Lecture 6: Convex Functions

goals:
- intro to convex functions
- many examples
- 1^{st} and 2^{nd} order convexity characterization
- operations that preserve convexity
- quasiconvexity

Definition A function $f: \mathbb{R}^n \to \mathbb{R}$ is called convex if its domain $\text{dom}(f)$ is a convex set and for all $x, y \in \text{dom}(f)$ we have

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y) \quad \forall \theta \in [0,1]$$

Definition A function f is concave if $-f$ is convex.

A function is strictly convex if the inequality holds strictly.

A function is strongly convex if $\exists \alpha > 0$ such that $f(x) - \alpha \|x\|_2^2$ is convex

FACT: Strong convexity $\Longrightarrow$ strict convexity $\Longrightarrow$ convexity

- Examples on $\mathbb{R}$

<u>Convex:</u>

- affine : $ax + b$, $\forall a, b \in \mathbb{R}$
- exponential : e^{ax} , $\forall a \in \mathbb{R}$
- powers : x^{α} on dom $\mathbb{R}_{++}$, $\forall \alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value : $|x|^p$, $\forall p \geq 1$
- negative entropy : $x \log x$ on $\mathbb{R}_{++}$

<u>Concave:</u>

- affine : $ax + b$, $\forall a, b \in \mathbb{R}$
- powers : x^{α} on $\mathbb{R}_{++}$, $\forall \alpha \in [0, 1]$
- logarithm : $\log x$ on $\mathbb{R}_{++}$

- Examples on $\mathbb{R}^n$
  - affine : $a^T x + b$, $\forall a \in \mathbb{R}^n$, $b \in \mathbb{R}$
  - quadratic : $x^T P x + q^T x + d$
    - convex iff $P \succeq 0$
    - strictly/strongly convex iff $P \succ 0$
    - concave iff $P \preceq 0$
  - any norm : $\|x\|_p$ $\forall p \geq 1$

<u>Proof:</u> $\forall \theta \in [0, 1]$, $\forall x, y \in \mathbb{R}^n$

$$f(\theta x + (1-\theta) y) \underset{\text{triangle inequality}}{\leq} f(\theta x) + f((1-\theta) y) \underset{\text{homogeneity}}{=} \theta f(x) + (1-\theta) f(y)$$

- Examples on $\mathbb{R}^{m \times n}$
    - affine: $\operatorname{trace}(A^T X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$
    - any matrix norm:
        - $\|X\|_2 = \sigma_{max}(X) = \sqrt{\lambda_{max}(X^T X)}$
        - $\|X\|_* = \sum_{i=1}^{\operatorname{rank}(X)} \sigma_i(X)$
    - $\log \det(X)$ (concave)

Restriction to a Line

Theorem A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex iff its evaluation along any line in its domain is convex

i.e.

$$g(t) = f(x + tv) \quad \text{is convex in } t$$

$\forall x \in \operatorname{dom}(f)$, $v \in \mathbb{R}^n$, where $\operatorname{dom}(g) = \{t \mid x + tv \in \operatorname{dom}(f)\}$

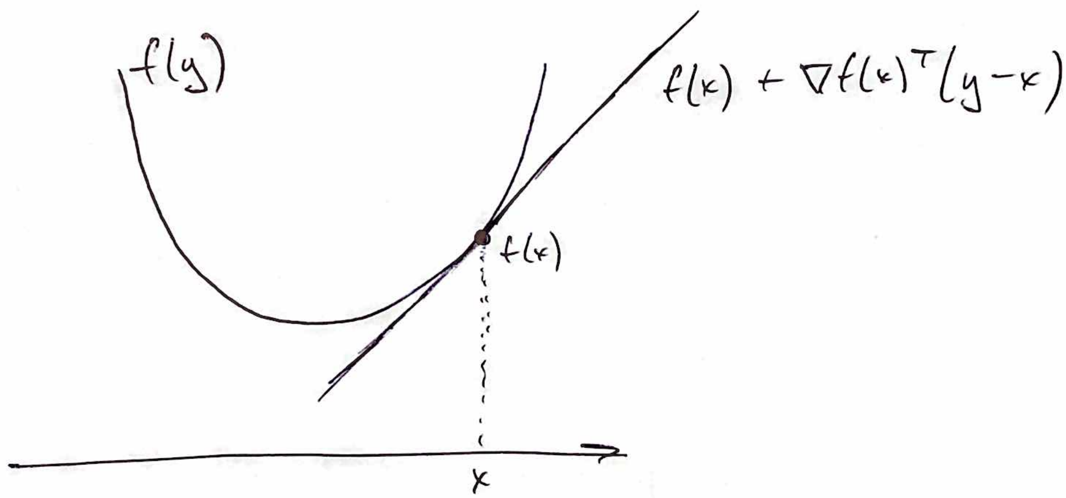
- i.e. convexity of f can be tested by checking functions of only one variable (though an infinite number)
- simplifies many proofs in convex analysis

First-Order Characterization of Convexity

Theorem: A differentiable function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with an open convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) \quad \forall x,y \in \mathrm{dom}(f)$$

Interpretation: The first-order Taylor expansion of f at any point is a global underestimator of f



Second-Order Characterization of Convexity

Theorem: A twice differentiable function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ with an open convex domain is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathrm{dom}(f)$$

i.e. it has nonnegative curvature everywhere

- If $\nabla^2 f(x) \succ 0 \quad \forall x \in \mathrm{dom}(f)$, then f is strictly convex
- Strict convexity ensures uniqueness of solutions

Corollary Consider an unconstrained optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

where f is convex and differentiable. Than any point $\bar{x}$ that satisfies $\nabla f(\bar{x}) = 0$ is a global minimum.

Proof: From the 1st-order convexity condition, we have

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \forall x, y$$

In particular,

$$f(y) \geq f(\bar{x}) + \nabla f(\bar{x})^T (y - x) \quad \forall y$$

Since $\nabla f(\bar{x}) = 0$, we have $f(y) \geq f(\bar{x}) \quad \forall y$ ∎

- With convexity $\nabla f(\bar{x}) = 0 \iff \bar{x}$ a global min
- Without convexity $\nabla f(\bar{x}) = 0$ not even sufficient for local min

---

Jensen's Inequality

Recall: f convex : $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y) \quad \forall \theta \in [0,1]$

More generally: f convex : $f(E z) \leq E f(z)$ (E is expectation)
for any random variable z

- basic inequality a special case with discrete distribution
  $\text{prob}(z = x) = \theta$, $\text{prob}(z = y) = 1 - \theta$

- Examples of 1^{st} + 2^{nd} order conditions
  - quadratic: $f(x) = \frac{1}{2}x^T P x + q^T x + r \quad P \in S^n$

    $\nabla^2 f(x) = P$, f convex iff $P \succeq 0$
  - least squares: $f(x) = \|Ax - b\|_2^2$

    $\nabla^2 f(x) = 2A^T A \succeq 0$ always convex
  - quadratic-over-linear: $f(x,y) = \frac{x^2}{y}$

    $\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$ convex for $y > 0$
  - log-sum-exp: $f(x) = \log \sum_{k=1}^{n} e^{x_k}$

    $\nabla^2 f(x) = \frac{1}{1^T z}\operatorname{diag}(z) - \frac{1}{(1^T z)^2} z z^T$, $z = e^{x_k}$

    not obvious, but can show $\nabla^2 f(x) \succeq 0$

    so f is convex
  - geometric mean: $f(x) = \left(\prod_{k=1}^{n} x_k\right)^{\frac{1}{n}}$ on $\mathbb{R}^n_{++}$

    is concave

Operations that Preserve Convexity

- How to prove convexity of a function
  1. use definition (often restricting to a line)
  2. use 1^{st} or 2^{nd} order conditions
     (usually showing $\nabla^2 f(x) \succeq 0$ for twice diff. f)
  3. show f is obtained from simple convex functions by operations that preserve convexity

- <u>Nonnegative weighted sum</u>
  - f convex $\Rightarrow$ αf convex $\forall \alpha \geq 0$
  - $f_1, \ldots, f_m$ convex $\Rightarrow$ $\alpha_1 f_1 + \cdots + \alpha_m f_m$ convex $\forall \alpha_i \geq 0$
    - extends to infinite sums and integrals

- <u>Composition with an affine function</u>
  - f convex $\Rightarrow$ $f(Ax+b)$ convex

- Examples:

  - $f(x) = \|Ax+b\|$ for <u>any</u> norm

  - <u>log barrier</u> for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

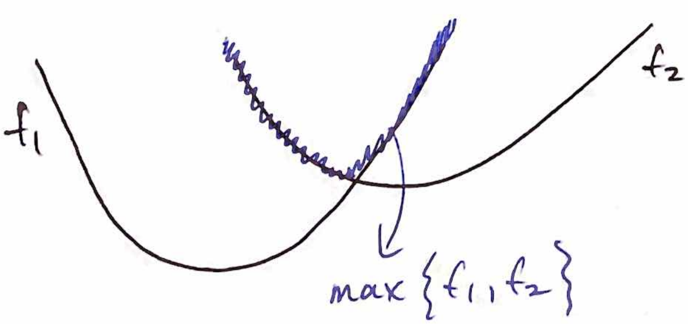
$$\mathrm{dom}(f) = \{x \mid a_i^T x < b_i,\ i=1,\dots,m\}$$

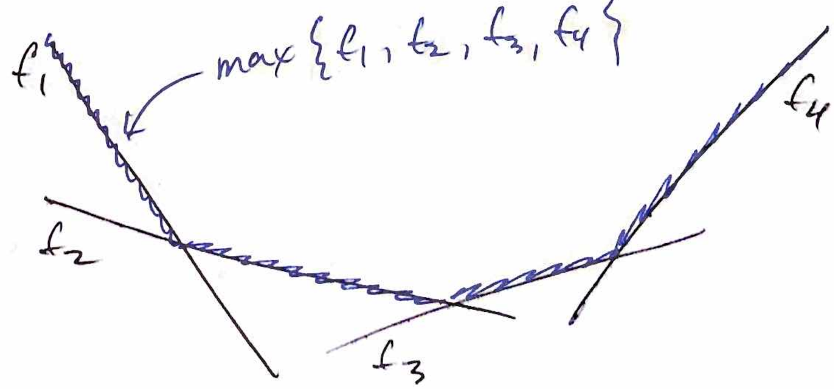
  - $f(x_1, x_2) = (x_1 - 2x_2)^4 + 2e^{3x_1 + 2x_2 - 5}$

* <u>NOT</u> obvious at first glance, but straightforward using convexity preserving operations

- <u>Pointwise Maximum</u>

  - $f_1, \dots, f_m$ convex $\implies f(x) = \max\{f_1(x), \dots, f_m(x)\}$ convex





- similarly, pointwise min of concave is concave
- pointwise min of convex <u>not</u> convex in general

- **Pointwise Supremum**

  - $f(x,y)$ convex in x for each $y \in \mathcal{Y}$

    $$\Longrightarrow \quad g(x) = \sup_{y \in \mathcal{Y}} f(x,y) \quad \text{convex}$$

  - Examples:

    - support function of a set $C \subseteq \mathbb{R}^n$:

      $$S_C(x) = \sup_{y \in C} y^T x$$

    - distance to farthest point in a set $C \subseteq \mathbb{R}^n$:

      $$f(x) = \sup_{y \in C} \|x - y\| \quad (\text{any norm})$$

    - max eigenvalue of a symmetric matrix $X \in S^n$:

      $$\lambda_{max}(X) = \sup_{\|y\|_2 = 1} y^T X y$$

- **Parametric/Partial Minimization**

  - $f(x,y)$ convex in (x,y) and C a convex set

    $$\Longrightarrow \quad g(x) = \inf_{y \in C} f(x,y) \quad \text{convex}$$

- Examples

  - $f(x,y) = x^T A x + 2x^T B y + y^T C y$

    with $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \quad C \succ 0$

    $$g(x) = \inf_y f(x,y) = x^T \underbrace{\left(A - BC^{-1}B^T\right)}_{\text{called \underline{Schur} \underline{complement}}} x \quad \text{convex}$$

  - distance to a convex set:

    $$\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$$

- <u>Scalar</u> <u>Composition</u>

  - Let $g: \mathbb{R}^n \to \mathbb{R}$, $h: \mathbb{R} \to \mathbb{R}$, $f(x) = h(g(x))$

    f convex if:

    ① g convex, h convex, $\tilde{h}$ nondecreasing

    ② g concave, h convex, $\tilde{h}$ nonincreasing

    <u>Proof</u>: ($n = 1$, differentiable f, g)

    $$f''(x) = h''(g(x))\, g'(x)^2 + h'(g(x))\, g''(x)$$

    (can prove <u>directly</u> in general)

- Extended Value Extension
  - a function f not defined everywhere (i.e. $\mathrm{dom}(f) \subset \mathbb{R}^n$) can be extended by defining

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathrm{dom}(f) \\ +\infty & x \notin \mathrm{dom}(f) \end{cases}$$

  - can simplify notation and preserves epigraph

- Examples
  - $e^{g(x)}$ convex if $g(x)$ convex
  - $\frac{1}{g(x)}$ convex if $g(x)$ concave and positive

Vector Composition

- Let $g: \mathbb{R}^n \to \mathbb{R}^k$, $h: \mathbb{R}^k \to \mathbb{R}$,

$$f(x) = h(g(x)) = h(g_1(x), \dots, g_k(x))$$

f convex if

① g_i convex, h convex, $\tilde{h}$ nondecreasing in each argument

② g_i concave, h convex, $\tilde{h}$ " " " "

Proof: ($n=1$, diff f, g)

$$f''(x) = g'(x)^T \nabla h(g(x))\, g'(x) + \nabla h(g(x))^T g''(x)$$

- Examples
  - $\sum_{i=1}^{m} \log(g_i(x))$ concave if g_i concave and positive
  - $\log \sum_{i=1}^{m} e^{g_i(x)}$ convex if g_i convex

Perspective

- the perspective of $f: \mathbb{R}^n \to \mathbb{R}$ is $g: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$

$$g(x,t) = t f\left(\frac{1}{t} x\right) \qquad \mathrm{dom}(g) = \left\{(x,t) \,\middle|\, \frac{1}{t}x \in \mathrm{dom}(f),\ t > 0\right\}$$

g convex if f convex

- Examples $\underbrace{f(x) = x^T x}_{\text{(convex)}} \Rightarrow g(x,t) = \frac{1}{t} x^T x$ convex ($t > 0$)

Quasiconvexity

- Recall the sublevel sets of a function $f: \mathbb{R}^n \to \mathbb{R}$

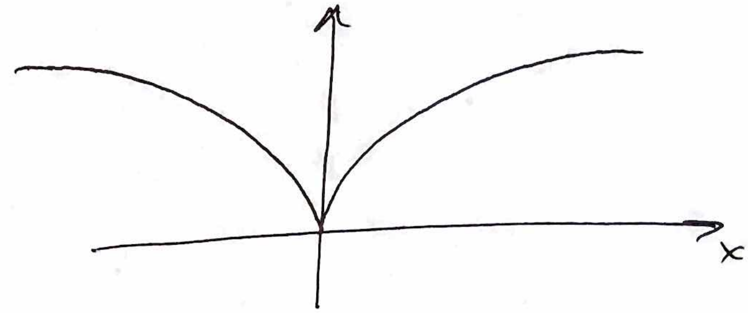
$$S_\alpha = \{ x \in \mathrm{dom}(f) \mid f(x) \leq \alpha \}$$

- Sublevel sets of convex functions are convex for any value of α (proof immediate from convexity defn)

- However, the converse is not true:

  a function can have all sublevel sets convex, but not be a convex function

  Ex $f(x) = \sqrt{|x|}$



Definition A function $f: \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex if all its sublevel sets are convex.

- f quasiconcave if $-f$ quasiconvex
- f quasilinear if both quasiconvex and quasiconcave

- Examples
  - $\mathrm{ceil}(x) = \inf\{z \in \mathbb{Z} \mid z \geq x\}$ quasilinear
  - $\log x$ quasilinear on $\mathbb{R}_{++}$
  - $f(x_1, x_2) = x_1 x_2$ quasiconcave on $\mathbb{R}^2_{++}$
  - linear fractional: $f(x) = \dfrac{a^T x + b}{c^T x + d}$ $\quad c^T x + d > 0$ quasilinear
  - distance ratio: $f(x) = \dfrac{\|x-a\|_2}{\|x-b\|_2}$ $\quad \|x-a\|_2 \leq \|x-b\|_2$ quasiconvex

- Quasiconvex optimization problems can be reduced to solving a (small) sequence of convex optimization problems