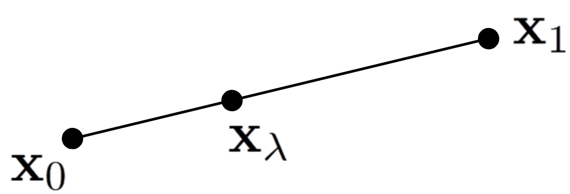
Learning from Data
Convex Sets, Functions, and Optimization

Spring 2017

Convex Set

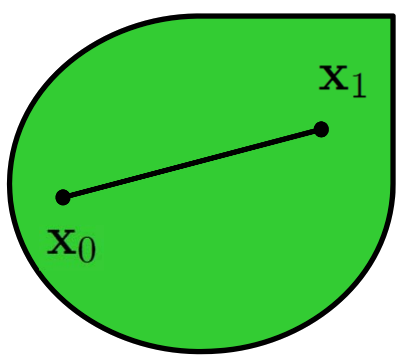
- Line-segment $[\boldsymbol{x}_0, \boldsymbol{x}_1]$ joining two points $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ in $\mathbb{R}^d$ is given by:

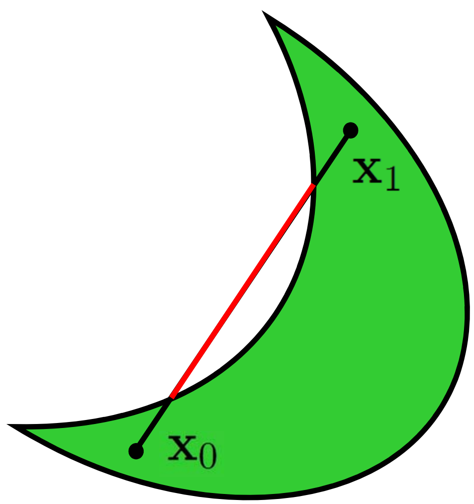
$$\{\boldsymbol{x}_\lambda = (1-\lambda)\boldsymbol{x}_0 + \lambda\boldsymbol{x}_1, \lambda \in [0,1]\}$$



- Convex set: A set $C \subseteq \mathbb{R}^d$ is called convex if for any two points $\boldsymbol{x}_0, \boldsymbol{x}_1 \in C$, the line segment $[\boldsymbol{x}_0, \boldsymbol{x}_1]$ joining them also lies in C:

Convex set

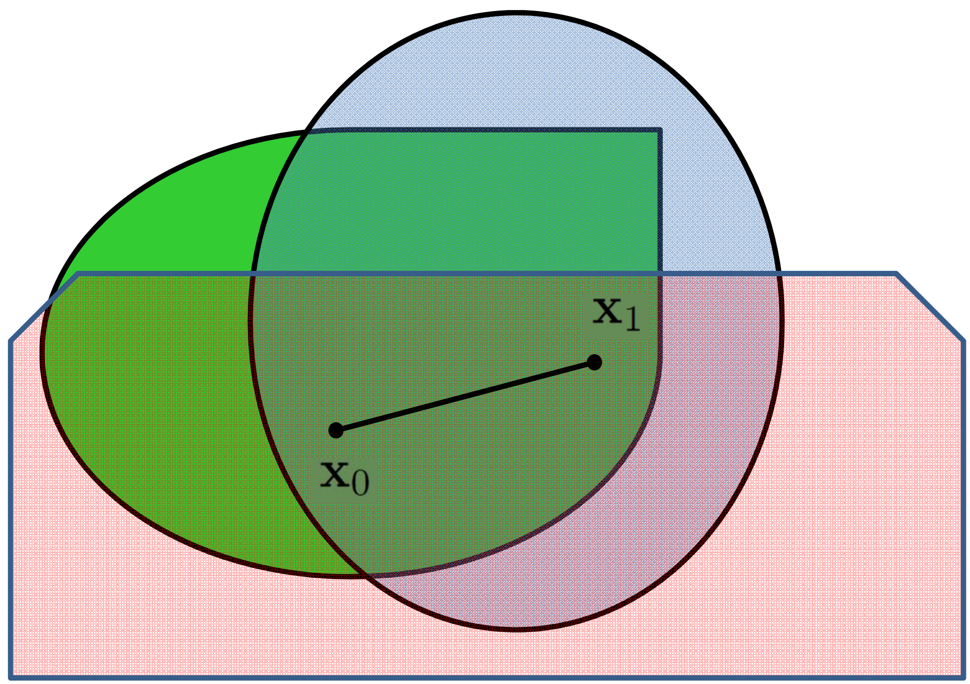




Non-convex set

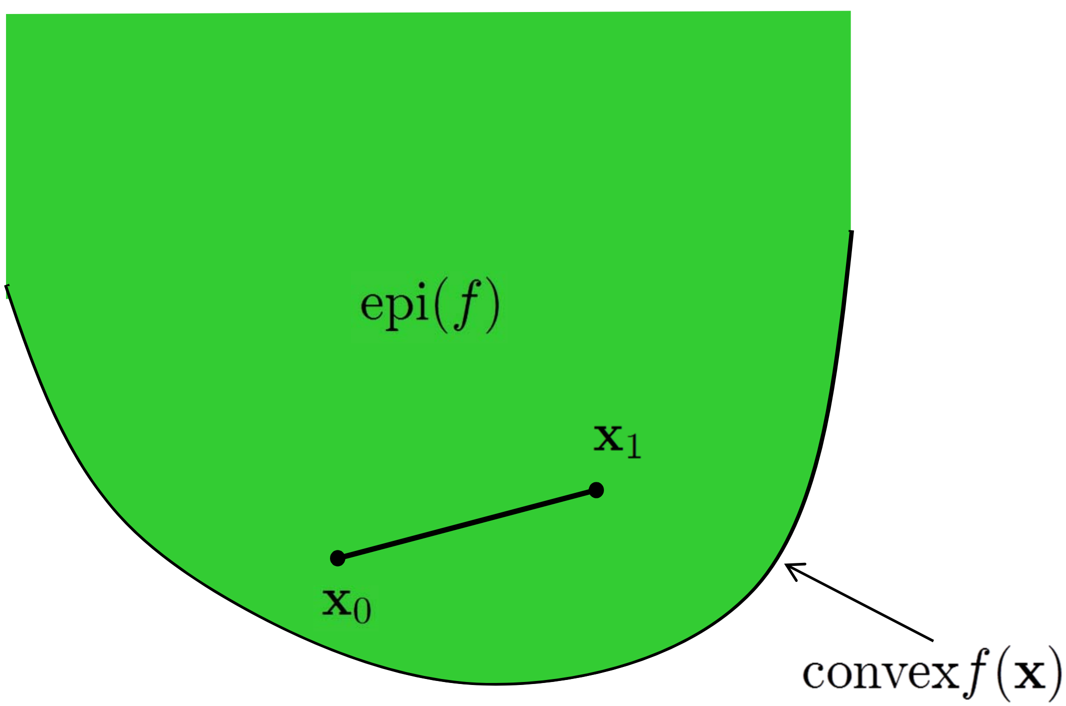
Convex Set

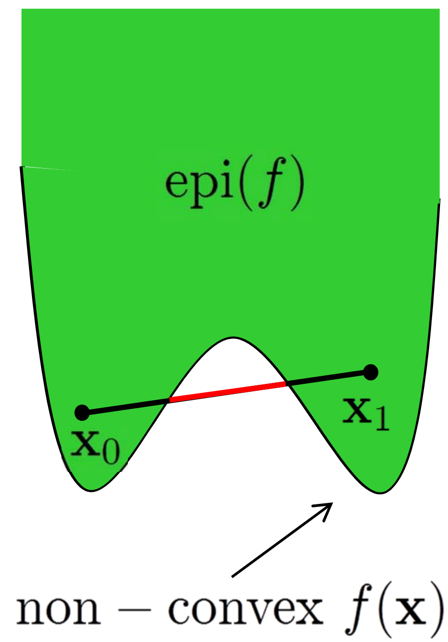
- The intersection of convex sets is convex



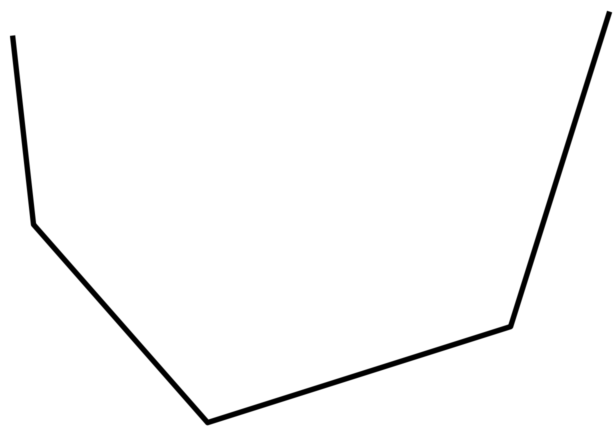
Convex function over a convex set

- Let C be a convex set and $f: C \to \mathbb{R}$ a real-valued function over C
- f is said to be a convex function over C if the region above the graph of the function (called its epigraph or epi(f)) is a convex set.



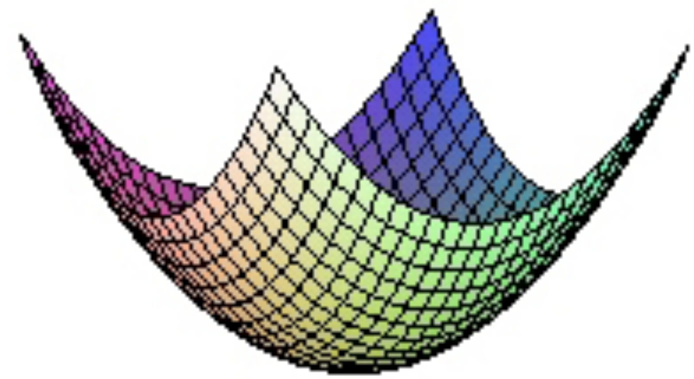


Convex function over a convex set

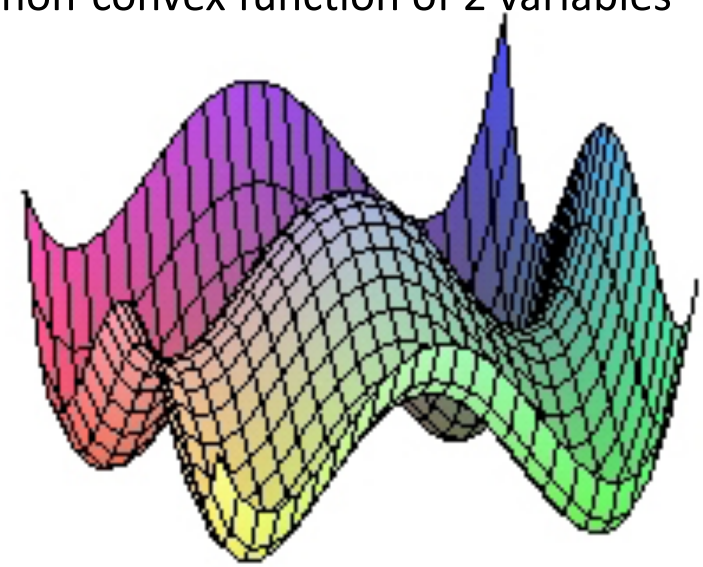


convex, but not
differentiable everywhere

convex function of 2 variables



non-convex function of 2 variables

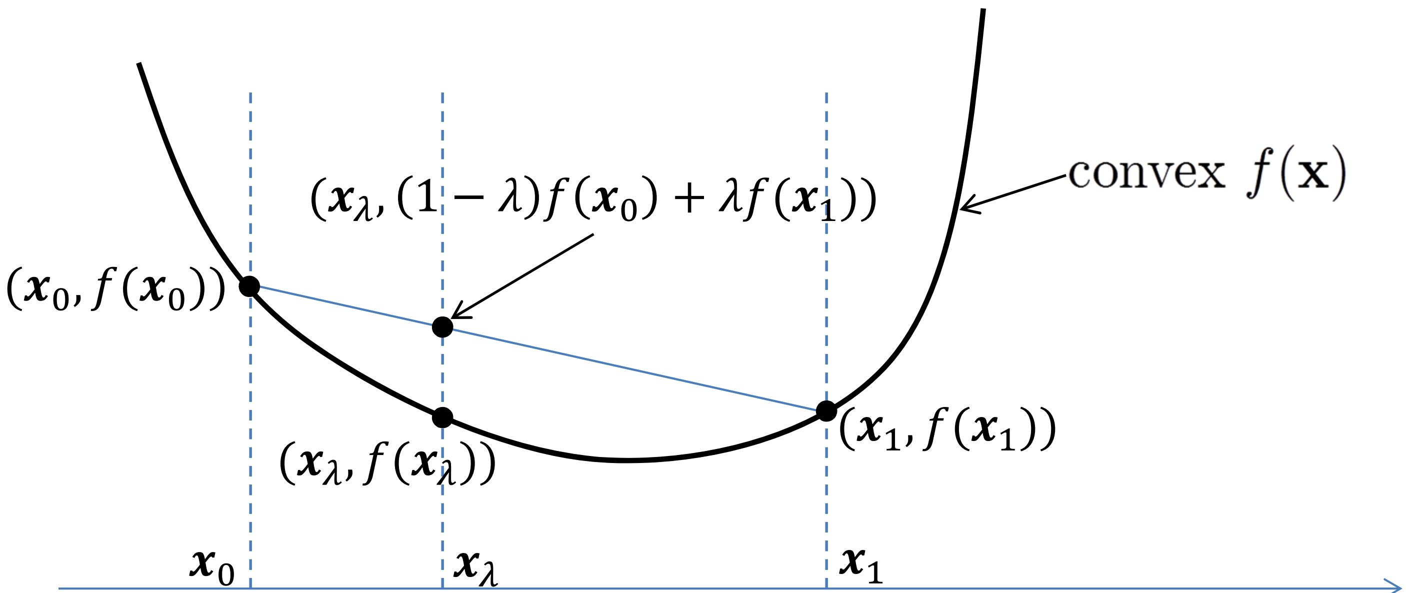


Convex function over a convex set

- Let C be a convex set and $f: C \to \mathbb{R}$ a real-valued function over C
- f is convex over C ⇔ chord joining any two points on the graph, never goes below the graph:

$$\forall \boldsymbol{x}_0, \boldsymbol{x}_1 \in C, \forall \lambda \in [0,1],$$

$$f(\boldsymbol{x}_\lambda) = f\big((1-\lambda)\boldsymbol{x}_0 + \lambda \boldsymbol{x}_1\big) \leq (1-\lambda) f(\boldsymbol{x}_0) + \lambda f(\boldsymbol{x}_1)$$



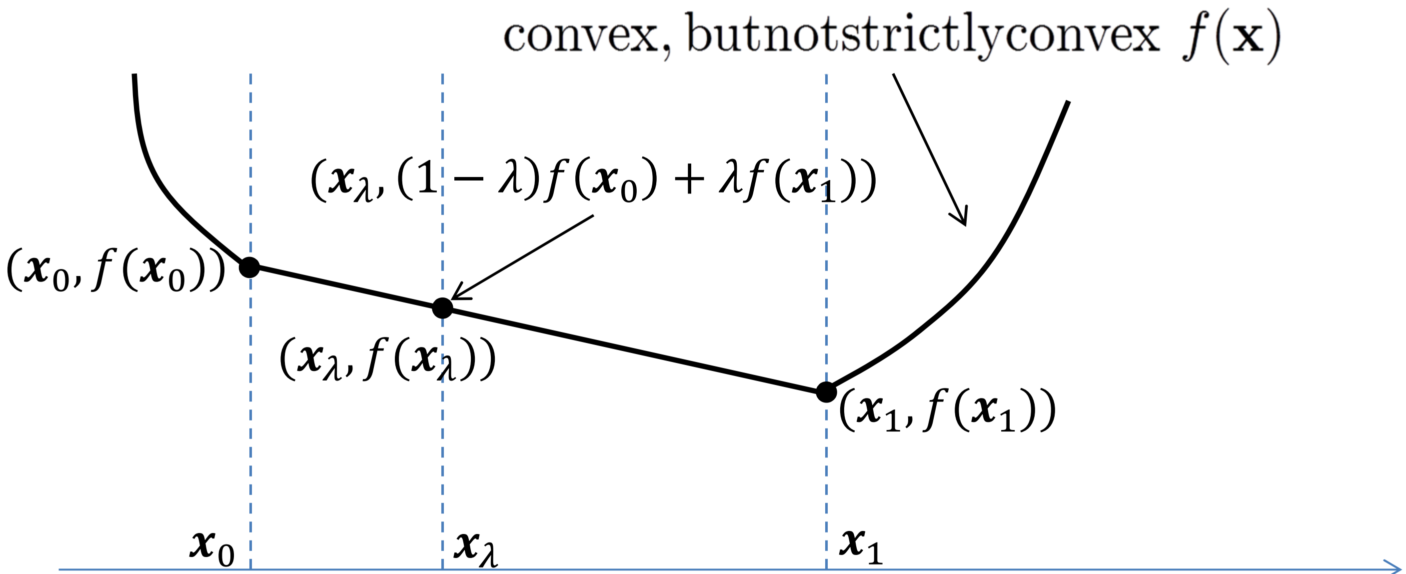
Convex function over a convex set

- f is convex over C ⇔ all chords joining any two points on the graph, never go below the graph:

$$\forall \boldsymbol{x}_0, \boldsymbol{x}_1 \in C, \forall \lambda \in [0,1],$$

$$f(\boldsymbol{x}_\lambda) = f\big((1-\lambda)\boldsymbol{x}_0 + \lambda \boldsymbol{x}_1\big) \le (1-\lambda) f(\boldsymbol{x}_0) + \lambda f(\boldsymbol{x}_1)$$

- If equality only when $\boldsymbol{x}_0 = \boldsymbol{x}_1$ or $\lambda = 0,1$ then f is strictly convex (no planar segments in graph)



Real-valued affine function

- A real-valued function of the form:

$$f(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x} + b,$$

where $\boldsymbol{x}, \boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R}$, is called a real-valued affine function over $\mathbb{R}^d$

- A real-valued affine function is a convex function, but it is not strictly convex

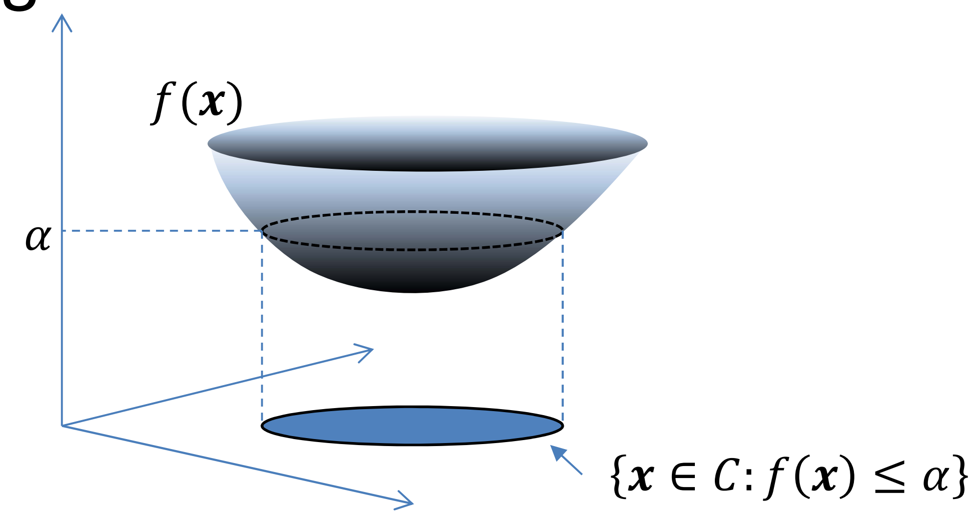
Sublevel sets of a convex function

- The set of points of the form

$$\{\boldsymbol{x} \in C : f(\boldsymbol{x}) \le \alpha\}$$

is called the α-sublevel set of the function f with domain C

- If f is convex over a convex set C, then all its α-sublevel sets are also convex. The reverse does not hold in general



Jensen's inequality for convex functions

- Let f be convex over a convex set C.
- Let $\boldsymbol{X}$ be any random vector whose probability distribution has a support contained in C
- Then,

$$f(E[\boldsymbol{X}]) \leq E[f(\boldsymbol{X})]$$

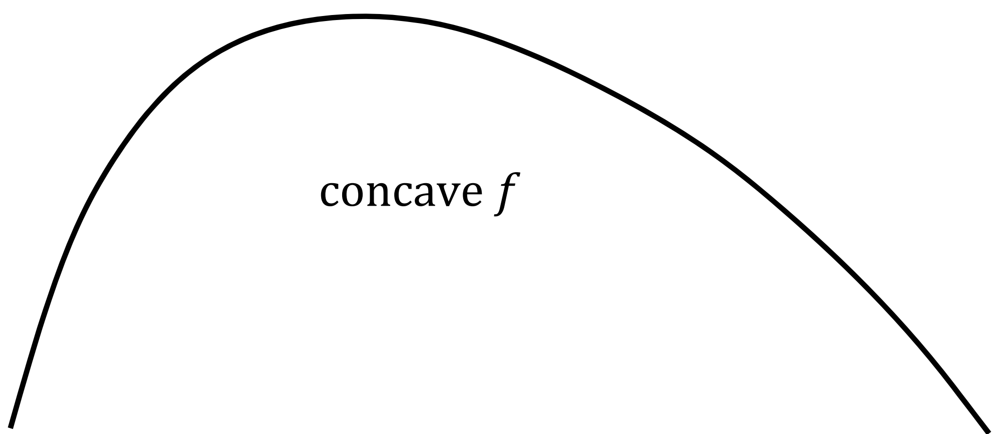
value at mean is not more than mean of values

- If f is also strictly convex, then equality can be attained in Jensen's inequality, if, and only if, $\boldsymbol{X} =$ constant with probability one.

Concave function over a convex set

- f is concave over a convex set $C \Leftrightarrow -f$ is convex over C

- f is strictly concave over a convex set $C \Leftrightarrow -f$ is strictly convex over C



Operations that preserve convexity

- If f is convex, so is αf, for any $\alpha \geq 0$

- If $f_1, f_2, \ldots, f_k$ are each convex, then so is
$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + \cdots + f_k(\boldsymbol{x})$$
The sum of convex functions is convex

- If $f_1, f_2, \ldots, f_k$ are each convex, then so is
$$f(\boldsymbol{x}) = \max(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))$$
The maximum of convex functions is convex

- If f is convex and g is non-decreasing and convex, then $h(\boldsymbol{x}) = g(f(\boldsymbol{x}))$ is convex: a nondecreasing convex function of a convex function is convex

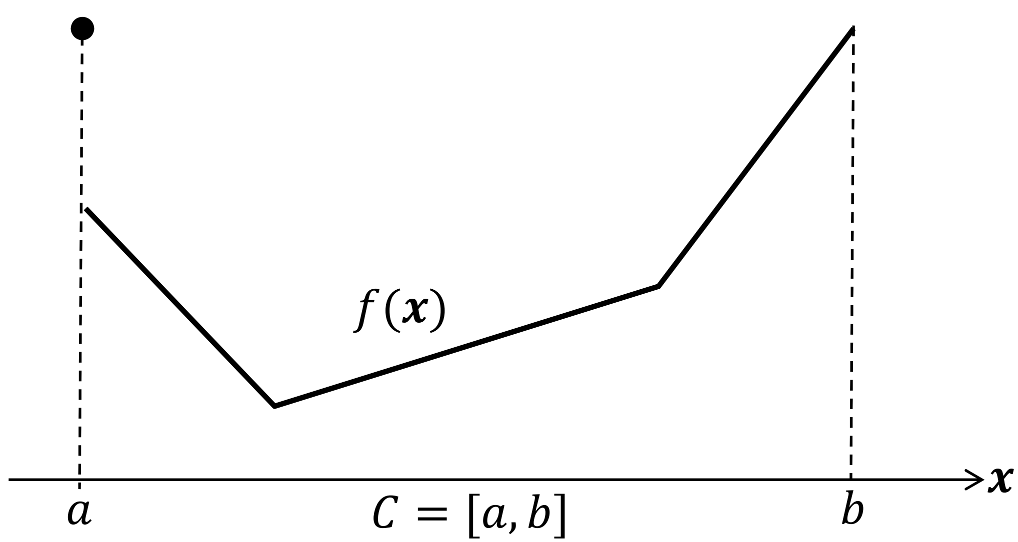
- If $f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^d$, is convex, so is: $h(\boldsymbol{z}) = f(A\boldsymbol{z} + \boldsymbol{b})$, for any $d \times k$ matrix A, $k \times 1$ vector variable $\boldsymbol{z}$, and $d \times 1$ constant vector $\boldsymbol{b}$: a convex function of an affine map is convex

Continuity of convex functions

- Let C be a convex set with a non-empty interior

- If f is convex over C then it is continuous over C's interior, but may have a jump discontinuity at C's boundary

f convex over $[a, b]$. It is continuous over (a, b), but discontinuous at a



Differentiable convex functions

- Let f is be differentiable over a convex set C with gradient vector of first-order partial derivatives:

$$\nabla f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_d}(\boldsymbol{x})\right)^T$$

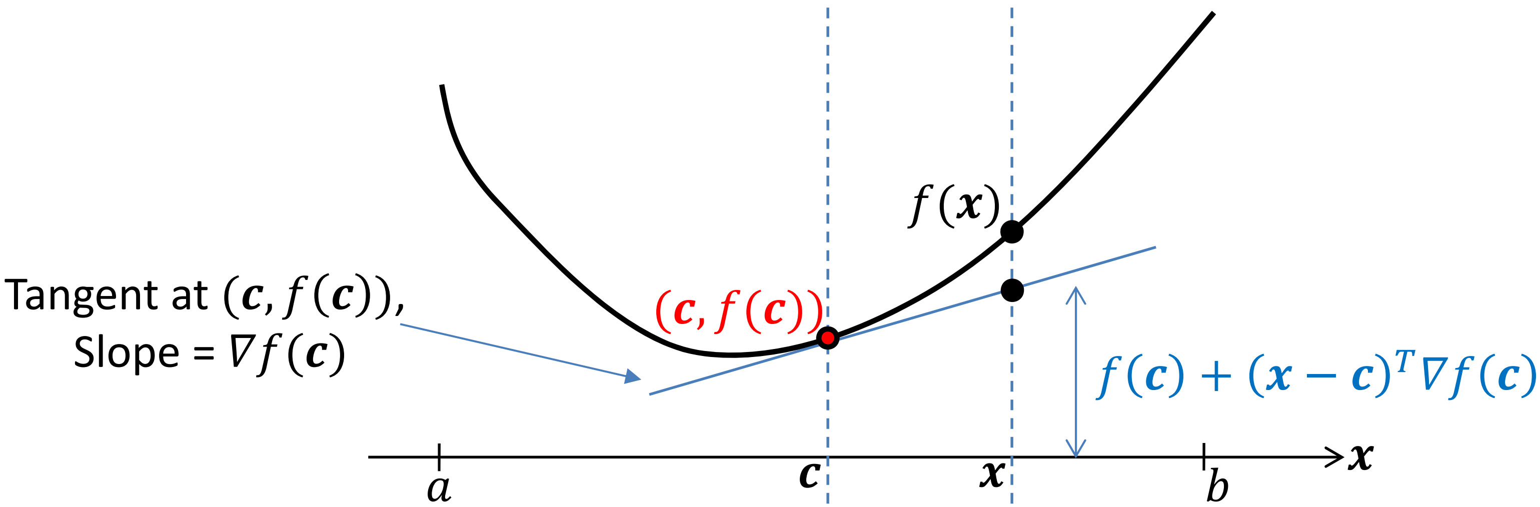
- Then f is convex over C, if, and only if, its graph never goes below the tangent plane constructed at any point:

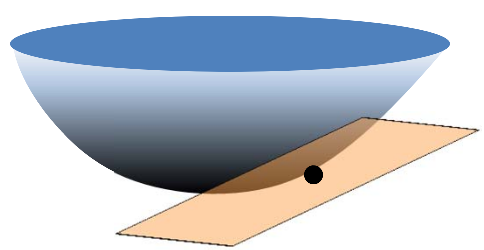
For all $\boldsymbol{x}, \boldsymbol{c} \in C, f(\boldsymbol{c}) + (\boldsymbol{x} - \boldsymbol{c})^T \nabla f(\boldsymbol{c}) \le f(\boldsymbol{x})$

Differentiable convex functions

- For all $\boldsymbol{x}, \boldsymbol{c} \in C, f(\boldsymbol{c}) + (\boldsymbol{x} - \boldsymbol{c})^T \nabla f(\boldsymbol{c}) \leq f(\boldsymbol{x})$

f convex over $C = [a, b]$ and differentiable over (a, b)





Twice-differentiable convex functions

- Let f is be twice-differentiable over a convex set C with a Hessian matrix of second-order partial derivatives:

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_d^2}(\boldsymbol{x}) \end{pmatrix}$$

- Then f is convex (strictly convex) over C, if, and only if, $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite (positive definite) for all $\boldsymbol{x} \in C$:

$$\text{For all } \boldsymbol{x} \in C, \nabla^2 f(\boldsymbol{x}) \geq \boldsymbol{0}$$

Examples

- $C = (0, \infty), f(x) = -\ln(x)$

$\frac{d^2 f(x)}{dx^2} = \frac{1}{x^2} > 0, \forall x \in C \Rightarrow f$ is (strictly) convex over C

$$\Rightarrow \forall x \in C, f(1) + (x-1)f'(1) \leq f(x)$$
$$\Rightarrow \forall x \in C, 0 + (x-1)(-1) \leq -\ln(x)$$
$$\Rightarrow \forall x \in C, \ln(x) \leq x - 1$$

- $C = (-\infty, \infty), f(x) = e^x$

$\frac{d^2 f(x)}{dx^2} = e^x > 0, \forall x \in C \Rightarrow f$ is (strictly) convex over C

$\Rightarrow$ By Jensen's inequality

$$f\left(\sum_{j=1}^{n} p_j x_j\right) = e^{\sum_{j=1}^{n} p_j x_j} \leq \sum_{j=1}^{n} p_j f(x_j) = \sum_{j=1}^{n} p_j e^{x_j}$$

If for all j we set $c_j = e^{x_j}, p_j = \frac{1}{n}$, then we get the Geometric-Mean – Arithmetic-Mean (GM-AM) inequality for nonnegative numbers :

$$\left(\prod_{j=1}^{n} c_j\right)^{1/n} \leq \frac{1}{n}\sum_{j=1}^{n} c_j$$

Examples

- $C = (0, \infty), f(x) = -\ln(x)$

$\frac{d^2 f(x)}{dx^2} = \frac{1}{x^2} > 0, \forall x \in C \Rightarrow f$ is (strictly) convex over C
$\Rightarrow$ By Jensen's inequality, for any two pmfs $p_1, \dots, p_n$ and $q_1, \dots, q_n$ over n values we have

$$f\left(\sum_{j=1}^{n} p_j \frac{q_j}{p_j}\right) = f(1) = 0 \leq \sum_{j=1}^{n} p_j f\left(\frac{q_j}{p_j}\right) = \sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right)$$

The quantity $\sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right)$ is called the Kullack-Liebler (KL) divergence or relative entropy of pmf p with respect to pmf q and the above inequality shows that it is always nonnegative.

Convex optimization

- An optimization problem of the form

$$\min_{\boldsymbol{x} \in C} f(\boldsymbol{x})$$

in which:

1. the constraint set C over which a function $f(\boldsymbol{x})$ is being minimized is a convex set and
2. the objective function $f(\boldsymbol{x})$ that is being minimized is a convex function over C

is called a convex optimization problem. Here,

- if C is a closed set then a minimizer $\boldsymbol{x} \in C$ is guaranteed to exist
- if C is a closed and f is strictly convex over C then the minimizer exists and is unique

Convex optimization

- Let $f, g_1, \dots, g_n$ be real-valued convex functions over a convex subset $C \subseteq \mathbb{R}^d$
- Then the optimization problem:

$$\min_{x \in C} f(\boldsymbol{x})$$

$$\text{subject to: } g_j(\boldsymbol{x}) \leq 0, j \in [1, n]$$

is a convex optimization problem. This is also called the primal optimization problem (even if it is non-convex).

- Lagrange function or Lagrangian associated to the primal optimization problem is defined as:

$$\forall \boldsymbol{x} \in C, \forall \lambda_1, \dots, \lambda_n \geq 0\,,$$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) + \sum_{j=1}^{n} \lambda_j g_j(\boldsymbol{x})$$

- Here, $\lambda_1, \dots, \lambda_n$ are called Lagrange or dual variables or multipliers

Slater's condition or weak constraint qualification

- A set C together with functions $g_1, \dots, g_n$ are said to satisfy Slater's condition or weak constraint qualification if
  1. there exists a point $\overline{\boldsymbol{x}}$ in the interior of C such that
  2. for all $j \in [1, n]$, either
     - $g_j(\overline{\boldsymbol{x}}) < 0$ or
     - $g_j(\overline{\boldsymbol{x}}) = 0$ and g_j is an affine function

Karush-Kuhn-Tucker Theorem

- Let $f, g_1, \ldots, g_n$ be real-valued convex and differentiable functions over a convex subset $C \subseteq \mathbb{R}^d$ satisfying Slater's condition. Then $\overline{\boldsymbol{x}}$ is a solution of the primal optimization problem if, and only if, the following conditions hold:
  - Primal feasibility conditions:
    $$\overline{\boldsymbol{x}} \in C \text{ and } \forall j \in [1, n], g_j(\overline{\boldsymbol{x}}) \leq 0$$
  - Stationarity of Lagrangian: there exist $\lambda_1, \ldots, \lambda_n \geq 0$ such that
    $$\nabla f(\overline{\boldsymbol{x}}) + \sum_{j=1}^{n} \lambda_j \nabla g_j(\overline{\boldsymbol{x}}) = 0,$$
  - Complementary slackness conditions:
    $$\forall j \in [1, n], \lambda_j g_j(\overline{\boldsymbol{x}}) = 0$$

Dual optimization problem

- The Lagrange dual function associated to the primal optimization problem is defined by:

$$F(\lambda_1, \dots, \lambda_n) = \inf_{\boldsymbol{x} \in C} L(\boldsymbol{x}, \boldsymbol{\lambda})$$
$$= \inf_{\boldsymbol{x} \in C} \left(f(\boldsymbol{x}) + \sum_{j=1}^{n} \lambda_j g_j(\boldsymbol{x}) \right)$$

for all $\lambda_1, \dots, \lambda_n \geq 0$.

- If $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_n)^T$, then $F(\boldsymbol{\lambda})$ is a concave function of $\boldsymbol{\lambda}$
- The optimization problem:

$$\max_{\boldsymbol{\lambda} \geq \boldsymbol{0}} F(\boldsymbol{\lambda})$$

is called the dual optimization problem associated to the primal optimization problem

Dual optimization problem

- If the conditions of the Karush-Kuhn-Tucker theorem are satisfied, then

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}} F(\boldsymbol{\lambda}) \quad = \quad \min_{\boldsymbol{x} \in C} f(\boldsymbol{x})$$

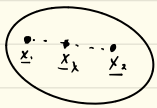
subject to: $g_j(\boldsymbol{x}) \leq 0, j \in [1, n]$

- Remarks: even if the primal problem is non-convex,
  - the dual function is concave
  - the dual function is never above the primal minimum
  - thus the dual maximum is never above the primal minimum
  - duality gap = (primal minimum) – (dual maximum) ≥ 0
  - for a convex primal problem that is "regular" (Salter's conditions), the duality gap is zero

Convex Sets $C \subseteq \mathbb{R}^d$

is convex if
for all $\underline{x}_1, \underline{x}_2 \in C$
for all $\lambda \in [0,1]$

$$\underline{x}_\lambda = (1-\lambda)\underline{x}_1 + \lambda \underline{x}_2 \in C$$



Convex function $f: \mathbb{R}^d \to \mathbb{R}$

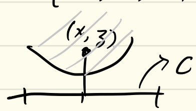
$C \subseteq \mathbb{R}^d$ is convex

f is a convex function over the convex set C if

1) Epigraph characterization

f is convex over C if

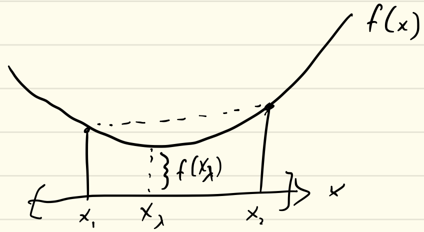
$$epi(f) := \{(x, z) : x \in C, z \geq f(x)\}$$



2) Chord Characterization

$f: \mathbb{R}^d \to \mathbb{R}$ is convex over the convex set $C \subseteq \mathbb{R}^d$ if
for all $\underline{x}_1, x_2 \in C$
for all $\lambda \in [0,1]$

$$f(x_\lambda) \leq (1-\lambda) f(x_1) + \lambda f(x_2)$$

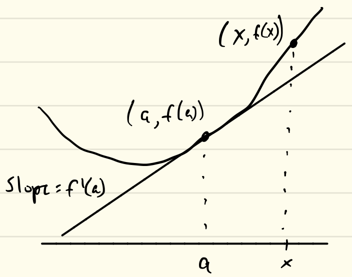


Tangent Characterization

$f : \mathbb{R}^d \to \mathbb{R}$, $C \subseteq \mathbb{R}^d$ convex

f differentiable over C

"tangent always below the graph"



tangent line

$$f(a) + (x-a) f'(a) \le f(x) \quad \forall a, x \in C$$

(scalar case)

$$f(\underline{a}) + (\underline{x} - \underline{a})^T \nabla f(\underline{a}) \le f(\underline{x})$$

(multiple dimensions)

2nd derivative characterization

$f : \mathbb{R}^d \to \mathbb{R}$, $C \subseteq \mathbb{R}^d$ convex

f is twice differentiable over C

"2nd derivative never negative"

$$\forall x \in C, \underbrace{\nabla^2 f(x)}_{\substack{\text{Hessian mtx} \\ d \times d}} \succeq 0$$

Positive semi-definite

$$\nabla^2 f(x) = \begin{array}{cc} & j \\ i & \left[\begin{array}{ccc} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} & & \\ & & \\ & & \end{array} \right]_{d \times d} \end{array}$$

Concave function $f: \mathbb{R}^d \to \mathbb{R}$

$C \subseteq \mathbb{R}^d$ is convex

f is a concave function over the convex set C if $-f$ is convex

Strictly Convex fn

1) f is convex
2) $=$ in chordal definition holds iff. $\begin{cases} \underline{x}_1 = \underline{x}_2 \text{ or} \\ \lambda = 0 \text{ or} \\ \lambda = 1 \end{cases}$

"No planar segment in graph of function"

Global / local minimum

If f is convex over C (& closed)
then it has a global minimum at some point in C
i.e. $\exists\, x_0 \in C$:
$\forall\, x \in C$
$f(x_0) \leq f(x)$

Steepest (gradient) descent

max of convex functions is convex

Initialization $x_0 \in C$

for $t = 1$ till t_{max} or convergence

$$x_t = x_{t-1} - \underbrace{\eta_t}_{\text{step size}} \nabla f(x_{t-1})$$

$g(x \in \mathbb{R})$ convex, non-decreasing

$f(x) \rightarrow$ convex

$g(f(x)) = h(x) \leftarrow$ convex

x_1, x_2, x_3

$$x_\lambda = (1-\lambda) x_1 + \lambda x_2$$

$$h(x_\lambda) = g(f(x_\lambda))$$

$$f(x_\lambda) \le (1-\lambda) f(x_1) + \lambda f(x_2)$$

$$g(f(x_1)) \le g((1-\lambda) f(x_1) + \lambda f(x_2))$$

$$\le (1-\lambda) g(f(x_1)) + \lambda g(f(x_2))$$

$$= (1-\lambda) h(x_1) + \lambda h(x_2)$$

$f(x) = x^2$

$f''(x) = 2 > 0$ f is strictly convex

$g(x) = e^x$

$g''(x) = e^x \ge 0 \ \forall x$

$g(f(x)) = e^{x^2} = h(x)$ is convex

$f(x) = \ln x$ So $-f(x)$ is strictly convex

$f'(x) = \frac{1}{x}$ $f(x)$ is " concave

$f''(x) = -\frac{1}{x^2} < 0$

$f(x) = -\ln x \leftarrow$ convex

Tangent Ch.

$$f(1) + (x-1) f'(1) \le f(x)$$

$$\underset{0}{\downarrow} \qquad \downarrow$$

$$(x-1)(-1) \le -\ln x$$

$$\Rightarrow \ln x \le x-1$$

Jensen's Inequality for convex fns $\forall$ Random vectors $\underline{x} \in C \subseteq \mathbb{R}^d$

If f is convex over C then

$f(E[x]) \le E[f(x)]$

$f(x) = -\ln x \longleftarrow$ Strictly convex

$x_1 = \frac{q_1}{p_1}, \quad x_2 = \frac{q_2}{p_2}, \ldots, \quad x_n = \frac{q_n}{p_n}$

$p_1 \cdots p_n \to$ PMF over n things

$q_1 \cdots q_n \to$ PMF over n things

$X =$ random variable, $P(X = x_i) = p_i$

By Jensen's inequality,

$$f(E[x]) \le E[f(x)]$$

$$= \sum_{i=1}^{n} \frac{q_i}{\cancel{p_i}} \cancel{p_i} \qquad \underset{0}{\underset{\shortparallel}{f(1)}} \le \sum_{i=1}^{n} p_i f(x_i)$$

$$= 1 \qquad\qquad = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right)$$