guarantees and known rates of convergence.

10.1.2 Examples of Nice Optimization Problems

Least Squares

The least squares problem:

$$\underset{x}{\text{minimize}} \quad ||Ax - b||_2^2$$

has an analytical solution of the form:

$$x^* = (A^T A)^{-1} A^T b$$

While this problem has an analytical solution, for most problems this does not exist. We will learn how to use an algorithm to get close to the solution.

Linear Program

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & a_i^T x \leq b_i \quad i = 1, ..., m \end{aligned} \tag{10.2}$$

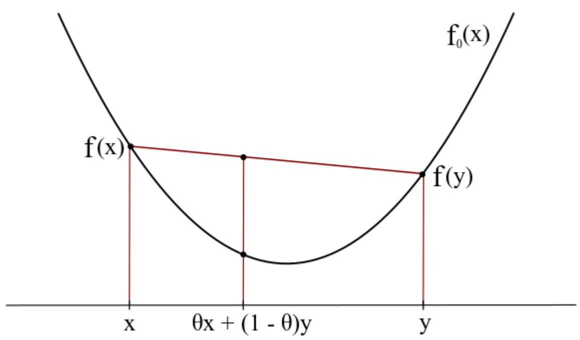
10.1.3 General Convex Optimization Problem

Recall the problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0 \quad i = 1, ..., m \end{aligned} \tag{10.3}$$

where f_i are all convex.

Def. A function is convex if:



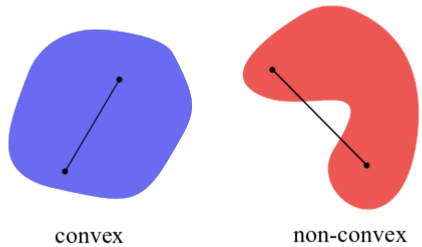
$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta) f(y) \tag{10.4}$$

for any $\theta \in [0, 1]$.

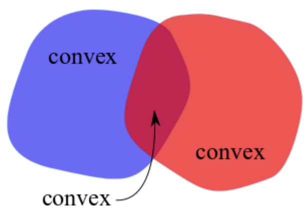
A function is strictly convex if the inequality is strict, and strongly convex if the Hessian has positive curvature.

The constraints $f_i(x) \leq 0$ define a **convex set.**

Def. A set $C \subset \chi$ is convex if the line between any two points remains in the set for any $x_1, x_2 \in C \rightarrow \theta x_1 + (1 - \theta) x_2 \in C$ for $\theta \in [0, 1]$.



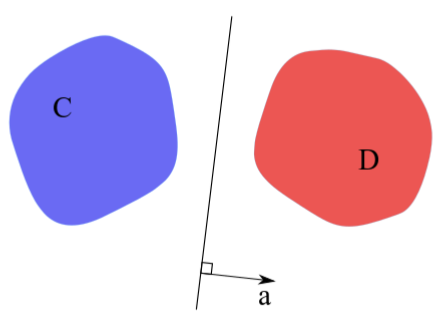
Fact: The intersection of convex sets is convex



Fact: If C and D are disjoint convex sets, then there is a hyperplane $a \in \mathbb{R}^d$ such that:

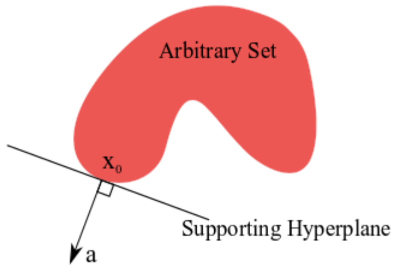
$$a^T x \leq b \quad x \in C$$
$$a^T x \geq b \quad x \in D$$

This is the **separating hyperplane theorem**



Fact: If C is a convex set, then there is a supporting hyperplane at every boundary point x_0.

$$a^T x \leq a^T x_0 \quad \text{for all } x \in C$$



10.1.4 Convex Functions

f is convex if $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$

For a general random variable z (for example, $z = x \; w.p. \; \theta, z = y \; w.p. \; 1-\theta$)

$$f(\mathbb{E}z) \leq \mathbb{E}(f(z)) \tag{10.5}$$

for a convex function. This is called **Jenson's Inequality**.

Examples of Convex Functions:

x, x^4, 0, e^x, negative entropy $(-x \log x)$

Example on $\mathbb{R}^n$: Norms

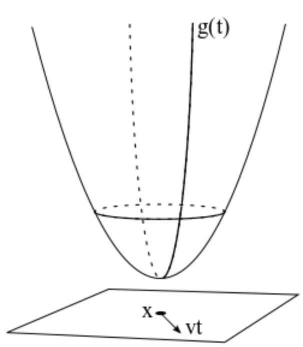
$$||x||_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$$

Example on $\mathbb{R}^{n\times m}$: Matrix Norms

$$||X||_2 = \sigma_{max}(X) = \underset{v}{max} \frac{||Xv||_2}{||v||_2} = f(x)$$

Fact: $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff $g : \mathbb{R} \to \mathbb{R}$

$g(t) = f(x+tv)$ is convex in t for any v. It is possible to check the convexity of f by checking 1-D convexity.



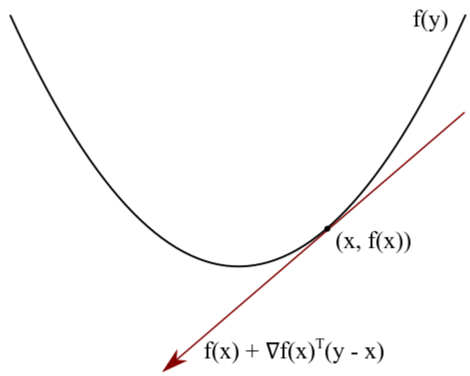
10.1.5 First Order Condition

f is differentiable if:

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), ..., \frac{\partial f}{\partial x_n}(x))$$

exists. A differentiable f (with convex domain) is convex iff:

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

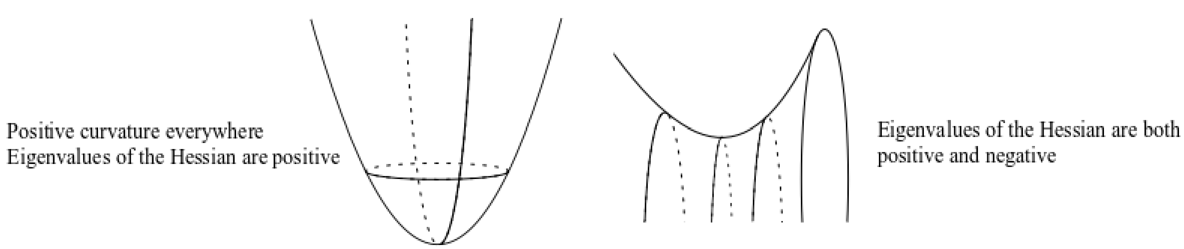


10.1.6 Second Order Condition

f is convex if $\nabla^2 f(x)_{ij} = \frac{\partial f}{\partial x_i \partial x_j}(x)\ \nabla f(x) \succeq 0$

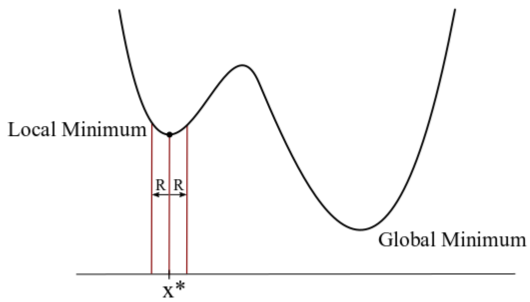
Matrix A is positive semi-definite ($A \succeq 0$) if $x^T A x \geq 0$ (i.e. all eigenvalues ≥ 0).

$\nabla^2 f$ is called the **Hessian** and it measures the curvature at x. If the Hessian $\nabla^2 f(x) \succ 0$ then f is strictly convex. If $\nabla^2 f(x) \succ \mu I$ then f is strongly convex.



Def. A point x^* is a local solution of an optimization problem if it is the optimal solution to:

$$\underset{x}{\text{minimize}} \quad f_0(x) \tag{10.6}$$
$$\text{subject to} \quad f_i(x) \leq 0$$
$$||x^* - x|| \leq R \tag{10.7}$$



For convex optimization problems **all local solutions are global solutions.**

10.1.7 Gradient Descent

Produces the "approximately optimal" point x_k where $f(x_k) \approx f(x^*)$ after k iterations.

Initialize x_0. Set:

$$x_1 = x_0 - \eta \nabla f_0(x_0)$$

where η is the step size and $-\eta \nabla f_0(x_0)$ is the direction of steepest descent.

$$x_{k+1} = x_k - \eta_k \nabla f_0(x_k)$$

Will always converge to a local minimum, and if the function is convex, a global minimum.

However, when close to the minimum, if the step size is too large the value may jump past the optimal point and oscillate about it. Thought needs to be put into selecting the step size, of which $\eta_k = \frac{1}{k}$ is the easiest. Next time: Improve convergence by fitting a quadratic approximation using Newton's method.