EE445 Mod4-Lec2: Convex Optimization

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

Topics for Module 4

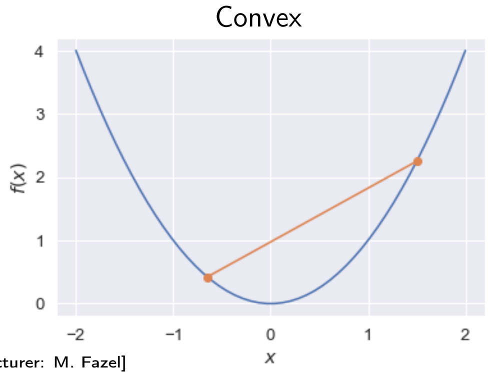
- Lec1: Convexity and Convex Sets
- Lec2: Convex Functions, Smooth Unconstrained Minimization & Gradient Descent
- Lec3: Convex Optimization Problems: ML models I
- Lec4: Convex Optimization Problems: ML models II

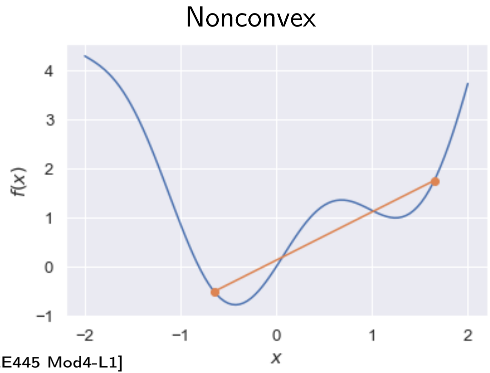
Convex functions

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 $0 \leq \lambda \leq 1$.





Examples on $\mathbf{R}$

- 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 $\alpha \leq 0$
- 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 $\mathbf{R}^n$

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

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: $f(x) = (1/2)x^T Px + q^T x + r$ (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: $f(x, y) = x^2/y$

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

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: $f(x, y) = x^2/y$

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

convex for $y > 0$

Examples

quadratic function: $f(x) = (1/2)x^T Px + q^T x + r$ (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: $f(x, y) = x^2/y$

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

convex for $y > 0$

Epigraph of a function

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

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

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

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 $0 \le \lambda \le 1$,

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

2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
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

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

(this extends to infinite sums, integrals)

2. Composition with an affine function

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: $g(x) = \|Ax + b\|$

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

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:

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

is convex if S is convex

More examples

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 $k = 1, 2, \ldots$,

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

Minimizing convex functions: Basic Solution Methods

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