1 Convex Functions

Check convexity: (1) definition $f(\lambda x+(1-\lambda)y) \leq \lambda f(x)+(1-\lambda)f(y)$ (2) 1-order, 2-order condition (3) function of any line in the domain must satisfy convexity conditions (4) epi$f$ must be convex.

1.1 Definition of Convexity

- 3.1 (a): Since the linear function $g(x)$ is always above the convex $f(x)$, hence applying the slope

  $$\frac{f(b)-f(a)}{b-a}=\frac{g(x)-f(a)}{x-a}$$

  gives

  $$g(x)=f(a)+(x-a)\frac{f(b)-f(a)}{b-a}=\frac{b-x}{b-a}f(a)+\frac{x-a}{b-a}f(b)\geq f(x)$$

  $\blacksquare$

- 3.1 (b): Subtract both sides by $f(a)$

  $$f(x)-f(a)\leq(x-a)\frac{f(b)-f(a)}{b-a}$$

  gives

  $$\frac{f(x)-f(a)}{x-a}\leq\frac{f(b)-f(a)}{b-a}$$

  $\blacksquare$

- 3.1 (c): When $x\to a$, the lhs inequality becomes $f'(a)$. Likewise the rhs inequality becomes $f'(b)$. $\blacksquare$

- 3.1 (d): Since $f''(a)=\lim\limits_{x\to a}\frac{f'(x)-f'(a)}{x-a}$, and $f'(x)\geq f'(a)$ when $x\geq a$, hence $f''(a)\geq 0$. $\blacksquare$

- 3.2: ***Level Sets of convex, concave, quasiconvex, and quasiconcave functions.***

- 3.3: ***Inverse of an "increasing" convex function.***
  g is an inverse function of an increasing convex function $f:\mathcal{R}\to\mathcal{R}$, where dom$f=(a,b)$, dom$g=(f(a),f(b))$, and $g(f(x))=x$. Prove that g is concave.

  Proof:
  Since f is increasing, given $x_2>x_1$, we know that $f(x_2)>f(x_1)$, set $x_1=\theta a+(1-\theta)b$ for $0<\theta<1$. Due to convexity of f, if $f(x_2)=\theta f(a)+(1-\theta)f(b)$ then $f(x_2)>f(\theta a+(1-\theta)b)=f(x_1)$ is satisfied. We want to show that g is concave by proving $g(\theta f(a)+(1-\theta)f(b))>\theta g(f(a))+(1-\theta)g(f(b))$. From $g(f(x))=x$, we know that the lhs of the inequality is $g(f(x_2))=x_2$ and rhs is $\theta a+(1-\theta)b=x_1$, therefore $x_2>x_1$. $\blacksquare$

- 3.4: $f:\mathcal{R}^n\to\mathcal{R}$ ***is continuous, show that*** f ***is convex iff given any line segment (with endpoints*** x ***and*** y***)***

  $$\int_0^1 f((1-\lambda)x+\lambda y)d\lambda\leq\frac{f(x)+f(y)}{2},$$

  i.e. the mean of f over the line segment is less than the average of the values on the endpoints. (Notice that the lhs comes from the mean

  $$\frac{1}{y-x}\int_x^y f(z)dz$$

by setting $z = (1-\lambda)x + \lambda y$, $dz = (y-x)d\lambda$, and $z = x$ when $\lambda = 0$ and $z = y$ when $\lambda = 1$, hence

$$\frac{1}{y-x}\int_0^1 f(1-\lambda)x + \lambda y)(y-x)d\lambda = \int_0^1 f((1-\lambda)x + \lambda y)d\lambda.$$

Proof:
The linear function $g(z)$ that intersects $f(x)$ and $f(y)$ is $g(z) = f(x) + \frac{f(y)-f(x)}{y-x}(z-x)$. We know that convexity of f gives $f(z) \le g(z)$, hence taking the mean of $g(z)$ over $[xy]$ gives $\frac{f(x)+f(y)}{2}$ which is $\ge$ the mean of f. ■

- 3.5: ***Running average of a convex function.***
  $f : \mathcal{R} \to \mathcal{R}$ is convex, show that the running avarage

  $$F(x) = \frac{1}{x}\int_0^x f(t)dt \tag{1}$$

  is convex with $\mathrm{dom}F = \mathcal{R}_{++}$.
  (Notice as 3.4, set $t = \lambda x$, $dt = xd\lambda$, $t = x$ when $\lambda = 1$, $t = 0$ when $\lambda = 0$, hence

  $$\frac{1}{x}\int_0^1 f(\lambda x)xd\lambda = \int_0^1 f(\lambda x)d\lambda \tag{2}$$

  is the moving average over $[0, x]$.)
  Proof:
  Since for any given λ, $f(\lambda x) = f(g(x))$ has a convex domain $\mathrm{dom}f = g(x)$, where $g(x)$ is an affine function. Therefore, given f is a convex function, $f(g(x))$ is convex over x. A nonnegative weighted (by 1's) sum of convex functions preserves convexity. Therefore, $F(x) = \int_0^1 f(\lambda x)d\lambda$ is convex. ■

- ***3.6: Functions and epigraphs.***
  When is the epigraph of a function a (1) halfspace (2) convex cone (3) polyhedron?

  Proof:
  $\mathrm{epi}f = \{(x,t) | t \ge f(x)\}$

  1. The normal vector $(\nabla f, -1)$ is a constant over the entire domain, and the $\mathrm{epi}f$ is bounded by the hyperplane formed by the graph $(x, f(x))$.
  2. Any conic combination of points (x, t) are in the epigraph.
  3. Intersection of hyperplanes on the graph $(x, f(x))$ forming halfspaces with normal vectors $(\nabla f, -1)$.

- 3.7: ***Prove that a convex function, with dom$f = \mathcal{R}^n$, bounded above is a constant.***
  Proof:
  Suppose f is non-differentiable and bounded above at the point y. Any line segment in $\mathcal{R}^n$ must give convexity of f, and suppose $z \succ y \succ x$ on this line segment. If $f(z) = f(y) > f(x)$, then the convex combination $\theta f(z) + (1-\theta)f(x) < f(y)$, which is non-convex. If $f(z) > f(y) > f(x)$, then f is unbounded. Therefore, f must be a constant, since all other possibilities are non-convex.

- 3.8: ***2-order condition for convexity.***
  If f is twice differentiable, then f is convex iff $\nabla^2 f(x) \succeq 0$ for all $x \in \mathrm{dom}f$.

  Proof:
  $(\Rightarrow)$ (1) $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$ (2) $\mathrm{epi}f$ is convex (3) any function of a line segment

in the domain is convex (4) 1-order condition $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ (5) sublevel set is convex. Use (4) The 2-order Taylor expansion which better approximates $f(y)$ is $f(x) + \nabla f(x)^T(y-x) + \nabla^2 f(x)^T(y-x)^2$, this is always greater than or equal to the 1-order expansion if $\nabla^2 f(x) \succeq 0$. ($\Leftarrow$) From the 2-order condition we know that $f(y) \geq f(x) + \nabla f(x)^T(y-x) + \nabla^2 f(x)^T(y-x)^2 \geq f(x) + \nabla f(x)^T(y-x)$, therefore f is convex.

- 3.9: ***2-order conditions for convexity on an affine set.***
  Suppose f twice differentiable with a convex domain $x = Fz + \hat{x} \in \mathbf{dom} f$.
  1) Show that $\tilde{f}(z) = f(Fz + x)$ is convex iff $F^T f(Fz + \hat{x})F \succeq 0$, where $F \in \mathcal{R}^{n \times m}$, $\hat{x} \in \mathcal{R}^n$, $z \in \mathcal{R}^m$, $Fz + \hat{x} \in \mathbf{dom} f$, and $z \in \mathbf{dom}\tilde{f}$. (Notice, this convexity condition of $F^T \nabla^2 f F$ on $z \in \mathbf{dom}\tilde{f}$ is analogous to that of $\nabla^2 f(x)$ on $x \in \mathbf{dom} f$.)
  2) $A \in \mathcal{R}^{p \times n}$ has null$A$=range$F$. Show that $\tilde{f}$ is convex iff there exists a $\lambda \in \mathcal{R}$ such that

$$\nabla^2 f(Fz + \hat{x}) + \lambda A^T A \succeq 0 \tag{3}$$

  Proof:

  (1) Since f is convex iff $\nabla^2 f(Fz + \hat{x}) \succeq 0$ for all $Fz + \hat{x} \in \mathbf{dom} f$. For a positive semidefinite matrix, the rhs is satisfied when $\Big(Fz+\hat{x}\Big)^T \nabla^2 f\Big(Fz+\hat{x}\Big) = z^T\big(F^T\nabla^2 fF\big)z + 2\hat{x}^T\nabla^2 fFz + \hat{x}\nabla^2 f\hat{x} \geq 0$. Since $\hat{x}$ and Fz are in $\mathbf{dom} f$ satisfying $\nabla^2 f \succeq 0$, therefore the first term must satisfies $z^T\big(F^T\nabla^2 fF\big)z \geq 0 \Rightarrow \big(F^T\nabla^2 fF\big) \succeq 0$.
  This shows that $\tilde{f}$ is convex iff $F^T\nabla^2 fF \succeq 0$ for all $z \in \mathbf{dom}\tilde{f}$. ■

  2) ($\Rightarrow$) $\lambda F^T A^T AFz$ is stricly zero for any λ, therefore from 1) we know that a convex f implies $F^T\nabla^2 fF \succeq 0 \Rightarrow \mathrm{null}F^{\perp} \in \mathbf{dom} f$, we know that $Az = A(\mathrm{null}F) + 0$, therefore $\lambda F^T A^T AFz$ is stricly zero for any λ. $F^T\nabla^2 fF + \lambda F^T A^T AF \succeq 0 \Rightarrow F^T\Big(\nabla^2 f + \lambda A^T A\Big)F \succeq 0 \Rightarrow \nabla^2 f(Fz+\hat{x}) + \lambda A^T A \succeq 0$.
  ($\Leftarrow$) $F^T\Big(\nabla^2 f(Fz + \hat{x}) + \lambda A^T A\Big)F \succeq 0$ gives the previous result in 1). ■

- 3.10: ***An extension of Jensens inequality.***
  Given a convex f with $\mathbf{E}f(x_0 + v) \geq f(x_0)$, where v is a random variable with zero mean.
  1) Find a counterexample that a higher variance v (i.e., randomization, more deviated from the mean) raises the mean value of f, i.e., $\mathbf{Var}(v) > \mathbf{Var}(w)$, but $\mathbf{E}f(x_0 + v) < \mathbf{E}f(x_0 + w)$. (The general case is supported by supposing $\mathbf{Var}(w) \to 0$, then $\mathbf{E}f(x_0 + w) \to \mathbf{E}f(x_0) = f(x_0)$, where $\mathbf{E}f(x_0 + v) \geq f(x_0)$ concludes that $\mathbf{E}f(x_0 + v) \geq \mathbf{E}f(x_0 + w)$, i.e., randomization will raise the mean of a convex function.)
  2) Show that $\mathbf{E}f(x_0 + tv)$ is monotonically increasing for $t \geq 0$, i.e. $w = tv$ just a scaling of the values of v without changing distribution.
  Proof:
  1) Given a convex function $f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$ and two distributions, $p(v = -4) = 0.1$ and $p(v = 4/9) = 0.9$ with $\mathbf{Var}(v) = 1.777$, and $p(w = -1) = 0.5$ and $p(w = 1) = 0.5$ with $\mathbf{Var}(w) = 1$. Therefore $\mathbf{Var}(v) > \mathbf{Var}(w)$, $\mathbf{E}f(v) = 0.4 < \mathbf{E}f(w) = 0.5$. ■
  2) $g(t) = \mathbf{E}f(x_0 + tv) = \theta_1 f_{v_1}(t) + \ldots + \theta_n f_{v_n}(t)$ is a convex function of t, indexed by v_i, since a nonnegative sum of convex functions is convex.

- 3.11 ***Monotone Mappings*** Definition of a "monotone" function $\psi : \mathcal{R}^n \to \mathcal{R}^n$

$$(\psi(x) - \psi(y))^T(x - y) \geq 0. \tag{4}$$

  i.e. ψ increases (decreases) with positive (negative) x, they are positively correlated.

Given f is a differentiable convex function. 1) Prove ∇f is monotone. 2) Is the converse true?

Proof:
1) From the 1-order condition, $f(y) \geq f(x) + \nabla f(x)^T(y-x)$ and $f(x) \geq f(y) + \nabla f(y)^T(x-y)$. Adding the two equations we get $(\nabla f(x) - \nabla(y))^T(x-y) \geq 0$. ■

2) Given a monotone $(\nabla f(x) - \nabla(y))^T(x-y) \geq 0$, is f convex? (UNPROOFED)

- 3.12 ***Fit an affine function between a convex and concave function with the same domain***
$g(x) \leq f(x)$, g is concave and f is convex.
Proof:
We know that hypo$g$ and epi$f$ are both convex sets which does not intersect, therefore there exists a hyperplane between the two sets. Suppose the two points $z_1 = (x_1, f(x_1))$ and $z_2 = (x_2, g(x_2)$ gives $\inf||u-v||$, where $u \in \{(x, f(x))\}$ and $v \in \{(x, g(x))\}$. $(\nabla f(x_1), -1)^T((x, h(x)) - (z_1+z_2)/2) = 0$ $\Rightarrow h(x) = \nabla f(x_1)^T x - (\nabla f(x_1), -1)^T(\frac{z_1+z_2}{2})$, where h is affine. ■

- 3.13 ***Kullback-Leibler divergence and the information inequality***
KL divergence:
$$\mathrm{D}_{\mathrm{KL}}(u,v) = \sum_{i=1}^{n}(u_i \log\frac{u_i}{v_i} - u_i + v_i), \tag{5}$$
becomes relative entropy when u and v are n dimensional probability vectors s.t. $1^T u = 1$ and $1^T v = 1$, n is also the sample size.
1) Prove *infomation inequality* $\mathrm{D}_{\mathrm{KL}}(u,v) \geq 0$ for all u, $v \in \mathcal{R}^n_{++}$.
2) $\mathrm{D}_{\mathrm{KL}}(u,v) = 0$ iff $u = v$.
Hint: $\mathrm{D}_{\mathrm{KL}} = f(u) - f(v) - \nabla f(v)^T(u-v) = \sum u_i \log u_i - \sum v_i \log v_i - (\log v_1 + 1, \ldots, \log v_n + 1)^T(u_1 - v_1, \ldots, u_n - v_n)$ where $f(v) = \sum_{i=1}^{n} v_i \log v_i$ is the negative entropy of v.
Proof:
Since negative entropy is strictly convex and differentiable, therefore for $u \neq v$, $f(u) > f(v) + \nabla f(v)^T(u-v) \Rightarrow \mathrm{D}_{\mathrm{KL}}(u,v) = f(u) - f(v) - \nabla f(v)^T(u-v) > 0$. ■

- 3.14 (unproof)

- 3.15 ***A family of Concave Utility functions***
For $0 < \alpha \leq 1$,
$$u_\alpha(x) = \frac{x^\alpha - 1}{\alpha} \tag{6}$$
with dom$u_\alpha = \mathcal{R}_+$, and $u_0 = \log x$ with domain $\mathcal{R}_{++}$. (a) Show that $\lim_{\alpha\to 0} u_\alpha = u_0$.
(b) Show that u_α are concave, monotone increasing, and $u_\alpha(1) = 0$.
Proof:
(a) $u' = x^{\alpha-1}$, at $\alpha = 0$ $u' = x^{-1} = u_0'$. ■
(b) $f(x) = (x^\alpha - 1)/\alpha$ is concave iff $f'' \leq 0$. $f'' = (\alpha-1)x^{\alpha-2}$, $0 < \alpha \leq 1$ and $x \in \mathcal{R}_{++}$ imply $f'' \leq 0$. $x \geq y$ implies $x^\alpha \geq y^\alpha$, therefore $(x^\alpha - 1)/\alpha \geq (y^\alpha - 1)\alpha$, which indicates monotonicity. $u_\alpha(1) = 0$ is trivial to prove. ■

- 3.16 ***Determine convex, concave, quasi-convex, quasi-concave***
(a) $f(x) = e^x - 1$ on $\mathcal{R}$.

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathcal{R}^2_{++}$.
(c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathcal{R}^2_{++}$.
(d) $f(x_1, x_2) = x_1/x_2$ on $\mathcal{R}^2_{++}$.
(e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathcal{R} \times \mathcal{R}_{++}$.
(f) $f(x_1, x_2) = x_1^a x_2^{1-a}$, where $0 \leq a \leq 1$, on $\mathcal{R}^2_{++}$.
Proof:
(a) Strictly convex, since $e^x - 1$ is the convex e^x, with $f'' = e^x > 0$, shifted by a constant, and monotonically increasing. Hence quasi-convex and quasi-concave (superlevel set is convex).
(b) The Hessian $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is not a positive semidefinite nor negative semidefinite matrix, hence neither convex nor concave. f is quasiconcave since the superlevel set $\{x | f(x) = x_1 x_2 \geq t, t \in \mathcal{R}\}$ is convex on $\mathcal{R}^2_{++}$.
(c) The Hessian $\frac{1}{x_1 x_2} \begin{pmatrix} 2x_2^{-2} & 1/(x_1 x_2) \\ 1/(x_1 x_2) & 2x_1^{-2} \end{pmatrix} \succeq 0$, by row reduction the diagonal elements are positive eigenvalues.
Therefore convex, and quasi-convex. The superlevel set $\{x | 1/(x_1 x_2) \geq t, t \in \mathcal{R}\}$ is not convex, hence not quasi-concave.
(d) The Hessian $\begin{pmatrix} 0 & -x_2^{-2} \\ -x_1^{-2} & 0 \end{pmatrix}$ is not PSD nor NSD. The level set $\{x | x_1/x_2 = t\}$ is a line (hyperplane in $\mathcal{R}^2$), which means f is quasilinear (i.e. both quasiconvex and quasiconcave).
(e) The Hessian $2/x_2 \begin{pmatrix} 1 & -x_1/x_2 \\ -x_1/x_2 & x_1^2/x_2^2 \end{pmatrix} \succeq 0$ by row reduction, thus convex and quasiconvex. It is the quadratic-over-linear function.
(f) The Hessian $a(1-a) x_1^a x_2^{-a} \begin{pmatrix} -x_1^{-2} x_2 & x_1^{-1} \\ x_1^{-1} & -x_2^{-1} \end{pmatrix} \preceq 0$ by row reduction, hence concave and quasi-concave. It is not quasi-convex.

- 3.17 ***Show*** $f(x) = \left(\sum\limits_{i=1}^n x_i^p \right)^{\frac{1}{p}}$ ***is concave with*** $p < 1$***,*** $p \neq 0$***, and dom***$f = \mathcal{R}^n_{++}$***.***

Special case: $f(x) = \left(\sum\limits_{i=1}^n x_i^{1/2} \right)^2$ and the *harmonic mean* $f(x) = \left(\sum\limits_{i=1}^n 1/x_i \right)^{-1}$. (Hint: prove by using Cauchy-Schwarz inequality $||a|| ||b|| \geq a^T b$, which is from triangular inequality $||a + b|| \leq ||a|| + ||b||$ taken squares on both sides $a^T a + 2a^T b + b^T b \leq a^T a + 2||a|| ||b|| + b^T b \Rightarrow a^T b \leq ||a|| ||b||$)
Proof:
$(\nabla f)_i = (\sum\limits_{k=1}^n x_k^p)^{\frac{1-p}{p}} x_i^{p-1}$
$(\nabla^2 f)_{ii} = -(1-p) x_i^{p-2} (\sum\limits_{k=1}^n x_k^p)^{\frac{1-p}{p}} + (1-p) x_i^{2p-2} (\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} = (1-p)(\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} \left(-x_i^{p-2} \sum\limits_{k=1}^n x_k^p + x_i^{2p-2} \right)$
$(\nabla^2 f)_{ij} = (1-p) x_i^{p-1} (\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} x_j^{p-1} = (1-p)(\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} \left(x_i^{p-1} x_j^{p-1} \right)$
$\nabla^2 f = (1-p)(\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} \left(-\operatorname{diag}(x_i^{p-2} \sum\limits_{i=1}^n x_k^p) + zz^T \right)$
where $z = (x_1^{p-1}, \ldots, x_n^{p-1})$. The condition $v^T \nabla^2 f v = (1-p)(\sum\limits_{k=1}^n x_k^p)^{\frac{1-2p}{p}} \Big(-\sum\limits_{i=1}^n (v_i^2 x_i^{p-2} \sum\limits_{i=1}^n x_k^p) + (\sum v_i z_i)^2 \Big) \leq 0$ must be satisfied for $\nabla^2 f$ to be NSD for concavity. Where the components in the bracket must be negative. Therefore, with $a_i = v_i x_i^{\frac{p-2}{2}}$, $b_i = x_i^{p/2}$ and $a_i b_i = v_i x_i^{p-1} = v_i z_i$, the component being negative becomes $||a||^2 ||b||^2 \geq (a^T b)^2$ which complies with Cauchy-Schwarz inequality. ■

- 3.18 ***Function of PSD matrices***
  Prove by using spectral theory (Hermitian self-adjoint matrices can be diagonalized w.r.t. orthonormal eigenbasis).
  (a) $f(X) = \text{tr}(X^{-1})$ is convex on $\text{dom}f = S^n_{++}$.
  (b) $f(X) = (\det X)^{1/n}$ is concave on $\text{dom}f = S^n_{++}$.
  Proof:
  (a) Consider an arbitrary line $X^{-1} = Y + tZ$, $g(t) = f(Y + tZ) = \text{tr}(Y + tZ) = \text{tr}Y^{1/2}(I + tY^{-1/2}ZY^{-1/2})Y^{1/2}$.
  (b) (unproof).

- 3.19 ***Nonnegative Weighted Sum and Integrals***
  (a) Show that $f(x) = \sum_{i=1}^{k} \alpha_i x_{[i]}$ is convex, where $x_{[i]}$ is the kth largest component of $x \in \mathcal{R}^n$, and $\alpha_1 \geq \alpha_2 \geq \ldots \geq 0$. Use the fact that $f(x) = \sum_{i=1}^{k} x_{[i]}$ is convex.
  (b) Show that $f(x) = -\int_0^{2\pi} \log T(x,\omega)d\omega$ is convex on $\{x \in \mathcal{R}^n | T(x,\omega) > 0\}$, where $T(x,\omega) = x_1 + x_2\cos\omega + x_3\cos 2\omega + \ldots + x_n\cos(n-1)\omega$
  Proof:
  (a) The nonnegative sum of the set of convex functions $f_j = \sum_{i=1}^{j} x_{[i]}$ is $\sum_{j=1}^{k} \alpha_j f_j = \sum_{i=1}^{k} \alpha_i x_{[i]} + \sum_{j=1}^{k-1} \alpha_{j+1} f_j$. Which indicates $\sum_{i=1}^{k} \alpha_i x_{[i]} = \sum_{j=1}^{k-1}(\alpha_j - \alpha_{j+1})f_j + \alpha_k f_k$ is also a nonnegative sum of convex functions which must also be convex. ■
  (b) $T(x,\omega) = \sum_{k=1}^{n} x_k \cos(k-1)\omega$.
  $(\nabla \log T)_i = -\cos(i-1)\omega / (\sum x_k \cos(k-1)\omega)$
  $(\nabla^2 \log T)_{ij} = -\cos(i-1)\omega\cos(j-1)\omega / (\sum x_k \cos(k-1)\omega)^2$
  Hence, $\nabla^2 \log T = -\frac{1}{(\sum x_k \cos(k-1)\omega)^2}\begin{pmatrix} 1 \\ \cos\omega \\ \vdots \\ \cos(n-1)\omega \end{pmatrix}[1, \cos\omega, \cdots, \cos(n-1)\omega] \preceq 0$
  Which shows that $\log T$ is concave, and $-\log T$ is convex. Therefore $f(x)$ is a continuous sum of convex functions indexed by ω. ■

- 3.20 ***Composition with an Affine Function***
  Show the followings are convex functions:
  (a) $f(x) = ||Ax - b||$, norm on $\mathcal{R}^m$.
  (b) $f(x) = -(\det(A_0 + x_1A_1 + \cdots + x_nA_n))^{1/m}$, on $\{x | A_0 + x_1A_1 + \cdots + x_nA_n \succ 0\}$, where $A_i \in S^m$.

  Proof:
  An affine function preserves the convexity of the points in $\text{dom}f$, i.e., $y = Ax - b$ is a convex set of points. (a) Since a norm is a convex function, therefore operating on a convex set of points $y = Ax - b$ preserves the convexity of the function.
  (b) (unproof)

- 3.21 ***Pointwise Maximum and Supremum***
  Show the followings are convex functions:
  (a) $f(x) = \max_{i=1,\cdots,k} ||A^{(i)}x - b^{(i)}||$.
  (b) $f(x) = \sum_{i=1}^{r} |x|_{[i]}$ on $\mathcal{R}^n$, where $|x|_{[i]}$ is the rth ordered maximum coordinates.
  Proof:
  (a) f is a pointwise maximum function evaluated at a fixed point x_1 over a set of convex norm

functions (composed over affine functions), which is convex. ■
(b) (unproof) hint: An affine function $f(x) = a^T x = [0,1,0,1,1]^T[x_1, \cdots, x_5] = x_2 + x_4 + x_5$ is convex.
$f(x) = \max_{1 \le i_1 \le \cdots \le n}\{x_{i_1} + \cdots + x_{i_r}\}$ is the pointwise maximum of fixed set of affine functions. Same thing applied to the function in (b).

- 3.22 ***Composition Rule***
Show the followings are convex functions:
(a) $f(x) = -\log(-\log(\sum_{i=1}^m e^{a_i^T x + b_i}))$ on $\mathrm{dom} f = \{x | \sum_{i=1}^m e^{a_i^T x + b_i} < 1\}$. Notice $\log(\sum_{i=1}^m e_i^y)$ is convex by $\nabla^2 f \succeq 0$.
(b) $f(x,u,v) = -\sqrt{uv - x^T x}$, $\mathrm{dom} f = \{(x,u,v) \mid uv > x^T x, u, v > 0\}$, and $-\sqrt{x_1 x_2}$ is convex on $\mathcal{R}^2_{++}$. Quadratic-over-linear is a convex function.
(c) $f(x,u,v) = -\log(uv - x^T x)$, same $\mathrm{dom} f$ as (b).
(d) $f(x,t) = -(t^p - ||x||_p^p)^{1/p}$ where $p > 1$ and $\mathrm{dom} f = \{(x,t) \mid t \ge ||x||_p\}$.
(e) $f(x,t) = -\log(t^p ||x||_p^p)$ where $p > 1$ and $\mathrm{dom} f = \{(x,t) \mid t > ||x||_p\}$. **Proof**:
(a) $h(x) = \log(\sum_{i=1}^m e^{a_i^T x + b_i})$ is a composition of log-sum-exp function over the affine functions $y_i = a_i T x + b_i$ (where any x point is transformed linearly to another point $y_i(x)$ indexed by i), where convexity is preserved. $h : \mathcal{R}^n \to \mathcal{R}$ and $g : \mathcal{R} \to \mathcal{R}$. We can restrict the domain of h to a line $x = x_0 + tv$, then $f(x) = g(h(x))$, where $g(h) = -\log(-h)$, has the condition $f'' = g'' h'^2 + g' h''$. $h(x) = k(z(x))$, where $k = \log(z)$ and $z = \sum e^{a_i^T x + b_i}$, hence $h'' = k'' z'^2 + k' z'' = \frac{1}{z^2}(\sum 1^T a_i e^{a_i^T x + b_i})^2 + (\sum 1^T a_i^2 e^{a_i^T x + b_i})/z \ge 0$ since a_i's are the only terms possibly negative. Therefore, $g'' = 1/h^2 > 0, g' = 1/h > 0, h'' > 0$, hence f is convex. ■

(b) $-\sqrt{x_1 x_2}$ is convex by the Hessian being $0.25 x_1^{-1/2} x_2^{-1/2}[x_1^{-1}, -x_2^{-1}][x_1^{-1}, -x_2^{-1}]^T \succeq 0$. $g(x,u) = x_1^2/u + x_2^2/u + \cdots + x_n^n/u = g_1(x_1,u) + \cdots + g_n(x_n,u)$ is a positive sum of quadratic-over-linear functions, hence the nonnegative sum is convex.
$f(x,u,v) = -\sqrt{u(v - g(x,u))}$, where both $g_1(x,u,v) = u$ and $g_2(x,u,v) = v - g(x,u)$ are positive, and $v - g(x,u)$ is an affine transformation of the concave function $-g$, hence also concave. Therefore, the composition $f = h(g(x,u,v))$ is convex. ■

(c) Suppose $h(x,y) = -\log(xy)$, $\nabla^2 h(x,y) = \begin{pmatrix} x^{-2} & 0 \\ 0 & y^{-2} \end{pmatrix} \succeq 0$, hence $h(x,y)$ is convex on $\mathcal{R}^2_{++}$. Following the same logic from (b), $f(x,u,v) = h(g(x,u,v))$ is convex. ■

(d) $-(t^{p-1}(t - \frac{||x||_p^p}{t^{p-1}}))^{1/p}$, where $\frac{||x||_p^p}{t^{p-1}}$ is convex by 3.23 (a). $h(x,y) = -(xy)^{1/p}$, both x and y are nonnegative, hence the Hessian

$$1/p^2 x^{1/p-1} y^{1/p-1}[\sqrt{(p-1)y/x}, -\sqrt{x/((p-1)y)}][\sqrt{(p-1)y/x}, -\sqrt{x/((p-1)y)}]^T \succeq 0$$

over the domain $\mathcal{R}^2_+$. Therefore, t^{p-1} is either convex or concave depending on p, and $t - \frac{||x||_p^p}{t^{p-1}}$ is concave. The composition by h over the convex domain gives convex f. ■

(e) $f(x,t) = -\log(t^{p-1}) - \log(t - (||x||_p/t)^p)$, where $-(p-1)\log(t)$ is convex and $-\log(t - (||x||_p/t)^p)$ is convex (by the composition rule $f'' = h'' g'^2 + g'' h' \ge 0$). The positive sum of convex functions is convex. ■

- 3.23 ***Perspective of a Function***
The perspective of a function g is $f(x,t) = tg(x/t)$ (Notice the perspective function $P(x,t) = \frac{x}{t}$ on

dom$P = \mathcal{R}^n \times \mathcal{R}_{++}$ preserves convex sets, from domain to image, but is not a convex function!) (a) Show that for $p > 1$,

$$f(x,t) = \frac{|x_1|^p + \cdots + |x_n|^p}{t^{p-1}} = \frac{||x||_p^p}{t^{p-1}} \tag{7}$$

is convex on $\{(x,t)|t > 0\}$.
(b) Show that $f(x) = \frac{||Ax+b||_2^2}{c^Tx+d}$ is convex on $\{x \mid c^Tx + d > 0\}$, where $A \in \mathcal{R}^{m\times n}$, $b \in \mathcal{R}^m$, $c \in \mathcal{R}^n$ and $d \in \mathcal{R}$.

Proof:
(a) $f(x,t) = t(\frac{||x||_p}{t})^p = t(||\frac{x}{t}||_p)^p = tg(x/t)$ From Minkowski's Inequality $||\lambda x + (1-\lambda)y||_p \leq ||\lambda x||_p + ||(1-\lambda)y||_p = \lambda||x||_p + (1-\lambda)||y||_p$, which shows that $||x/t||_p$ is convex. Since $p > 1$, $g(x/t) = ||x/t||_p^p$ is convex. Therefore $f(x,t) = tg(x/t)$ is convex (proven by using perspective function preserving convexity by mapping epi$f = \{(x,t,v) \mid v \geq tg(x/t)\}$ to epi$g = \{(x/t, v/t) \mid v/t \geq g(x/t)\}$, by knowing g is convex we know that f must be convex). ■

(b) $\frac{||Ax+b||_2^2}{c^Tx+d} = (c^Tx+d)\left(\frac{||Ax+b||_2}{c^Tx+d}\right)^2 = (c^Tx+d)\left(||\frac{Ax+b}{c^Tx+d}||_2\right)^2$, $\frac{Ax+b}{c^Tx+d}$ is a perspective function preserving convex affine sets. From (a), the perspective of a convex function is a convex function. ■

- 3.24 ***Functions on the Probability Simplex***
A probability simplex is the n-dim'l probability vector space that satisfies

$$\{p \in \mathcal{R}_+^n \mid 1^T(p_1, \cdots, p_n) = 1, \text{ where } \ p_i = p(x = a_i), a_i < a_{i+1}\},$$

which is a convex set (the triangular surface on a simplex spanned by p_1, p_2 and p_3, if $n = 3$). Determine the convexity or quasi-convexity of the functions:
(a) $\mathbf{E}x$
(b) $\mathbf{prob}(x \geq \alpha)$
(c) $\mathbf{prob}(\beta \geq x \geq \alpha)$
(d) $f(p) = \sum_{i=1}^n p_i \log p_i$, the negative entropy of the distribution.
(e) $\mathbf{Var}x = \mathbf{E}(x - \mathbf{E}x)^2$
(f) quartile$(x) = \inf\{\beta \mid \mathbf{prob}(x \leq \beta) \geq 0.25\}$.
(g) The cardinality of the smallest set $A \subseteq \{a_1, \cdots, a_n\}$ with $\mathbf{prob}(A) \geq 0.9$. (Disprove quasiconvexity by an example!)
(h) The minimum width interval that contains 90% of the probability, i.e., $\inf\{\beta - \alpha \mid \mathbf{prob}(\alpha \leq x \leq \beta) \geq 0.9\}$.
(f),(g), and (h) is crutial in thinking quasi-convexity, and is proven by giving counter-examples. **Logic of proof: fix the p used for sub- (super-) level set, then prove if given the higher- (lower-) levels whether the p's forms a halfspace.**

Proof:
quasiconvex function must satisfy the 1^{st}-order condition $f(y) \leq f(x) \ \Rightarrow \ \nabla f(x)^T(y - x) \leq 0$, and 2^{nd}-order condition $y^T\nabla f(x) = 0 \ \Rightarrow \ y^T\nabla^2 f(x)y \geq 0$ (i.e., whenever the slope is zero, the curvature is nonnegative). (a) $\mathbf{E}x = f(p) = \sum_i a_i p(x = a_i)$, $f : \mathcal{R}_+^n \to \mathcal{R}$, is an affine/linear transformation on the simplex (a convex set) which preserves convexity. Hence $f(p)$ is a convex line in $\mathcal{R}$ which satisfies convex, concave, quasi-convex and quasi-concave. ■
(Notice if $f(x)$ is a convex function, a random variable of x, and the probability vector is fixed on each outcome of x, then $\mathbf{E}f(x)$ is just a point, we talk only about whether randomization will hurt the mean.)

(b) $\mathbf{prob}(x \geq \alpha) = f(p) = \sum_{\{i|a_i \geq \alpha\}} p(x = a_i)$, where the set $\{p \in \mathcal{R}^{n-m+1} \mid 1^T(p_m, \cdots, p_n) \leq 1$, and $\alpha \leq a_m < \cdots < a_n\}$ is a convex subset of the probability simplex set. Hence this is just a linear function of a convex set of p, which is convex, concave, quasiconvex and quasiconcave. To visualize the domain when part of the probability basis vectors are chopped off, we can focus on $n = 3$, a triangular simplex set. If we chop off $p(x = a_3)$, we get a triangle area projected on p_1-p_2-plane with $p_1 + p_2 \leq 1$, $p \in \mathcal{R}^n_+$. ■

(c) $\mathbf{prob}(\beta \geq x \geq \alpha) = \sum_{\{i|\beta \geq a_i \geq \alpha\}} p(x = a_i)$ is a linear function of the set of p, which is convex, concave, quasiconvex and quasiconcave. ■

(d) Since $x \log x$ is a convex function, $f(p)$ is convex by a nonnegative sum of convex functions, quasiconvex. ■

(e) $\mathbf{Var}x = f(p) = \sum_i x_i^2 p_i - (\sum_i x_i p_i)^2$ is a linear function subtracting a sum of nonnegative quadratic functions, which is concave, quasiconcave. ■

(f) quartile$(x) = \inf\{\beta \mid \mathbf{prob}(x \leq \beta) = \sum_i p_i \geq 0.25\}$. If p is fixed, the quartile only give a single number. When p changes, the quartile may change, i.e., different distributions may have the same or different quartiles. If quartile is a_1, then $p_1(a_1) \geq 0.25$. If quartile is a_2, then $p_1(a_1) + p_2(a_2) \geq 0.25$. Hence this is a function of p, $f(p) =$ quartile$(x) = \inf\{\beta \mid \mathbf{prob}(x \leq \beta) = \sum_i p_i \geq 0.25\}$. Given a point p in the simplex set, we can find the minimum β that satisfies the condition $\mathbf{prob}(x \leq \beta) = \sum_i p_i \geq 0.25$. Since β is picked from one of the values of a_i's, thus the function is not continuous, which is not convex or concave. ■ Therefore, we check if the sub-or-super-level sets are convex for quasi-convex-or-concave. The superlevel set is the p satisfying quartile$(x) \geq \alpha$. If p's satisfies quartile$(x) \geq \alpha$, then using the same p's with the levels $\beta < \alpha$ will gaurantee $\mathbf{prob}(x \leq \beta) < 0.25$, hence defines an open halfspace in the domain, which is convex. ■ If quartile$(x) \leq \alpha$, then with $\beta > \alpha$, $\mathbf{prob}(x \leq \beta) = \sum_i p_i > 0.25$ is gauranteed, hence the domain is strictly convex. ■

(g) $C = \{$ card$(A) \mid A \subseteq \{a_1, \ldots, a_n\}$ with $\mathbf{prob}(A) = \sum_i p_i \geq 0.9\}$, and $f(p) =$ minimum$\{C\}$. Same logic as (f), the function gives integer values which is not continuous over the domain, hence not convex or concave. ■ Given a card(A), there will be multiple A in the set C. If p's satisfy $f(p) \geq \alpha$, then for all card$(B) < \alpha$, $\mathbf{prob}(B) = \sum_i p_i < 0.9$. Hence the superlevel set is a strictly convex set bounded by in a halfspace, quasiconcave. ■ If p's satisfy $f(p) \leq \alpha$, then for card$(B) > \alpha$, $\mathbf{prob}(B) = \sum_i p_i$ not necessary greater than 0.9, no halfspace is defined (The solution gives an example!). Hence the sublevel set is not convex, not quasiconvex. ■

(h) $f(p) = \inf\{\beta - \alpha \mid \mathbf{prob}(\alpha \leq x \leq \beta) \geq 0.9\}$. Since β and α are integers, the function cannot be convex or concave. ■ If p's satisfy $f(p) \leq \gamma$, then any $\beta - \alpha$ with wider range does not necessary contain the 90% probability using the same p, hence not necessary quasiconvex. Now disprove quasiconvexity by an example. Suppose $n = 3, a_1 = 1, a_2 = 2, a_3 = 3, p' = (p'_1, p'_2, p'_3) = (0.1, 0.2, 0.7)$, $p'' = (0.7, 0.2, 0.1)$, $f(p') = 1$, $f(p'') = 1$. Given $\gamma = 1$, then both p' and p'' are in the sublevel set, but $p'\,'' = 0.5{*}p' + 0.5{*}p'' = (0.4, 0.2, 0.4)$ with $f(p''') = 2$ is not in the sublevel set, hence not quasiconvex. ■ If p's satsify $f(p) \geq \gamma$, then any $\beta - \alpha$ with narrower range satisfies $\mathbf{prob}(\alpha \leq x \leq \beta) < 0.9$, which is quasiconcave. ■

- 3.25 ***Maximum probability distance between distributions.***

$$d_{mp}(p, q) = \max\{|\mathbf{prob}(p, C) - \mathbf{prob}(q, C)| \mid C \subseteq \{1, \cdots, n\}\} \tag{8}$$

where the two probability vectors $p, q \in \mathcal{R}^n_+$, and $\mathbf{prob}(p, C) = \sum_{i \in C} p_i$.
(a) Find an expression for d_{mp} involving $||p - q||_1 = \sum_{i=1}^n |p_i - q_i|$.
(b) Show d_{mp} is convex on $\mathcal{R}^n \times \mathcal{R}^n$.

Proof
(a) Suppose $C^+ = \{i \mid p_i \geq q_i,\}$, $C^- = \{i \mid p_i < q_i\}$, $i \in 1, \cdots, n$. Since $\mathbf{prob}(p, C^+) + \mathbf{prob}(p, C^-) = 1$, then $\mathbf{prob}(p, C^+) + \mathbf{prob}(q, C^+) = 1 - \mathbf{prob}(p, C^-) - (1 - \mathbf{prob}(q, C^-)) = -(\mathbf{prob}(p, C^-) - \mathbf{prob}(q, C^-))$ $d_{mp}(p, q) = ||p - q||_1^{C^+} = \sum_{i \in C^+} |p_i - q_i|$ ■

(b) Since $||p - q||_1^{C^+} = \sum_{i \in C^+} |p_i - q_i| = 1^T p - 1^T q$ is a linear function of the domain, hence convex. To make it even more simplified, notice that since $\sum_{i \in C^+}(p_i - q_i) = -\sum_{i \in C^-}(p_i - q_i)$, then $d_{mp}(p, q) = 1/2 \sum_{i \in C^+}(p_i - q_i) + 1/2 \sum_{i \in C^+}(p_i - q_i) = 1/2 \sum_{i \in C^+}(p_i - q_i) - 1/2 \sum_{i \in C^-}(p_i - q_i) = 1/2 \sum_{i \in C^+}(p_i - q_i) + 1/2 \sum_{i \in C^-}(q_i - p_i) = 1/2 \sum_i |p_i - q_i| = 1/2||p - q||_1$. ***Maximum probability distance between distributions is*** $1/2||p - q||_1$ ■

- 3.26 ***Functions of eigenvalues***
  Let $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$, where $X \in S^n$. Several convex or concave functions of the eigenvalues:
  - **Maximum (minimum) eigenvalue of a symmetric matrix is convex (concave)**. $f(X) = \lambda_{\max}(X) = \sup\{y^T X y \mid ||y||_2\}$. Given $y = y_1$, $f_{y_1}(X) = y_1^T X y_1$ is linear for all X. Given a point X, there are infinite numbers of linear functions indexed by unit vectors y, and $f(X)$ is just finding the function that gives the maximum value. Hence it is a pointwise maximum of linear functions, which is convex.
  - **Sum of the eigenvalues (trace) is linear**
  - **Sum of the inverses of the eigenvalues (trace of the inverse) is convex on** S^n_{++}
  - **Geometric mean** $(\det X)^{1/n}$**, and the logorithm of the product of the eigenvalues** $\log \det X$