Convex Functions, part 4: examples

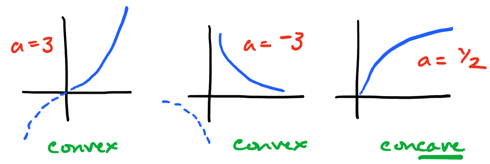
Tuesday, February 2, 2021 6:45 PM

from §3.1.5 BV'04 * why important? Building blocks / shortcuts ... also in cvx/cvxpy

$f:\mathbb{R}\to\mathbb{R}$, examples of convex functions

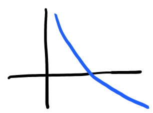
* we plan to cover 3.1, 3.2, 3.3
skip 3.4, 3.5, 3.6

- e^{ax} any $a\in\mathbb{R}$
- x^a on $x\in\mathbb{R}_{++}$ if $a\le 0$ or $a\ge 1$
  (if $0\le a\le 1$, it's concave)

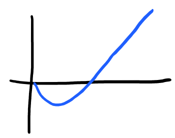


...but not convex (nor concave) on all of $\mathbb{R}$

- $|x|^a$ on all of $\mathbb{R}$, if $a\ge 1$
- $-\log_b(x)$ on $\mathbb{R}_{++}$ if $b>1$



- $\begin{cases} x\cdot\log(x) & x>0 \\ 0 & x=0 \end{cases}$ on $\mathbb{R}_+$ since $f''(x)=\frac{1}{x}>0$



$f:\mathbb{R}^n\to\mathbb{R}$ examples of convex functions

- any norm or seminorm $\boxed{Q1}$ prove this

$\boxed{A1}$ WTS $\forall x,y$ $\forall t\in[0,1]$, $\|tx+(1-t)y\| \le t\|x\|+(1-t)\|y\|$

Take $\|tx+(1-t)y\| \le \|tx\| + \|(1-t)y\|$ by Δ-ineq.
$= |t|\cdot\|x\| + |1-t|\cdot\|y\|$ by pos. homogeneous
$= t\,\|x\| + (1-t)\|y\|$ since $t\ge 0$, $1-t\ge 0$ since $t\in[0,1]$.

- $f(\vec{x}) = \max(x_1, x_2, \ldots, x_n)$
- $f(x,y) = x^2/y$, $\mathrm{dom}(f) = \mathbb{R}\times\mathbb{R}_{++}$ "quadratic-over-linear"

  and more generally,
  $f(\vec{x},y) = \|x\|^2/y$, $\mathrm{dom}(f) = \mathbb{R}^{n-1}\times\mathbb{R}_{++}$ is convex
  and even more generally,
  $f(\vec{x},Y) = x^T Y^{-1} x$ on $\mathrm{dom}(f) = \mathbb{R}^n\times S^n_{++}$ "matrix fractional function"

$\boxed{Q2}$ What about the "linear fractional function" (p.41) $g(x) = \frac{Ax+b}{c^Tx+d}$
$\mathrm{dom}(g) = \{x : c^Tx+d>0\}$

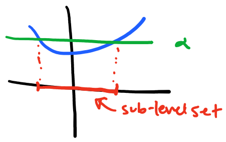
(ex: $g(p) = p_i / \sum_j p_j$ renormalization)

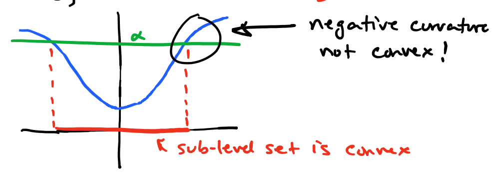
$\boxed{A2}$ No, not convex. See §3.4 p.97 Ex. 3.32, shows $\dfrac{a^T x + b}{c^T x + d}$ is quasi-convex

(quasi-convex: the sub-level sets, $\{ x : f(x) \le \alpha \}$, are all convex)

Ex. $f(x) = e^{-x}$ or any 1D monotonic fcn

Convex $\Rightarrow$ quasi-convex
$\cancel{\Leftarrow}$





... in fact, it's quasi-convex and quasi-concave
i.e., quasi-linear.

⚠ Quasi-convex fcn can be quasi-solved* via convex programming e.g.,
*not a word
use constraint $\{ x : f(x) \le \alpha \}$, do bisection (or other root-finding) on α

... but f, g cvx $\Rightarrow$ $f+g$ cvx. f, g both quasi-convex $\cancel{\Rightarrow}$ $f+g$ quasi-cvx Fragile!

- "log-sum-exp" aka "soft-max" $\longrightarrow$ used in machine learning. Smooth... but be careful numerically!

$$f(\vec{x}) = \frac{1}{\alpha} \log \left(e^{\alpha x_1} + \dots + e^{\alpha x_n} \right), \quad \alpha > 0 \text{ a parameter}$$

$$(\longrightarrow \max(x_1, \dots, x_n) \text{ as } \alpha \to \infty)$$

- geometric mean $f(\vec{x}) = \left(\prod_{i=1}^{n} x_i \right)^{1/n}$ on $\mathbb{R}^n_{++}$

- $-\log \det(X)$ on S^n_{++} (s. pos. def)

$$= -\log \left(\prod \lambda_i \right) = -\sum \log(\lambda_i)$$ λ_i are eigenvalues of X

Jensen's Inequality, i.e., definition of convexity

If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, then for any $t \in [0,1]$

$$f(tx + (1-t)y) \le \underbrace{t}_{p_1} \cdot f(x) + \underbrace{(1-t)}_{p_2} f(y)$$

$p_1, p_2 \ge 0$, $p_1 + p_2 = 1$] Discrete prob. distribution

$$f(p_1 x + p_2 y) \le p_1 f(x) + p_2 f(y)$$

Interpret: X a random variable, $X = \begin{cases} x & \text{w.p. } p_1 \\ y & \text{w.p. } p_2 \end{cases}$

Expected value

so $f(\mathbb{E} x) \le \mathbb{E} f(x)$

In fact, this is true for any probability distribution, not just discrete ones

$$\boxed{f(\mathbb{E} x) \le \mathbb{E} f(x)}$$ Jensen's Inequality

Ex in machine learning, stats, we often prove something like

$$\mathbb{E}\,\|error\|^2 \le \varepsilon$$

think of $f(x) = x^2$, $\mathbb{E}\, f(\|error\|)$, so via Jensen's ...

$$(\mathbb{E}\,\|error\|)^2 \le \mathbb{E}\,\|error\|^2 \le \varepsilon$$

or

$$\mathbb{E}\,\|error\| \le \sqrt{\mathbb{E}\,\|error\|^2} \le \sqrt{\varepsilon}$$

NOT typically an equality.

why would you do this?

$\|\vec{e}\|^2 = \sum_{i=1}^{n} e_i^2$ is separable, nice to work with

also smooth. In fact... nicest function ever!

vs

$\|\vec{e}\| = \sqrt{\sum_{i=1}^{n} e_i^2}$ $\sqrt{\ }$ ruins everything

not separable

not differentiable at 0

not strongly smooth

not strongly convex

Hölder-Inequality / Cauchy-Schwarz

proof via Jensen's.

$\frac{1}{\infty} = 0$ for this

If $\frac{1}{p} + \frac{1}{q} = 1$, $\quad |\langle x, y \rangle| \le \|x\|_p \cdot \|y\|_q$

$p = q = 2$ is special case of Cauchy-Schwarz

$$\left.\begin{matrix}(1, \infty) \\ (\infty, 1) \\ (2, 2)\end{matrix}\right\}\text{ common Hölder conjugates}$$