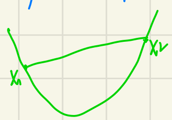
Lec 3: Convex functions

- Convex function

$f: \mathbb{R}^n \to \mathbb{R}$ is cvx if dom f is cvx and
$$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$$
for all $x, y \in$ dom f, $0 \le \theta \le 1$



f is concave $\Leftrightarrow$ $-f$ is cvx

f is strict cvx if dom f is cvx and $f(\theta x + (1-\theta) y) < \theta f(x) + (1-\theta) f(y)$ for $x \ne y$ $\theta \in [0,1]$

eg cvx:

- affine : $ax+b$ on $\mathbb{R}$
- exponential : e^{ax} for any $a \in \mathbb{R}$
- power : x^a on $\mathbb{R}_{++}$ for $a \ge 1$ or $a \le 0$
- negative entropy : $x \log x$ on $\mathbb{R}_{++}$

eg. concave:

- affine: $ax+b$ on $\mathbb{R}$
- powers : x^a on $\mathbb{R}_{++}$ for $0<a<1$
- log : $\log x$ on $\mathbb{R}_{++}$

· Egs on $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$

all norms are convex

egs on $\mathbb{R}^n$:

- affine: $f(x) = a^T x + b$
- norms: $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ for $p \ge 1$

egs on $\mathbb{R}^{m \times n}$

- affine: $f(X) = \underbrace{tr(A^T X)}_{\langle A, X \rangle} + b = \sum_i \sum_j A_{ij} X_{ij} + b$

- spectral norm: $f(X) = \|X\|_2 = \sigma_{max}(X) = \left(\lambda_{max}(X^T X)\right)^{1/2}$
  (max singular value)

· Line restriction

$f: \mathbb{R}^n \to \mathbb{R}$ is cvx

$\Updownarrow$

$g: \mathbb{R} \to \mathbb{R}$ $\quad g(t) = f(x+tv)$ $\quad \mathrm{dom}\, g = \{t \mid x+tv \in \mathrm{dom} f\}$

is cvx in t for any x, v

eg. $f: S^n \to \mathbb{R}$ $\quad f(X) = \log\det X$ $\quad \mathrm{dom} f = S^n_{++}$ $\qquad$ (V is symmetric but not has to be psd)

$$\begin{aligned} f(X+tV) &= \log\det(X+tV) \\ &= \log\det\left[X^{1/2}\left(I + t\cdot X^{-1/2} V X^{-1/2}\right) X^{1/2}\right] \\ &= \log\det X + \log\det\left(I + t\cdot X^{-1/2} V X^{-1/2}\right) \end{aligned}$$

$X^{-1/2}\cdot V\cdot X^{-1/2} = U\Lambda U^T$

$$= \log\det + \sum_i \log(1 + t\cdot\lambda_i)$$

λ_i is ith eigenvalue of $X^{-1/2}\cdot V\cdot X^{-1/2}$

concave in t

· Eexended-value extension

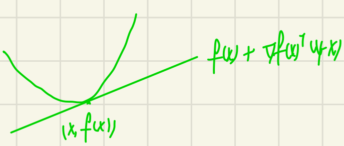
$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathrm{dom} f \\ +\infty & x \notin \mathrm{dom} f \end{cases}$$

· First-order condition

f is differentiable if $\mathrm{dom} f$ is open and gradient $\nabla f(x)$ at each $x \in \mathrm{dom} f$

differentiable f with cvx domain is cvx iff

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



from local information (gradient), get global conclusion (under estimator)

· Second-order condition

f is twice differentiable if dom f is open and the Hessian

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

exists at each $x \in$ dom f

twice differentiable f with cvx domain is cvx if $\nabla^2 f(x) \succeq 0$ for all $x \in$ dom f

Strictly cvx if $\nabla^2 f(x) \succ 0$

Eg.

1. quadratic function $f(x) = \frac{1}{2} x^T P x + q^T x + r \quad (P \in S^n)$

cvx iff $P \succeq 0$

2. least-square objective $\|Ax-b\|_2^2$

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

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

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

cvx if $y > 0$

sth to do with exponential family

4. log-sum-exp: $f(x) = \log \sum_i \exp(x_i)$ (Softmax is cvx)

$$\frac{\partial f}{\partial x} = \frac{e^{x_i}}{\sum e^{x_i}} \qquad \frac{\partial^2 f}{\partial x_i \partial x_i} = \frac{e^{x_i} \cdot \sum e^{x_i} - e^{x_i} \cdot e^{x_i}}{(\sum e^x)^2} = \frac{e^{x_i}}{\sum e^x} - \frac{e^{x_i} \cdot e^{x_i}}{(\sum e^x)^2}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{0 \cdot \sum e^{x_i} - e^{x_i} \cdot e^{x_j}}{(\sum e^x)^2} = 0 - \frac{e^{x_i} \cdot e^{x_j}}{(\sum e^x)^2}$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{1^T e^x} \cdot \mathrm{diag}(e^x) - \frac{1}{(1^T e^x)^2} \cdot (e^{x_i})(e^{x_j})^T$$

$\Downarrow \; z = e^x$

$$v^T \frac{\partial^2 f}{\partial x^2} v = \frac{1}{1^T z} \cdot v^T z v - \frac{1}{(1^T z)^2} v^T z \cdot z^T v$$

$$= \frac{1}{(\mathbf{1}^T z)^2} \cdot \left[\left(\sum v_i^2 z_i\right) \cdot \left(\sum z_i\right) - \left(\sum v_i \cdot z_i\right)^2 \right]$$

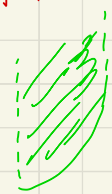
$$= \frac{1}{(\mathbf{1}^T z)^2} \cdot \left[\left(\sum (v_i \sqrt{z_i})^2\right) \cdot \left(\sum (\sqrt{z_i})^2\right) - \left(\sum v_i \cdot \sqrt{z_i} \cdot \sqrt{z_i}\right)^2 \right]$$

≥ 0 (cauchy-schwarz)

5. Geometric mean $f(x) = \left(\prod_{k=1}^{n} x_k\right)^{1/n}$ is concave on $\mathbb{R}^n_{++}$
   similar proof as for log-sum-exp

- sublevel set & Epigraph
  - α-sublevel set of $f: \mathbb{R}^n \to \mathbb{R}$
    $C_\alpha = \{x \mid f(x) \leq \alpha\}$
    sublevel sets of a cvx function is cvx (converse is false)

  - epigraph of $f: \mathbb{R}^n \to \mathbb{R}$
    $\text{epi } f = \{(x,t) \in \mathbb{R}^{n+1} \mid x \in \text{dom} f,\ f(x) \leq t\}$
    f is cvx iff epi$f$ is a cvx set



- Jensen's Inequality
  - if f is cvx, then
    $f(E[X]) \leq E[f(x)]$

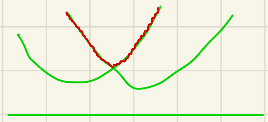
for cvx functions, wiggling hurts

Operations that preserve convexity

- non-negative multiple : αf is cvx if f is cvx ($\alpha \geq 0$)
- sum. $f_1 + f_2$ is cvx if f_1 f_2 are cvx
- Affine composition : $f(Ax+b)$ is cvx if x is cvx

- Egs.
  1. log-barrier for linear inequalities
  $$f(x) = \sum_{i=1}^{m} \log(b_i - a_i^T x) \qquad \operatorname{dom} f = \{x \mid a_i^T x < b_i \text{ for all } i\}$$
  2. Any norm of affine function $f(x) = \|Ax + b\|$

- Pointwise - maximum
  if $f_1 \cdots f_m$ are convex, then $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$ is cvx



$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{f_1(\theta x + (1-\theta)y),\ f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta f_1(x) + (1-\theta)f_1(y),\ \theta f_2(x) + (1-\theta)f_2(y)\} \\ &\leq \theta \cdot \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\} \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

intersection of 2 epigraphs

eg. piecewise-linear function : $f(x) = \max_i (a_i^T x + b_i)$ is cvx

eg. sum of r largest components of $x \in \mathbb{R}^n$
$$f(x) = x_{[1]} + \cdots x_{[r]}$$
$$= \max_i (a_i^T x) \quad \text{where } a_i \text{ is permutation vector}$$

- Pointwise supremum
  if $f(x,y)$ is cvx in x for each $y \in A$ then
  $$g(x) = \sup_{y \in A} f(x,y)$$
  is cvx

eg. maximum eigenvalue of symmetric matrix : $X \in S^n$

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

· Composition

composition of $g: \mathbb{R}^n \to \mathbb{R}$ and $h: \mathbb{R} \to \mathbb{R}$

$$f(x) = h(g(x))$$

$$f'(x) = h'(g(x)) \cdot g'(x)$$

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

$$f \text{ is cvx if } \begin{cases} g \text{ cvx},\ h \text{ cvx},\ h \text{ nondecreasing} \\ g \text{ concave},\ h \text{ cvx},\ h \text{ nonincreasing} \end{cases}$$

eg exp of cvx function is cvx

eg. $1/g(x)$ is cvx if g is concave and positive

· Vector Composition

composition of $g: \mathbb{R}^n \mapsto \mathbb{R}^k$ and $h: \mathbb{R}^k \to \mathbb{R}$

$$f(x) = h(g(x)) = h(g_1(x) \cdots g_k(x))$$

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

$$\begin{array}{c} \\ \scriptstyle 1\times k \end{array}\begin{bmatrix} \nabla h \end{bmatrix} \begin{array}{c}\scriptstyle k\times n\end{array}\begin{bmatrix} - \nabla g_1(x) - \\ \vdots \end{bmatrix} \begin{array}{c}\scriptstyle n\times 1\end{array}\begin{bmatrix} dx \end{bmatrix}, \quad \begin{bmatrix} dg \end{bmatrix}_{k}, \quad [df]$$

$$f''(x) = \nabla g(x)^T \cdot \nabla h^2(g(x)) \cdot \nabla g(x) + \nabla h(g(x))^T \cdot g''(x)$$

$$\underset{n\times k}{\begin{bmatrix} \nabla g_1(x) & \cdots & \nabla g_k(x) \end{bmatrix}} \underset{k\times k}{\begin{bmatrix} \nabla h^2(g) \end{bmatrix}} \underset{k\times n}{\begin{bmatrix} - \nabla g_1(x) - \\ \vdots \\ - \nabla g_k(x) - \end{bmatrix}} + \underset{1\times k}{\begin{bmatrix} \nabla h \end{bmatrix}} \underset{k\times n}{\begin{bmatrix} - g_i''(x) - \end{bmatrix}}$$

$$f \text{ is cvx if } \begin{cases} g_i \text{ is cvx},\ h \text{ is cvx},\ \tilde{h} \text{ is non decreasing in each argument} \\ g_i \text{ is concave},\ h \text{ is cvx},\ \tilde{h} \text{ is non increasing in each argument} \end{cases}$$