Convexity

Daniel P. Robinson
Department of Applied Mathematics and Statistics
Johns Hopkins University

September 10, 2020

Notes

Outline

Notes

- Convex sets are central to constrained optimization, stochastic optimization, nonsmooth structured optimization, and even unconstrained optimization

Definition (convex sets)

A set $\mathcal{S}$ is said to be a convex set if for any two points in $\mathcal{S}$, the entire line segment joining the points is also contained in $\mathcal{S}$, i.e., a set $\mathcal{S}$ is convex if and only if for all x and y in $\mathcal{S}$, it follows that

$$\alpha x + (1-\alpha)y \in \mathcal{S} \quad \text{for all } \alpha \in [0,1]$$

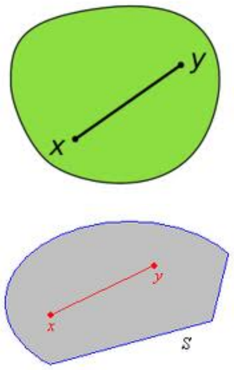


Figure: Convex sets.

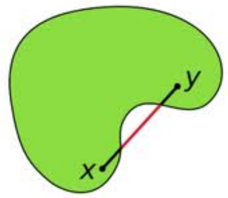


Figure: Nonconvex set.

Notes

Intersections

Theorem

The intersection of convex sets is itself a convex set, i.e., if $\mathcal{S}_i$ is convex for $i \in \mathcal{I}$, then

$$\mathcal{S} \stackrel{\text{def}}{=} \cap_{i\in\mathcal{I}} \mathcal{S}_i$$

is convex.

Proof:
Let x and y be in $\mathcal{S}$. Thus,

$$x \in \mathcal{S}_i \quad \text{and} \quad y \in \mathcal{S}_i \quad \text{for all } i \in \mathcal{I}.$$

Since each $\mathcal{S}_i$ is convex, we know that

$$\alpha x + (1-\alpha)y \in \mathcal{S}_i \quad \text{for all } \alpha \in [0,1] \text{ and } i \in \mathcal{I}.$$

Thus,

$$\alpha x + (1-\alpha)y \in \mathcal{S} \quad \text{for all } \alpha \in [0,1]$$

which completes the proof. ∎

Notes

Definition (convex combination)

A point x is a convex combination of $x_1, x_2, \ldots, x_k$ if there exist $\{\alpha_1, \alpha_2, \ldots, \alpha_k\} \geq 0$ such that

$$\sum_{i=1}^{k} \alpha_i = 1 \quad \text{and} \quad x = \sum_{i=1}^{k} \alpha_i x_i$$

Definition

The convex hull of a set $\mathcal{S}$, denoted by $\mathbf{conv}(\mathcal{S})$, is defined as the smallest (with respect to set inclusion) convex set that contains $\mathcal{S}$. More formally, for any convex set X such that $\mathcal{S} \subseteq X$, we must have $\mathbf{conv}(\mathcal{S}) \subseteq X$. The convex hull $\mathbf{conv}(\mathcal{S})$ may be characterized in two equivalent ways:

1. the intersection of all convex sets containing $\mathcal{S}$
2. the set of all convex combinations of points in $\mathcal{S}$

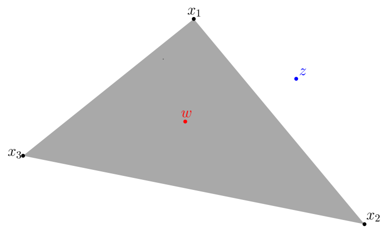


Figure: Convex combinations.

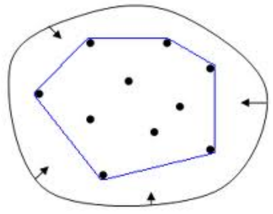


Figure: Convex hull.

Notes

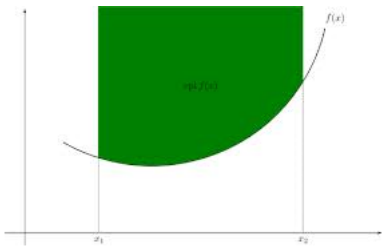
Notes

- Convex functions are ubiquitous in many ares of optimization theory
  - line-search methods for unconstrained and constrained optimization
  - trust-region methods for unconstrained and constrained optimization
  - stochastic optimization
  - (structured) nonsmooth optimization
  - ℓ_1 regularization for inducing sparsity
  - hinge-loss function associated with the Support Vector Machine
- Convex functions have important properties and are very important in optimization even when the objective f is nonconvex (e.g., line-search methods)
- Optimization algorithms make use of the properties of convex functions

Definition

If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ where
$$\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{+\infty\}$$
then

- the effective domain of f is the set of points in $\mathbb{R}^n$ over which it is not equal to positive infinity, i.e.,
$$\mathbf{dom}(f) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : f(x) < +\infty\}$$
- the epigraph of f is the set of points in $\mathbb{R}^n \times \mathbb{R}$ that "lies above" f, i.e.,
$$\mathbf{epi}(f) \stackrel{\text{def}}{=} \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y \geq f(x)\}$$



Notes

There are several equivalent definitions of a convex function

Definition 1: Convex function

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if its epigraph is a convex set.

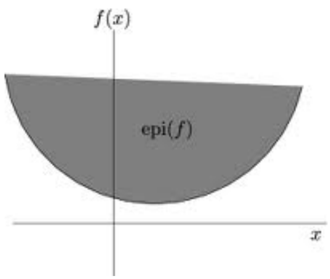


Figure: Convex function.

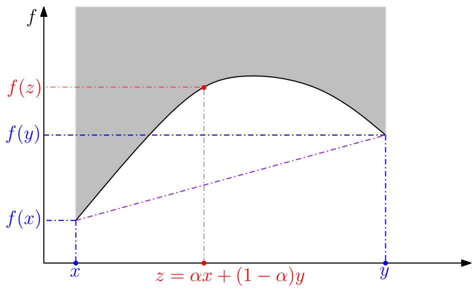


Figure: Nonconvex function.

Notes

Definition 2: Convex function

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex if for all x_1 and x_2 in $\mathbb{R}^n$ and $\alpha \in (0, 1)$, it follows that

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$$

In other words, the function f is convex if it "lies below" the line segment joining any two function values.

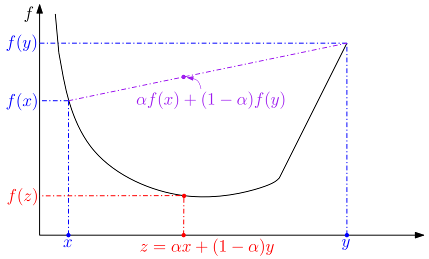


Figure: Convex function.

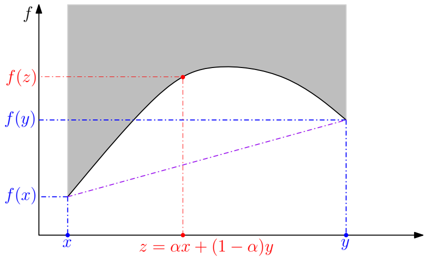


Figure: Nonconvex function.

- f is strictly convex if the above inequality is strict for all $x_1 \neq x_2$

Notes

Definition

A function f is (strictly) concave if $-f$ is (strictly) convex.

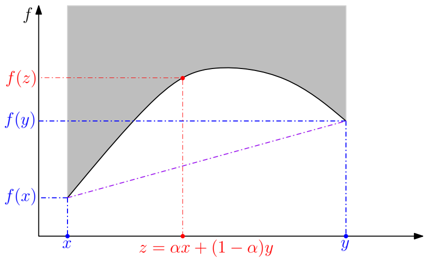


Figure: A concave function.

- A function is simultaneously convex and concave if and only if it is affine
- Just because a function is not convex, does not mean that it is concave!
- Nonlinearity and nonconvexity are not the same thing!

Notes

Some examples

Convex functions

- Affine: $\langle a, x \rangle + b$
- Powers: x^a for $x > 0$ and $a \notin (0, 1)$
- Negative entropy: $x \log(x)$ for $x > 0$
- p-norm: $\left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ for $p \geq 1$

Concave functions

- Affine: $\langle a, x \rangle + b$
- Powers: x^a for $x > 0$ and $a \in [0, 1]$
- Logarithms: $\log(x)$ for $x > 0$

Notes

Operations preserving convexity

- Addition: if $f_1, f_2, \ldots, f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ are convex and $\{\alpha_1, \alpha_2, \ldots, \alpha_k\} > 0$, then

$$f(x) = \sum_{i=1}^{k} \alpha_i f_i(x) \quad \text{is convex}$$

- Maximization: if $f_1, f_2, \ldots, f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ are convex, then

$$f(x) = \max\{f_1(x), f_2(x), \ldots, f_k(x)\} \quad \text{is convex}$$

- Pre-Composition with Affine function : if $T : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function, i.e., $T(x) = Ax + b$ for some matrix $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, and $h : \mathbb{R}^m \to \overline{\mathbb{R}}$ is convex, then

$$f(x) = h\big(T(x)\big) \quad \text{is convex}$$

- Post-Composition with nondecreasing, convex function: if $g : \mathbb{R} \to \overline{\mathbb{R}}$ is nondecreasing and convex, and $h : \mathbb{R}^n \to \mathbb{R}$ is convex, then

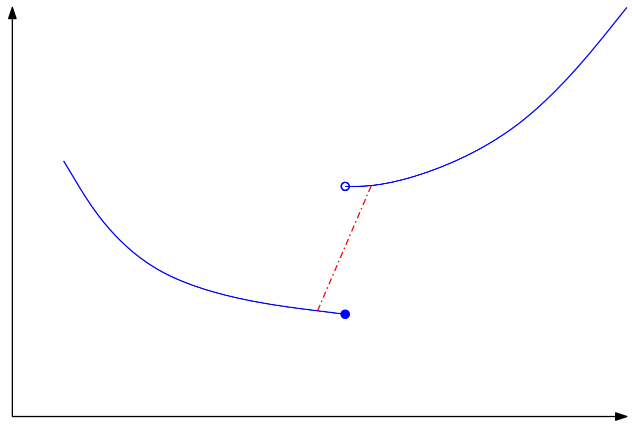
$$f(x) = g\big(h(x)\big) \quad \text{is convex}$$

Notes

Theorem

If $f : \mathcal{S} \subseteq \mathbb{R}^n \to \mathbb{R}$ is a convex function for some open convex set $\mathcal{S}$, then it is continuous on $\mathcal{S}$.

Proof (by picture):



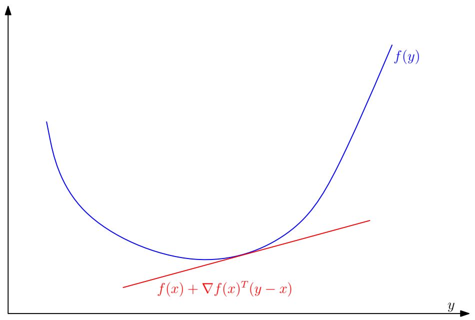
Notes

Convexity using first derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable everywhere. Then the following are equivalent:

1. f is convex.
2. $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ for all $x, y \in \mathbb{R}^n$.
3. $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$ for all $x, y \in \mathbb{R}^n$.

A characterization of strict convexity is obtained if all the above inequalities are considered strict for all $x \neq y \in \mathbb{R}^n$.



- Note: if f is convex, then for any given x the affine function

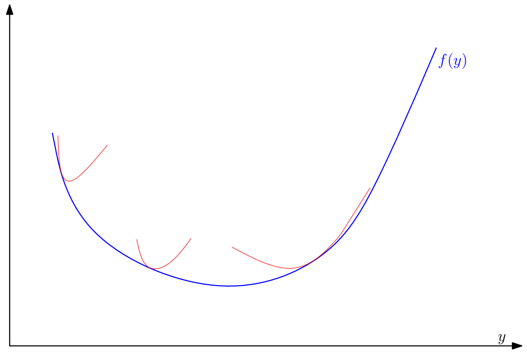
$$f(x) + \nabla f(x)^T(y - x) \quad \text{is a linear underestimator for } f(y).$$

Notes

Convexity using second derivatives

Suppose that f is twice-continuously differentiable. It follows that

- f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$
- if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbb{R}^n$, then f is strictly convex



Question: Why is the second statement not an "if and only if"?

Notes

Notes