EE445 Mod4-Lec2: Convex Optimization

References: [Optimization Models: Calafiore & El Ghaoui] Chapter 8

Topics for Module 4

HW5 – start early

→ HW6 – extra credit – included in final exam

- Lec1: Convexity and Convex Sets
- Lec2: Convex Functions, Smooth Unconstrained Minimization & Gradient Descent
- Lec3: Convex Optimization Problems: ML models I
- Lec4: Convex Optimization Problems: ML models II

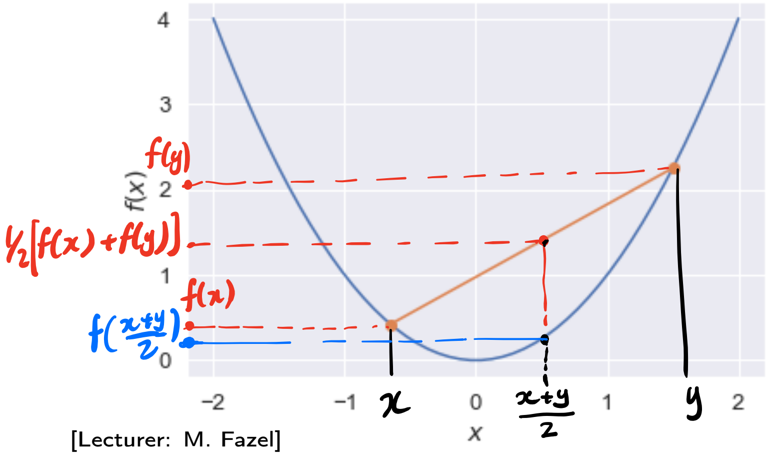
Convex functions

From last lecture: $f : \mathbf{R}^n \mapsto \mathbf{R}$ is a **convex function** if

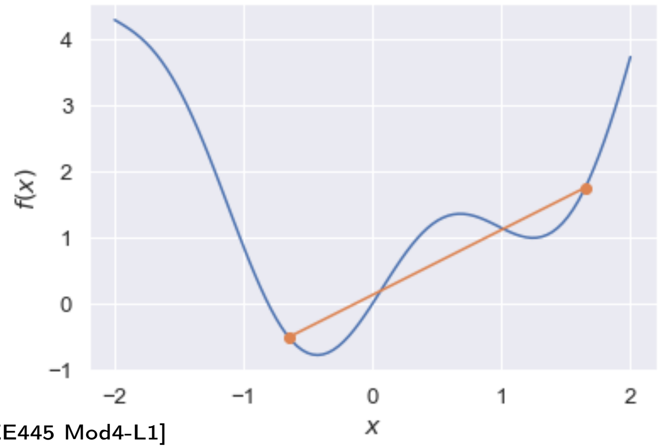
$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda) f(y)$$

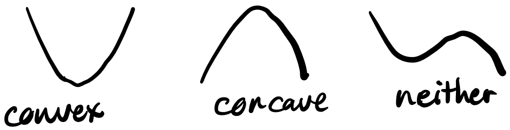
for all $x, y \in \mathbf{R}^n$ and all $0 \leq \lambda \leq 1$.

Convex

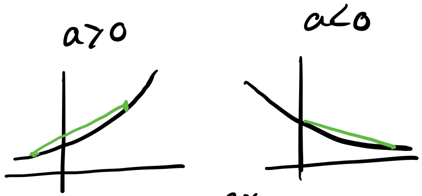


Nonconvex



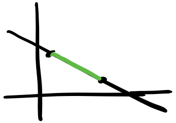


Examples on $\mathbf{R}$



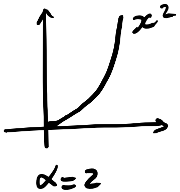
$f'(x) = ae^{ax}$
$f''(x) = a^2 e^{ax} > 0$

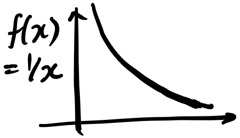
- f is called **concave** if $-f$ is convex

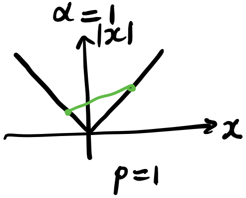


convex:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$ $x > 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$

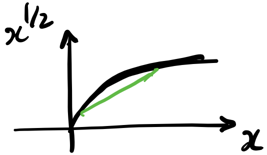


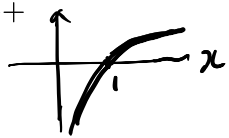




concave:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$





Examples on $\mathbf{R}^n$

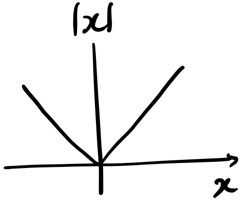
$f : \mathbb{R}^n \to \mathbb{R}^m$ $f(x) = Ax + b$

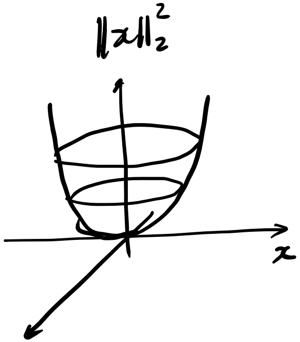
affine functions are both convex and concave:

- affine function $f : \mathbf{R}^n \mapsto \mathbf{R}$, $f(x) = a^T x + b$

all norms are convex, e.g.,

- ℓ_p norms: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$
- ∞-norm: $\|x\|_\infty = \max_k |x_k|$





First-order convexity condition

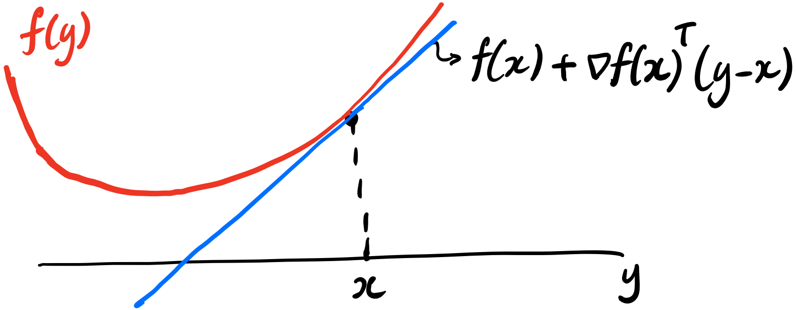
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

exists at each $x \in \mathbf{dom}\, f$

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$



underestimator

Taylor expansion around x:

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

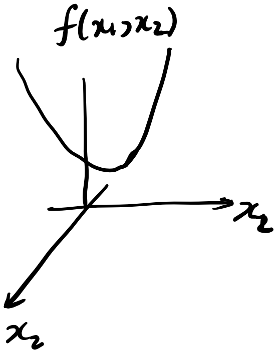
1st order (linear) approximation

Second-order convexity condition

f is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n,$$

exists at each $x \in \mathbf{dom}\, f$

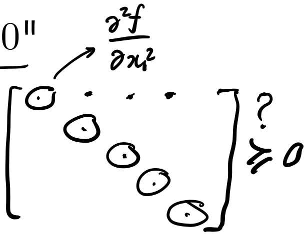


2nd-order conditions: for twice differentiable f with convex domain

- f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

- in 1D: means $f''(x) \geq 0$ for all $x \in \mathbf{dom}\, f$
- note the distinction between: $\nabla^2 f(x) \succeq 0$ versus "diag entries ≥ 0"



Examples

quadratic function: $f(x) = (1/2)x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$) $x \in \mathbb{R}^n$

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2$

$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

convex (for any A)

quadratic-over-linear: $f(x, y) = x^2/y$

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

convex for $y > 0$

Examples

quadratic function: $f(x) = (1/2)x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2 = (Ax-b)^T(Ax-b)$

$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A \succeq 0$$

convex (for any A)

quadratic-over-linear: $f(x, y) = x^2/y$

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

convex for $y > 0$

Examples

quadratic function: $f(x) = (1/2)x^T P x + q^T x + r$ (with $P \in \mathbf{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

convex if $P \succeq 0$

least-squares objective: $f(x) = \|Ax - b\|_2^2$

$$\nabla f(x) = 2A^T(Ax - b), \qquad \nabla^2 f(x) = 2A^T A$$

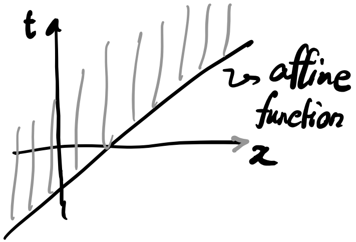
convex (for any A)

quadratic-over-linear: $\underline{f(x,y) = x^2/y}$ (See Friday 5/20 session for example details)

$$\underbrace{\nabla^2 f(x,y)} = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

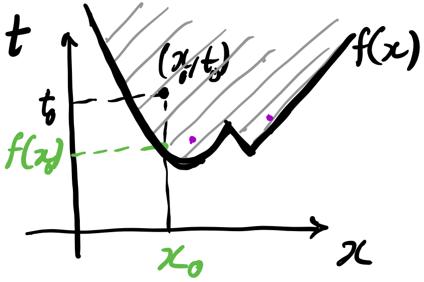
convex for $y > 0$

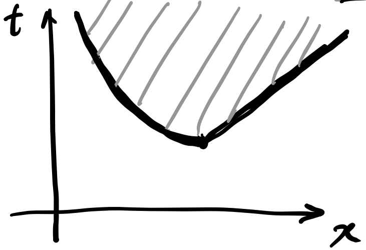
Epigraph of a function



epigraph of $f : \mathbf{R}^n \to \mathbf{R}$:

$$\mathbf{epi}\, f = \{(x,t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\, f,\ f(x) \le t\}$$





this notion connects the definitions of convex functions with convex sets:

function f is convex if and only if $\mathbf{epi}\, f$ is a convex set

Proof of: if f is convex function, then the set $\mathbf{epi}\, f$ is a convex set (suppose $\mathbf{dom} f = \mathbb{R}^n$ for simplicity)

given: $(x_1, t_1) \in \mathbf{epi}\, f$ (i.e., $f(x_1) \le t_1$)
and $(x_2, t_2) \in \mathbf{epi}\, f$ (i.e., $f(x_2) \le t_2$)

show that: $\forall\ 0 \le \theta \le 1$: $(\theta x_1 + (1-\theta) x_2,\ \theta t_1 + (1-\theta) t_2) \in \mathbf{epi}\, f$

from def. of epigraph, this is the same as:

$$f(\theta x_1 + (1-\theta) x_2) \le \theta t_1 + (1-\theta) t_2$$

to show it, first use convexity of f:

$$f(\theta x_1 + (1-\theta) x_2) \le \theta \underbrace{f(x_1)}_{\le t_1} + (1-\theta) \underbrace{f(x_2)}_{\le t_2} \le \theta t_1 + (1-\theta) t_2$$

then use $(x_1, t_1), (x_2, t_2) \in \mathbf{epi}\, f$

DONE

Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify the definition: show for all $x, y \in \mathbf{dom}\, f$ and all $0 \leq \lambda \leq 1$,

$$f\Big(\lambda x + (1-\lambda)y\Big) \leq \lambda f(x) + (1-\lambda)f(y)$$

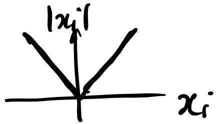
2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$ → positive semidefinite
3. show that f is obtained from simple convex functions by operations that preserve convexity:
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum → • partial minimization
   - ~~more general composition~~

1. Positive weighted sum

nonnegative multiple: αf is convex if f is convex, $\alpha \geq 0$

sum: $f_1 + f_2$ convex if f_1, f_2 convex

Q: how about $f_1 - f_2$? no, in general



- $f(x) = \underbrace{\|x\|_2} + 3\underbrace{\|x\|_1}$ $\quad \sum_{i=1}^{n} |x_i|$

  $\Rightarrow$ convex

- $f(x) = \sum_{i=1}^{n} e^{x_i}$

  $\Rightarrow$ convex

(this extends to infinite sums, integrals)

2. Composition with an affine function

$\mathbb{R}^n \to \mathbb{R}^m$

Consider the affine function $x \mapsto Ax + b$, with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, then the function $g(x) = f(Ax + b)$ is a convex function if f is convex

examples

- (any) norm of an affine function: $g(x) = \|Ax + b\|$

$x \to Ax+b$

$x_i \to a_i^T x + b_i$

$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$

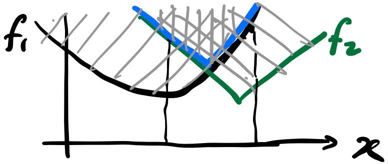
- $f(x) = \sum_{i=1}^{n} e^{x_i}$

$g(x) = f(Ax+b) = \sum_{i=1}^{n} e^{(a_i^T x + b_i)} \Rightarrow$ convex

3. Pointwise maximum

If $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

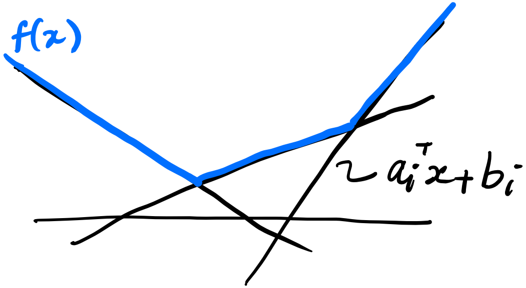
note: this is maximum is taken pointwise, meaning for every x, look at the value of $f_1(x)$,...,$f_m(x)$ and take the largest of them (at that x)



examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ is convex

$f(x) = \max\{f_1(x), f_2(x)\}$

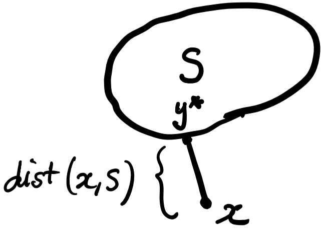


4. Partial minimization

if $f(x,y)$ is convex in (x,y) (note that this means jointly convex in the variables) and C is a convex set, then

$$g(x) = \min_{y \in C} f(x,y)$$

is also convex



example

distance from a point x to a set S: $x \in \mathbb{R}^n$, $S \subset \mathbb{R}^n$

$$f_0(x) = \mathbf{dist}(x,S) = \min_{y \in S} \|x - y\|$$

is convex if S is convex

proof: first check if $f(x,y) = \|x-y\|$ is convex (jointly) in (x,y):

$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \underbrace{[I \ \ -I]}_{A} \begin{bmatrix} x \\ y \end{bmatrix}$ is a linear map, and $\|x-y\| = \|A\begin{bmatrix} x \\ y \end{bmatrix}\|$ is composition of linear map $A\begin{bmatrix} x \\ y \end{bmatrix}$ with convex function $\|\cdot\|$, so f is convex in (x,y).

More examples

- show that the following $f : \mathbb{R}^n \to \mathbb{R}$ is convex:

$$f(x) = \max_{i=1,\dots,k} \{ \|A^{(i)}x - b^{(i)}\| \}$$ where $A^{(i)} \in \mathbb{R}^{m \times n}$, $b^{(i)} \in \mathbb{R}^m$, and $\|\cdot\|$ is a norm

<u>proof</u>: for each i, $g_i(x) = \|A^{(i)}x - b^{(i)}\|$ is convex, because norm $\|x\|$ is a convex function and $g(x)$ is its <u>composition with affine map</u> $x \to A^{(i)}x - b^{(i)}$.

Then $f(x)$ is convex because it is the <u>pointwise max</u> of convex functions $g_i(x)$, $i = 1, \dots, k$.

Minimizing convex functions: Basic solution methods

Very few optimization problems have a closed-form solution (e.g., least-squares); most problems are solved using iterative methods.

One important **iterative method** is **gradient descent** (for unconstrained minimization of a differentiable, convex f):

given a starting point x^0, run the following iterations for $k = 1, 2, \ldots,$

$$x^{k+1} = x^k - \underbrace{\alpha}_{\text{step size}} \nabla f\left(x^k\right)$$

$$x^1 = x^0 - \alpha \nabla f(x^0)$$

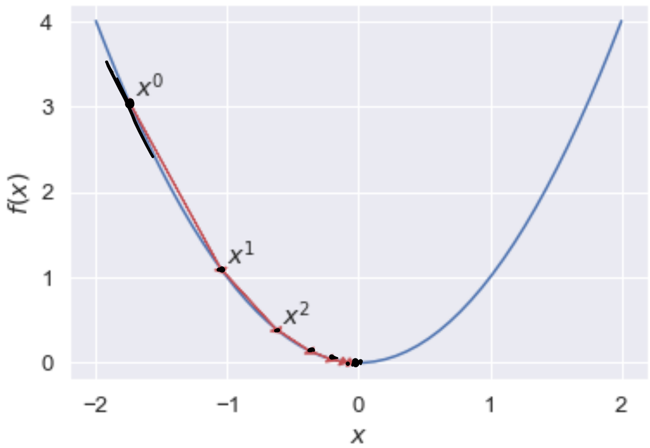
$$x^2 = x^1 - \alpha \nabla f(x^1)$$

$$\vdots$$

Convex Optimization: Basic Solution Methods

(see Python notebook & TA session on 5/20)

$$x^{k+1} = x^k - \alpha \nabla f\left(x^k\right)$$



$f(x_1, x_2)$