1 Convex Sets

A convex set is a set $\mathcal{D}$ with the following property: $\forall x, y \in \mathcal{D}$

$$\alpha x + (1-\alpha)y \in \mathcal{D}, \quad \forall \alpha \in [0,1]$$

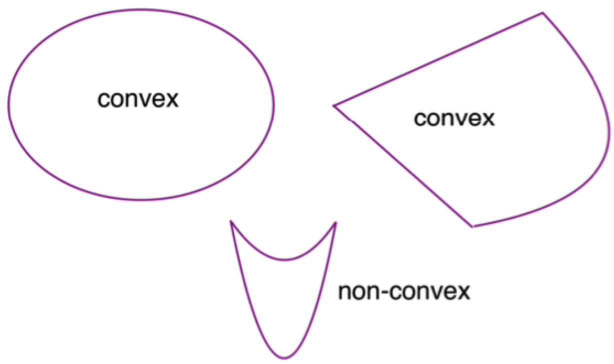

Figure 1: Examples of convex sets. From 10-725 slides by Prof. Yuanzhi Li.

Example: Simple convex sets

- a line
- the empty set
- a single point

Example: Let $\mathcal{C}_i$ be a convex set for $\forall i$. Is the conjunction of the convex sets $\mathcal{C} = \cap\, \mathcal{C}_i$ a convex set? What about the union of convex sets $\mathcal{C}' = \cup\, \mathcal{C}_i$?

2 Convex Functions

A function f over a convex set D is convex if $\forall x, y \in \mathcal{D}$,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \quad \forall \alpha \in [0,1]$$

A function f over a convex set D is *strictly* convex if $\forall x, y \in \mathcal{D}$,

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y), \quad \forall \alpha \in [0,1]$$

A function f over a convex set D is *concave* if $\forall x, y \in \mathcal{D}$,

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y), \quad \forall \alpha \in [0,1]$$

Important Note: Non-convex $\neq$ concave.

Examples:

- $f(x) = x$
- $f(x) = x^2$
- $f(x) = \frac{1}{x},\ x > 0$

Example: If f and g are convex functions, please show that $f + g$ is also a convex function.

Example: If f and g are convex functions, please show that $\max(f, g)$ is also a convex function. Is $\min(f, g)$ a convex function?

Other properties of convex functions: In one-dimensional case, i.e. $f, g : \mathbb{R} \to \mathbb{R}$

- If f and g are convex and g is non-decreasing, $g \circ f$ is convex, e.g. $f(x) = \frac{1}{x}, x > 0$ and $g(x) = x^2$.
- If f is concave and g is convex and non-increasing then $g \circ f$ is convex.

3 Aside: Gradients and Hessians

In machine learning, we're typically concerned with computing derivatives of functions with vector-valued inputs. To formalize this notion, we'll introduce the *gradient* vector and *Hessian* matrix.

3.1 Definitions

For a function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient vector $\nabla f(\mathbf{x}) \in \mathbb{R}^n$ is defined as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots & \frac{\partial f}{\partial x_m} \end{bmatrix}^T$$

On the other hand, for a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the Hessian matrix generalizes second derivatives for scalar functions with vector inputs and is defined as $\nabla^2 f(\mathbf{x})_{ij} = \frac{\partial f}{\partial x_i \partial x_j}$. In this handout, we will present all the derivatives in denominator layout - i.e., each row i of the gradient vector, which is a column vector, corresponds to the derivative of f with respect to an input variable x_i. The definition above uses denominator layout. In numerator layout, the gradient vector would be a row vector, with each column i corresponding to the derivative with respect to an input variable x_i.

3.2 Examples

1. Let $y = f(\mathbf{x}) = 3x_1^2 \sin x_2$. What are $\nabla f(x)$ and $\nabla^2 f(\mathbf{x})$?

2. Let $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$. What are $\nabla f(x)$ and $\nabla^2 f(\mathbf{x})$?

Additional Notes: In single variable calculus, we have the Taylor series, which is an approximation of $f(y)$ using an infinite sum of terms that include derivatives at a particular point, i.e. for a constant x,

$$f(y) \approx \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(y-x)^n$$

For multivariable functions, we can write the following second order Taylor expansion:

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\mathbf{x})(\mathbf{y}-\mathbf{x}).$$

Inequalities relating the LHS and RHS as well as a bound on the Hessian term are frequently used in convex optimization.

4 Check Convexity

First Order Condition: Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is convex if and only if $\forall x, y \in \mathbb{R}^n$

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

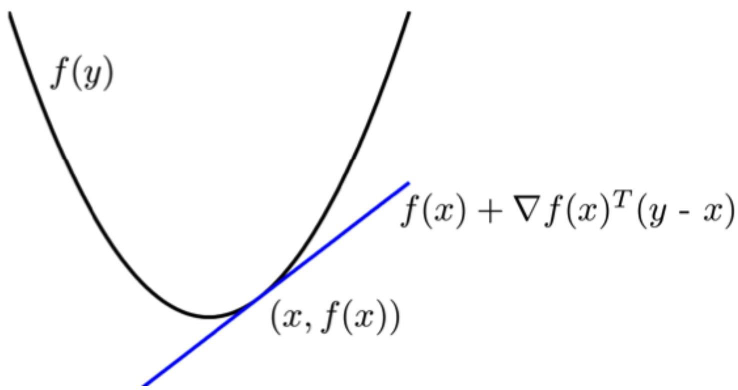

Figure 2: First Order Condition. From 10-725 slides by Prof. Zhiyuan Li.

Second Order Condition: Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then f is convex if and only if $\forall x \in \mathbb{R}^n$

$$\nabla^2 f(x) \succcurlyeq 0$$

i.e. the eigenvalues of $\nabla^2 f(x)$ are all non-negative.

Example: Show the MLE target function $f(\theta) = \log \prod_{i=1}^{n} \theta^{X_i} (1 - \theta)^{1 - X_i}$ is concave.

5 Minimum & Maximum

Local minimum If f is differentiable and twice differentiable, $f(x^*)$ is the local minimum of the function f

$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \succcurlyeq 0$$

Global minimum If f is convex, then any local minimum of f is also a global minimum.

Saddle point If f is differentiable and twice differentiable, $f(x^*)$ is a saddle point if

$$\nabla f(x^*) = 0 \quad \text{but not} \quad \nabla^2 f(x^*) \succcurlyeq 0$$

Example: Saddle point $f(x) = x^3$ at $x = 0$

Another example: Potato chips