Convexity

Andrew Nobel

September, 2021

Convex Sets and Functions

Convex Sets

Definition: A set $C \subseteq \mathbb{R}^d$ is *convex* if for every pair of points $x, y \in C$ and every $\alpha \in [0, 1]$ the point $\alpha x + (1 - \alpha)y \in C$.

Interpretation

- Vector $\alpha x + (1 - \alpha)y$ called *convex combination* of x, y with weight α
- Set $\{\alpha x + (1 - \alpha)y : \alpha \in [0, 1]\}$ is just the line between x and y
- So C is convex if the line between any two points in C is contained in C

Examples of Convex Sets

- In 1-d convex sets are intervals, e.g., $[0, 1]$, $(-2, 5)$, $(0, \infty)$, and $\mathbb{R}$
- More generally: $\emptyset$, $\{0\}$, and line $C = \{\gamma\, x : \gamma \in \mathbb{R}\}$ are convex in $\mathbb{R}^d$
- Ball of radius r centered at x_0, $B(x_0, r) := \{x : ||x - x_0|| < r\}$
- Halfspace with direction w and offset b, $H(w, b) = \{x : w^t x \geq b\}$
- Hyperplane $\partial H(w, b) = \{x : w^t x = b\}$, boundary of halfspace $H(w, b)$
- Polyhedron $\{x : Ax \leq c\}$ where $\leq$ understood componentwise
- Probability simplex $\{u : u_i \geq 0$ and $\sum_{i=1}^{d} u_i = 1\}$

Basic Properties of Convex Sets

Fact: If $C_1, C_2, \ldots$ are convex sets, then so is their intersection $\cap_{i \geq 1} C_i$

Note: A union of convex sets is generally *not* convex

Fact: Other ways of combining convex sets to get new ones

- If $A, B \subseteq \mathbb{R}^d$ are convex then so is $A + B = \{u + v : u \in A$ and $v \in B\}$
- If $A \subseteq \mathbb{R}^d$ is convex and $c \in \mathbb{R}$ then $cA = \{cu : u \in A\}$ is convex

Convex Functions

Definition: Let $C \subseteq \mathbb{R}^d$ be convex. A function $f : C \to \mathbb{R}$ is *convex* if for every $x, y \in C$ and every $\alpha \in (0, 1)$,

$$f(\alpha x + (1-\alpha)y) \;\leq\; \alpha f(x) + (1-\alpha) f(y) \quad (*)$$

Convexity of C ensures that $f(\cdot)$ is defined at each point $\alpha x + (1-\alpha)y$

Interpretation: Line connecting $(x, f(x))$ and $(y, f(y))$ lies *above* the graph of f

Related Definitions

- $\alpha x + (1-\alpha)y$ is called a *convex combination* of x and y
- $f : C \to \mathbb{R}$ is *strictly convex* if $(*)$ holds with $\leq$ replaced by $<$
- $f : C \to \mathbb{R}$ is *concave* if $(*)$ holds with $\leq$ replaced by $\geq$

How to Verify Convexity or Concavity of Functions

1. Check the definition: In many cases it is possible to directly check the definition

2. Second derivative condition:

- $f : C \to \mathbb{R}$ is convex if the matrix $\nabla^2 f(x)$ of second partials derivatives is non-negative definite for each $x \in C$
- $f : C \to \mathbb{R}$ is concave if the matrix $\nabla^2 f(x)$ of second partial derivatives is non-positive definite for each $x \in C$

Special case: A twice-differentiable function $f : \mathbb{R} \to \mathbb{R}$ is convex if $f'' \geq 0$ and is concave if $f'' \leq 0$

Examples of Convex/Concave Functions

Case $d = 1$

- $f(x) = |x|$ is convex, but *not* strictly convex
- $f(x) = x^2$, e^x, e^{-x}, x^{-1}, and $x \log x$ are strictly convex
- $f(x) = \log x$, $\sqrt{x}$ are strictly concave

Case $d \geq 2$

- $f(x) = ||x||$ is convex
- $f(x) = \langle x, u \rangle + b$, affine function, is convex and concave
- $f(x) = \max_{u \in A} \langle x, u \rangle$, where $A \subseteq \mathbb{R}^d$ is bounded, is convex
- $f(x) = x^t Ax$, quadratic form, is convex if $A \geq 0$, concave if $A \leq 0$

Basic Properties of Convex Functions

Fact

(a) If $f : C \to \mathbb{R}$ is convex then $-f$ is concave, and vice-versa

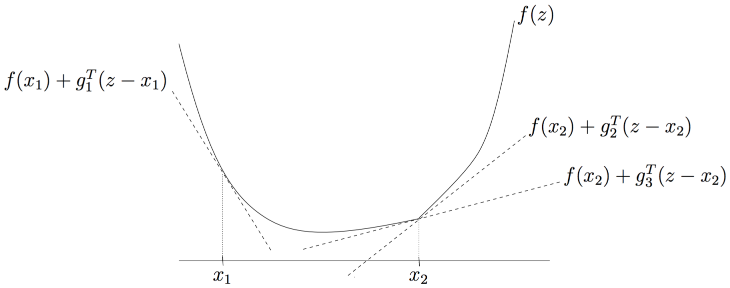
(b) If $f_1, \ldots, f_m : C \to \mathbb{R}$, are convex then $f(x) = \max_{1 \leq i \leq m} f_i(x)$ is convex

Fact: If $f : C \to \mathbb{R}$ is convex, then for every $u \in C$ there is a vector $v \in \mathbb{R}^d$ such that

$$f(x) \;\geq\; f(u) + \langle v, x - u \rangle \text{ for each } x \in C$$

- The vector v is called a *subgradient* of f at u (not necessarily unique)
- Lower bound $h_u(x) := f(u) + \langle v, x - u \rangle$ is affine in x with $h_u(u) = f(u)$
- Graph of function h_u is a hyperplane supporting the graph of f at u

Subgradient Illustration (credit: John Lambert)



Jensen’s Inequality

Jensen's Inequality in 1-Dimension

Theorem: Let $X \in (a, b)$ be a random variable

(1) The expected value $\mathbb{E}X \in (a, c)$

(2) If $f : (a, b) \to \mathbb{R}$ is convex then $f(\mathbb{E}X) \leq \mathbb{E}f(X)$.

(3) If $f : (a, b) \to \mathbb{R}$ is concave then $f(\mathbb{E}X) \geq \mathbb{E}f(X)$.

Proof: In increasing generality

- When X has two possible values, (2) is just the definition of a convex function
- Case of finite valued X follows by induction
- General case follows from the existence of subgradient at the point $u = \mathbb{E}X$

Some Applications of Jensen's Inequality

Fact: Provided all expectations are well defined

- $\mathbb{E}|X| \geq |\mathbb{E}X|$ and $\mathbb{E}X^2 \geq (\mathbb{E}X)^2$ and $\mathbb{E}e^X \geq e^{\mathbb{E}X}$
- If $X > 0$ then $\mathbb{E}(X \log X) \geq (\mathbb{E}X) \log(\mathbb{E}X)$
- If $X > 0$ then $\mathbb{E} \log X \leq \log \mathbb{E}X$ and $\mathbb{E}\sqrt{X} \leq \sqrt{\mathbb{E}X}$

AM-GM inequality: If $a_1, \ldots, a_n > 0$ then $\left(\prod_{i=1}^n a_i\right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n a_i$

Cauchy-Schwartz: If X and Y are r.v. then $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2 \, \mathbb{E}Y^2}$

Somewhat Fancier Stuff

Jensen's Inequality in Higher Dimensions

Recall: The expected value of a random vector $X = (X_1, \ldots, X_d)^t$ is defined by

$$\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_d)^t \in \mathbb{R}^d$$

Jensen's Inequality: Let $C \subseteq \mathbb{R}^d$ be convex and suppose that $X \in C$. Provided that all expectations are well-defined, the following hold.

(1) The expectation $\mathbb{E}X \in C$

(2) If $f : C \to \mathbb{R}$ is convex then $f(\mathbb{E}X) \leq \mathbb{E}f(X)$. If f is strictly convex and X is not constant then the inequality is strict.

(3) If $f : C \to \mathbb{R}$ is concave then $f(\mathbb{E}X) \geq \mathbb{E}f(X)$. If f is strictly concave and X is not constant then the inequality is strict.

Note: Definition of convexity is a special case of (2) for a random vector $X \in C$ with $\mathbb{P}(X = x) = \alpha$ and $\mathbb{P}(X = y) = 1 - \alpha$

Jensen's Inequality, Case $d \geq 2$

Fact: Let $X \in \mathbb{R}^d$ be a random vector. Provided that the expectations are well defined

- $\mathbb{E}(\langle X, \mathbf{u} \rangle + b) = \langle \mathbb{E}X, \mathbf{u} \rangle + b$ (by linearity)
- $\mathbb{E}||X|| \geq ||\mathbb{E}X||$
- $\mathbb{E}(X^t \mathbf{A} X) \leq (\mathbb{E}X)^t \mathbf{A} \, (\mathbb{E}X)$ if $\mathbf{A} \leq 0$

Holder's Inequality

Fact: Let $a, b \geq 0$ and $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. Then

$$p^{-1} a^p + q^{-1} b^q \geq ab$$

Holder's Inequality: Let $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. If X, Y are random variables such that $\mathbb{E}|X|^p$ and $\mathbb{E}|Y|^q$ are finite, then

$$|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|Y|^q)^{1/q}$$

Corollary (Cauchy-Schwartz): If $\mathbb{E}X^2$ and $\mathbb{E}Y^2$ finite then $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2 \, \mathbb{E}Y^2}$

Integral Version of Holder's Inequality

Theorem: If $p, q \geq 0$ satisfy $1/p + 1/q = 1$ and $f, g, h : \mathbb{R}^d \to \mathbb{R}$ with $h \geq 0$ then

$$\int |f(x)g(x)|h(x)\,dx \;\leq\; \left(\int |f(x)|^p h(x)\,dx\right)^{1/p} \left(\int |g(x)|^q h(x)\,dx\right)^{1/q}$$

provided that all the integrals are finite.