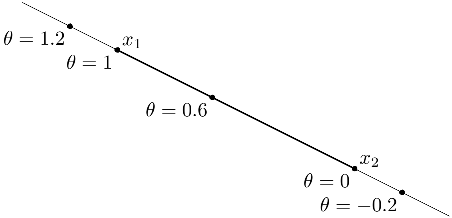
Operations Research
05. Convex Sets & Functions

Affine Set

- **line** through x_1, x_2: all points of the form $x = \theta x_1 + (1 - \theta)x_2$, $\theta \in \mathbb{R}$

- **affine set** contains the line through any two distinct points in the set

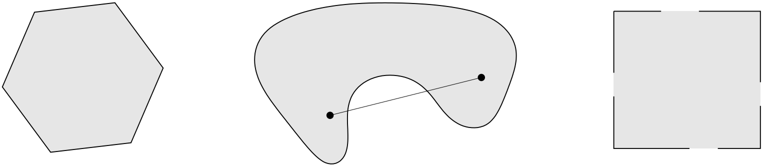
- e.g. solution set of linear equations $\{x \mid Ax = b\}$; every affine set can be expressed as solution set of system of linear equations



Convex Set

- **line segment** through x_1, x_2: all points of the form
  $x = \theta x_1 + (1-\theta)x_2$, $0 \leqslant \theta \leqslant 1$

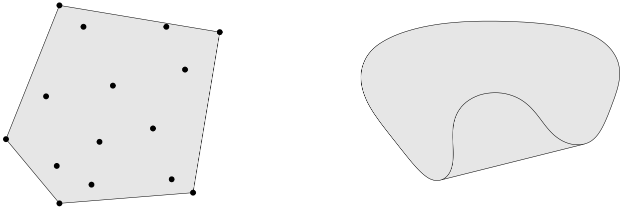
- **convex set** contains the line segment between any two distinct points in the set:
  $x_1, x_2 \in S \implies \forall\, 0 \leqslant \theta \leqslant 1,\ \theta x_1 + (1-\theta)x_2 \in S$



Convex Combination, Convex Hull

- **convex combination** of x_1, x_2, ..., x_k: any point x of the form $x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$ with $\theta_1 + \theta_2 + \cdots + \theta_k = 1$, $\theta_i \geqslant 0$

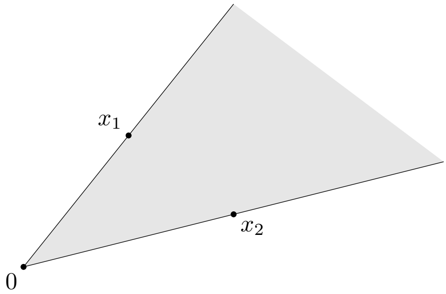
- **convex hull** $\operatorname{conv} S$: sets of all convex combinations of points in S



Convex Cone

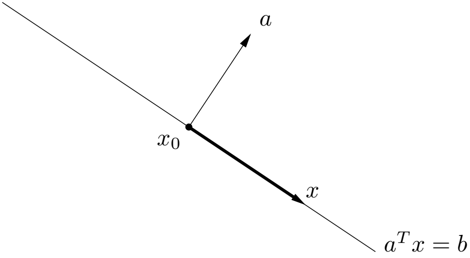
- **conic (nonnegative) combination** of x_1 and x_2: any point x of the form $x = \theta_1 x_1 + \theta_2 x_2$ with $\theta_i \geqslant 0$

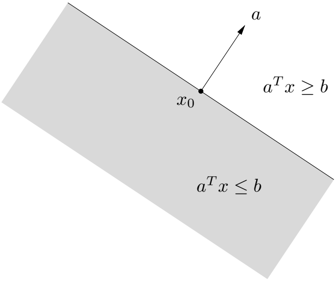
- **convex cone** set that contains all conic combinations of points in the set



Hyperplane, Halfspace

- **hyperplane**: set of the form $\{x \mid a^\top x = b\}$ with $a \neq 0$
- **halfspace**: set of the form $\{x \mid a^\top x \leqslant b\}$ with $a \neq 0$
- a: normal vector
  hyperplanes are affine and convex, halfspaces are convex

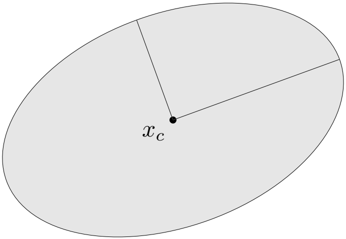




Euclidean Ball, Ellipsoid

- **(Euclidean) ball** with center x_c and radius r:
  $B(x_c, r) = \{x \mid \|x - x_c\|_2 \leqslant r\} = \{x_c + r\,u \mid \|u\|_2 \leqslant 1\}$

- **ellipsoid**: set of the form $\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leqslant 1\}$ with $P \in \mathsf{S}^n_{++}$ (P symmetric positive definite), or $\{x_c + A\,u \mid \|u\|_2 \leqslant 1\}$ with nonsingular A



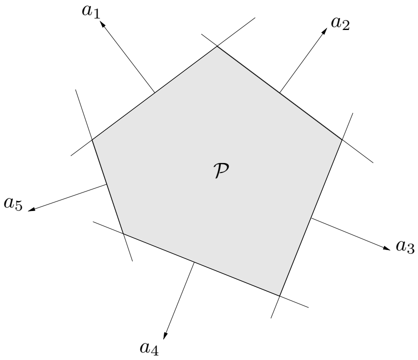
Norm Ball, Norm Cone

- **norm**: a function $\|\cdot\|$ that satisfies
  - $\|x\| \geqslant 0$; $\|x\| = 0 \iff x = 0$
  - $\|tx\| = |t|\|x\|$, $\forall\, t \in \mathbb{R}$
  - $\|x + y\| \leqslant \|x\| + \|y\|$
- **norm ball** with center x_c and radius r: $\{x \mid \|x - x_c\| \leqslant r\}$
- **norm cone**: $\{(x, t) \mid \|x\| \leqslant t\}$
- norm balls and norm cones are convex
- notation for different norms: $\|\cdot\|_2$, $\|\cdot\|_{\mathrm{symb}}$

Polyhedra

- **polyhedron**: solution set of finitely many linear equalities and inequalities $\{x \mid A\,x \preceq b,\ C\,x = d\}$, where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{p\times n}$, $\preceq$ is componentwise inequality

- intersection of finite number of halfspaces and hyperplanes

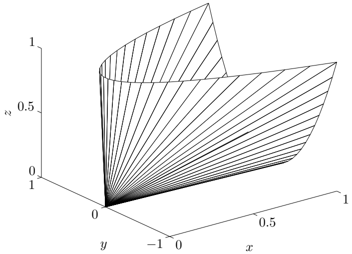


Positive Semidefinite Cone

- S^n: set of symmetric $n \times n$ matrices

- $\mathsf{S}^n_+ = \{X \in \mathsf{S}^n \mid X \succcurlyeq 0\}$: set of positive semidefinite (symmetric) $n \times n$ matrices; $X \in \mathsf{S}^n_+ \iff z^\top X z \geqslant 0 \; \forall\, z$; a convex cone, the **positive semidefinite cone**; Below: $\begin{pmatrix} x & y \\ y & z \end{pmatrix} \in \mathsf{S}^2_+$

- $\mathsf{S}^n_{++} = \{X \in \mathsf{S}^n \mid X \succ 0\}$: set of positive definite (symmetric) $n \times n$ matrices

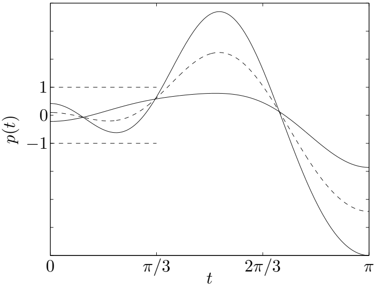


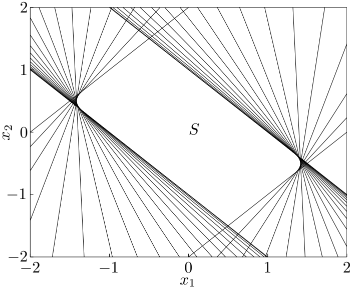
Showing a Set is Convex

- apply definition: $x_1, x_2 \in S \implies \theta x_1 + (1-\theta)x_2 \in S, \forall\, 0 \leqslant \theta \leqslant 1$
  recommended only for simple sets
- use convex functions (later)
- show that the set is obtained from other simple convex sets (e.g. hyperplanes, halfspaces, norm balls) by operations that preserve convexity:
  - intersection
  - affine mapping
  - perspective mapping
  - linear-fractional mapping
- mostly using last two

Intersection

- intersection of (any number of) convex sets is convex
- e.g. $S = \left\{ x \in \mathbb{R}^m \,\middle|\, |p(t)| \leqslant 1 \ \forall\, |t| \leqslant \frac{\pi}{3} \right\}$, $p(t) = \sum_{k=1}^{m} x_k \cos kt$ is convex by $S = \bigcap_{|t| \leqslant \frac{\pi}{3}} \{x \mid |p(t)| \leqslant 1\}$; intersection of convex slabs. Below: $m = 2$.





Affine Mappings

- suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is **affine**, i.e.

$$f(x) = A\,x + b \quad \text{with } A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$$

- the **image** of a convex set under f is convex:

$$S \subseteq \mathbb{R}^n \text{ convex} \implies f(S) = \{f(x)\,|\,x \in S\} \text{ convex}$$

- the **inverse image** of a convex set under f is convex:

$$C \subseteq \mathbb{R}^m \text{ convex} \implies f^{-1}(C) = \{x \in \mathbb{R}^n\,|\,f(x) \in C\} \text{ convex}$$

- e.g. scaling $aS + b = \{ax + b\,|\,x \in S\}$, a, $b \in \mathbb{R}$ is convex
- e.g. projection $\underset{x}{\operatorname{proj}}(S) = \{x\,|\,(x, y) \in S\}$ is convex

Perspective and Linear-Fractional Function

- **perspective function** $p : \mathbb{R}^{n+1} \to \mathbb{R}^n$:

$$p(x,t) = \frac{x}{t} \qquad \operatorname{dom} p = \{(x,t) \mid t > 0\}$$

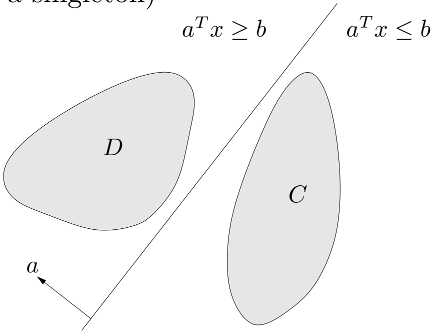
- **linear-fractional function** $f : \mathbb{R}^n \to \mathbb{R}^m$:

$$f(x) = \frac{A\,x + b}{c^\top x + d} \qquad \operatorname{dom} f = \{x \mid c^\top x + d > 0\}$$

- images and inverse images of convex sets under perspective and linear-fractional functions are all convex

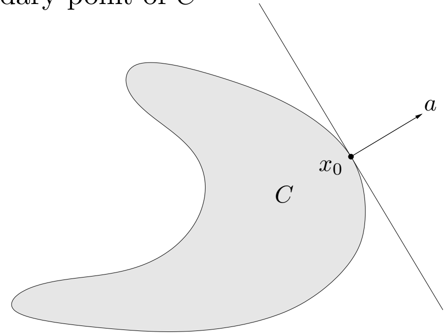
Separating Hyperplane Theorem

- if C, D are nonempty disjoint ($C \cap D = \emptyset$) convex sets, $\exists\, a \neq 0$, b s.t. $a^\top x \leqslant b$ for $x \in C$, $a^\top x \geqslant b$ for $x \in D$
- the hyperplane $\{x \mid a^\top x = b\}$ **separates** C and D
- strict separating requires additional assumptions (e.g. C is closed; D is a singleton)



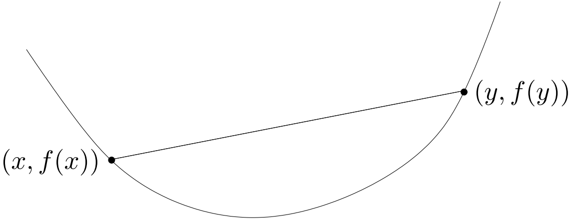
Supporting Hyperplane Theorem

- suppose x_0 is a boundary point of $C \subseteq \mathbb{R}^n$
- **supporting hyperplane** to C at x_0: $\{x \mid a^\top x = a^\top x_0\}$, where $a \neq 0$ and $a^\top x \leqslant a^\top x_0 \ \forall\, x \in C$.
- if C is convex, then there exists a supporting hyperplane at every boundary point of C



Convex Function

- $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if $\operatorname{dom} f$ is convex and $\forall\, x,\, y \in \operatorname{dom} f$, $0 \leqslant \theta \leqslant 1$, $f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y)$
- $f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if $\operatorname{dom} f$ is convex and $\forall\, x,\, y \in \operatorname{dom} f$, $x \neq y$, $0 < \theta < 1$, $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$
- f is **concave** if $-f$ is convex



Example Functions on $\mathbb{R}$

- convex functions
  - affine: $ax + b$, $\forall\, a,\ b \in \mathbb{R}$
  - exponential: e^{ax}, $\forall\, a \in \mathbb{R}$
  - power: x^α on $x > 0$, $\forall\, \alpha \geqslant 1 \ \vee\ \alpha \leqslant 0$
  - power of absolute value: $|x|^\alpha$, $\forall\, \alpha \geqslant 1$
  - positive part (relu): $\max\{x, 0\}$
- concave functions
  - affine: $ax + b$, $\forall\, a,\ b \in \mathbb{R}$
  - power: x^α on $x > 0$, $\forall\, 0 \leqslant \alpha \leqslant 1$
  - logarithm: $\log x$ on $x > 0$
  - entropy: $-x \log x$ on $x > 0$
  - negative part: $\min\{x, 0\}$

Example Convex Functions on $\mathbb{R}^n$

- affine: $a^\top x + b$
- any norm
  - $\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$, $\forall\, p > 1$
  - $\|x\|_\infty = \max\left\{|x_1|, |x_2|, \ldots, |x_n|\right\}$
- sum of squares: $\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$
- max function: $\max(x) = \max\left\{x_1, x_2, \ldots, x_n\right\}$
- softmax / log-sum-exp: $\log\left(e^{x_1} + e^{x_2} + \cdots + e^{x_n}\right)$

Example Functions on $\mathbb{R}^{m\times n}$

- Let $X \in \mathbb{R}^{m\times n}$ be the variable
- general affine function

$$f(X) = \mathrm{tr}(A^\top X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b, \quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}$$

- spectral norm (maximum singular value) is convex:

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = \sqrt{\lambda_{\max}(X^\top X)}$$

- log determinant is concave:

$$f(X) = \log\det X, \quad X \in \mathsf{S}^n_{++}$$

Extended-Value Extension

- suppose f is convex on $\mathbb{R}^n$
- its extended-value extension $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is defined as

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

- this often simplifies notation; e.g. the condition

$$0 \leqslant \theta \leqslant 1 \implies \tilde{f}(\theta x + (1-\theta)y) \leqslant \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$$

(as an inequality in $\mathbb{R} \cup \{\infty\}$), means the same as the two conditions combine

  - $\operatorname{dom} f$ is convex
  - $x,\, y \in \operatorname{dom} f,\ 0 \leqslant \theta \leqslant 1 \implies$
    $f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y)$

Restriction of a Convex Function to a Line

- $f : \mathbb{R}^n \to \mathbb{R}$ is convex (concave) $\iff$ $g : \mathbb{R} \to R$,

$$g(t) = f(x + t\,v), \quad \mathrm{dom}\, g = \{t \,|\, x + t\,v \in \mathrm{dom}\, f\}$$

  is convex (concave) in t for all $x \in \mathrm{dom}\, f$ and $v \in \mathbb{R}^n$

- useful for checking convexity / concavity of multivariate f; e.g. to check the concavity of log determinant: Let $X \in \mathsf{S}^n_{++}$, $V \in \mathsf{S}^n$,

$$\begin{aligned} g(t) = f(X + t\,V) &= \log\det(X + t\,V) \\ &= \log\det\left(X^{\frac{1}{2}}(I + t\,X^{-\frac{1}{2}}V X^{-\frac{1}{2}})X^{\frac{1}{2}}\right) \\ &= \log\det X + \log\det\left(I + t\,X^{-\frac{1}{2}}V X^{-\frac{1}{2}}\right) \\ &= \log\det X + \sum_{i=1}^{n} \log(1 + t\lambda_i) \end{aligned}$$

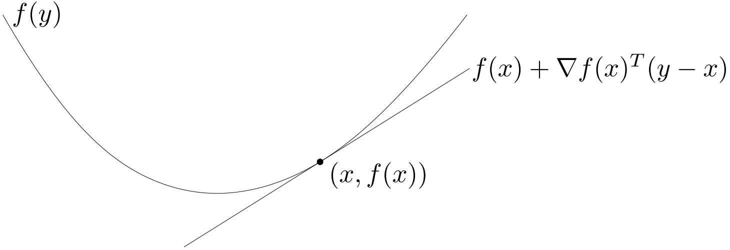
  where λ_i are the eigenvalues of $X^{-\frac{1}{2}}V X^{-\frac{1}{2}}$; g is concave in t

First-Order Condition

- $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable if $\mathrm{dom}\, f$ is open and the gradient ∇f exists at each $x \in \mathrm{dom}\, f$.

- **first-order condition** differentiable f with convex domain is convex $\iff f(y) \geqslant f(x) + \nabla f(x)^\top (y - x), \ \forall\, x,\, y \in \mathrm{dom}\, f$

- first order Taylor approximation of convex f is a **global underestimator** of f



Second-Order Condition

- $f : \mathbb{R}^n \to \mathbb{R}$ is **differentiable** if $\operatorname{dom} f$ is open and the Hessian matrix $\nabla^2 f \in \mathsf{S}^n$ exists at each $x \in \operatorname{dom} f$:

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

- **second-order condition** for twice differentiable f with convex domain is convex:
  - f is convex $\iff$ $\nabla^2 f \succcurlyeq 0,\ \forall\, x \in \operatorname{dom} f$
  - $\nabla^2 f \succ 0,\ \forall\, x \in \operatorname{dom} f \implies f$ is strictly convex

Examples

- **quadratic function**: $f(x) = \dfrac{1}{2} x^\top P x + q^\top x + r$ with $P \in \mathsf{S}^n$

$$\nabla f(x) = Px + q, \quad \nabla^2 f(x) = P$$

  convex if $P \succeq 0$ (concave if $P \preceq 0$)

- **least-squares objective**: $f(x) = \|A\,x - b\|^2$

$$\nabla f(x) = 2A^\top (A\,x - b), \quad \nabla^2 f(x) = 2A^\top A$$

  convex for any A

- **quadratic-over-linear function**: $f(x, y) = \dfrac{x^2}{y}$, $y > 0$

$$\nabla f(x,y) = \begin{pmatrix} \frac{2x}{y} & -\frac{x^2}{y^2} \end{pmatrix}, \quad \nabla^2 f(x,y) = \frac{2}{y^3} \begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}$$

  convex for $y > 0$

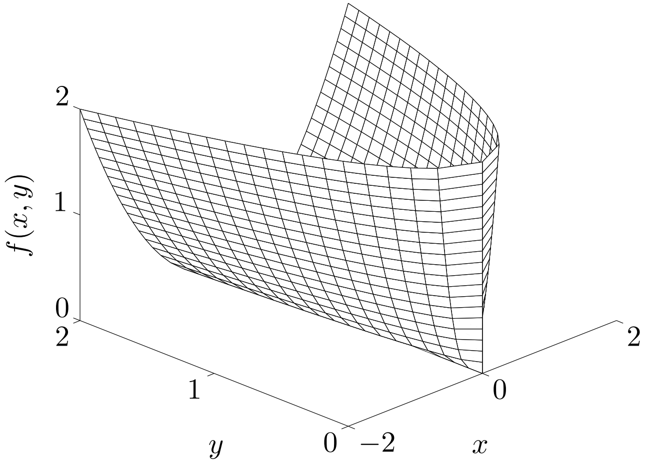


Figure 1: Graph of quadratic-over-linear function $f(x, y) = \dfrac{x^2}{y}$, $y > 0$

- **log-sum-exp function**: $f(x) = \log\Big(\sum_{k=1}^{n} e^{x_k}\Big)$ is convex:

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^\top z}\,\mathrm{diag}(z) - \frac{1}{(\mathbf{1}^\top z)^2} z z^\top, \quad z_k = e^{x_k}$$

- to show that $\nabla^2 f(x) \succcurlyeq 0$, one must verify $v^\top \nabla^2 f(x)\, v \geqslant 0\ \forall\, v$:

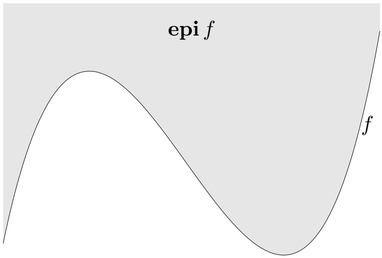
$$v^\top \nabla^2 f(x)\, v = \frac{\big(\sum_k z_k v_k^2\big)\big(\sum_k z_k\big) - \big(\sum_k v_k z_k\big)^2}{\big(\sum_k z_k\big)^2} \geqslant 0$$

  by Cauchy-Schwarz inequality $\big(\sum_k z_k v_k^2\big)\big(\sum_k z_k\big) \geqslant \big(\sum_k v_k z_k\big)^2$

- **geometric-mean function**: $f(x) = \Big(\prod_{k=1}^{n} x_k\Big)^{\frac{1}{n}}$ on $x \succ 0$ is concave

Epigraph, Sublevel Set

- α**-sublevel set** of $f : \mathbb{R}^n \to \mathbb{R}$: $C_\alpha = \{x \in \mathrm{dom}\, f \mid f(x) \leqslant \alpha\}$
- sublevel sets of convex functions are convex sets
- **epigraph** of $f : \mathbb{R}^n \to \mathbb{R}$:
  $\mathrm{epi}\, f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \mathrm{dom}\, f,\ f(x) \leqslant t\}$
- f is convex $\iff$ $\mathrm{epi}\, f$ is a convex set



Jensen's Inequality

- **basic form**: if f is convex, then for x, $y \in \operatorname{dom} f$, $0 \leqslant \theta \leqslant 1$

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y)$$

- **extension**: if f is convex and z is a random variable on $\operatorname{dom} f$,

$$f(\,\mathsf{E}\, z\,) \leqslant \mathsf{E}\, f(z)$$

- basic form is special case with discrete distribution

$$\mathsf{P}\{z = x\} = \theta, \quad \mathsf{P}\{z = y\} = 1 - \theta$$

- e.g. for $z \sim \mathsf{N}(\mu, \sigma^2)$, let $f(x) = e^x$, then

$$f(\,\mathsf{E}\, z\,) = f(\mu) = e^{\mu} \leqslant e^{\mu + \frac{\sigma^2}{2}} = \mathsf{E}\, f(z)$$

Showing Convexity of a Function

- apply definition (often simplified by restricting to a line)
- for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
- show that f is obtained from simple convex functions by operations that preserve convexity
  - nonnegative multiple, sum, integral
  - composition with affine function
  - pointwise maximum and supremum
  - partial minimization
  - composition
  - perspective

Nonnegative Multiple, Sum, Integral

- **nonnegative multiple**: αf is convex if f is convex and $\alpha \geqslant 0$
- **sum**: $f_1 + f_2$ is convex if f_1, f_2 is convex
- **infinite sum**: if each of f_i is convex, then $\sum_{i=1}^{\infty} f_i$ is convex
- **integral**: if $f(x, \alpha)$ is convex in x for each $\alpha \in \mathcal{A}$, then

  $$\int_{\alpha \in \mathcal{A}} f(x, \alpha) \, \mathrm{d}\alpha$$

  is convex
- analogous rules for concave functions

Composition with Affine Function

- $f(Ax+b)$ is convex if f is convex
- e.g.
  - log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log\left(b_i - a_i^\top x\right)$$
$$\operatorname{dom} f = \{x \mid a_i^\top x < b_i,\ i = 1, 2, \ldots, m\}$$

  - norm approximation error (any norm)

$$f(x) = \|Ax - b\|$$

Pointwise Maximum

- $f(x) = \max\{f_1(x), f_2(x), \ldots, f_m(x)\}$ is convex if each f_i is convex

- e.g.

  - piecewise linear function

$$f(x) = \max_i \left(a_i^\top x + b_i\right)$$

  - sum of r largest components of $x \in \mathbb{R}^n$

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

    where $x_{[i]}$ is i-th largest component of x. Note that

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n\}$$

Pointwise Supremum

- $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$

- e.g.

  – distance to farthest point in a set C

  $$f(x) = \sup_{y \in C} \|x - y\|$$

  – maximum eigenvalue of symmetric matrix

  $$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^\top X\, y, \quad X \in \mathsf{S}^n$$

  – support function of a set C

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

Partial Minimization

- the function $g(x) = \inf_{y \in C} f(x, y)$ is called the **partial minimization** of f w.r.t. y
- if $f(x, y)$ is convex in (x, y) and C is a convex set, then partial minimization g is convex
- e.g.
  - let $f(x, y) = x^\top A\,x + 2x^\top B\,y + y^\top C\,y$ with $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \succcurlyeq 0$, $C \succ 0$; minimizing over y gives
  $$g(x) = \inf_{y \in C} f(x, y) = x^\top (A - BC^{-1}B^\top)\,x$$
  g is convex, hence Schur complement $A - BC^{-1}B^\top \succcurlyeq 0$
  - distance to a convex set S
  $$\mathrm{dist}(x, S) = \inf_{y \in S} \|x - y\|$$

Composition with Scalar Functions

- composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ is $f(x) = h(g(x))$ $(f = h \circ g)$

- composition f is convex if
  - g convex, h convex, $\tilde{h}$ nondecreasing; or
  - g concave, h convex, $\tilde{h}$ nonincreasing

- proof for $n = 1$, differentiable g, h

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

- e.g.
  - $f(x) = e^{g(x)}$ is convex if g is convex
  - $f(x) = \dfrac{1}{g(x)}$ is convex if g is concave and positive

Composition: General

- composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^n \to \mathbb{R}$ is $f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$
- composition f is convex if h is convex and for each i, one of the following holds:
  - g_i convex, $\tilde{h}$ nondecreasing in its i-th argument
  - g_i concave, $\tilde{h}$ nonincreasing in its i-th argument
  - g_i affine
- e.g.
  - $\log\left(\sum_{i=1}^{m} e^{g_i(x)}\right)$ is convex if each g_i is convex
  - $\dfrac{p(x)^2}{q(x)}$ is convex if p is nonnegative and convex and q is positive and concave

Perspective

- perspective of $f : \mathbb{R}^n \to \mathbb{R}$ is the function $g(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined as

$$g(x,t) = t\, f\Big(\frac{x}{t}\Big), \quad \operatorname{dom} g = \Big\{(x,t) \;\Big|\; \frac{x}{t} \in \operatorname{dom} f,\ t > 0\Big\}$$

- g is convex if f is convex

- e.g.

  - $f(x) = x^\top x$ is convex, so $g(x,t) = \dfrac{x^\top x}{t}$ is convex if $t > 0$
  - $f(x) = -\log x$ is convex, so the **relative entropy**

$$g(x,t) = t\, \log t - t\, \log x$$

    is convex on $x > 0$, $t > 0$

Convexity Verification: An Example

- test the convexity of $f(x, y) = \dfrac{(x-y)^2}{1 - \max(x, y)}$, $x < 1$, $y < 1$

- x, y, and 1 are affine

- $\max(x, y)$ is convex; $x - y$ is affine

- $1 - \max(x, y)$ is concave

- $\dfrac{u^2}{v}$ is convex, monontone decreasing in v for $v > 0$

- f is composition of $\dfrac{u^2}{v}$ with $u = x - y$, $v = 1 - \max(x, y)$, hence convex

Convexity Verification: A Caveat

- test the convexity of $f(x) = \sqrt{1+x^2}$
- $\sqrt{\cdot}$ is concave
- 1, x^2 are convex
- $\sqrt{1+x^2}$ is ... indefinite ?
- but, note that $\|\cdot\|_2$ is convex
- $\sqrt{1+x^2}$ can be represented as the 2-norm of vector $(1,x)$ — $\|(1,x)\|_2$, hence is convex
- The general composition rules are only sufficient, not necessary