Lecture 2: Convex Sets

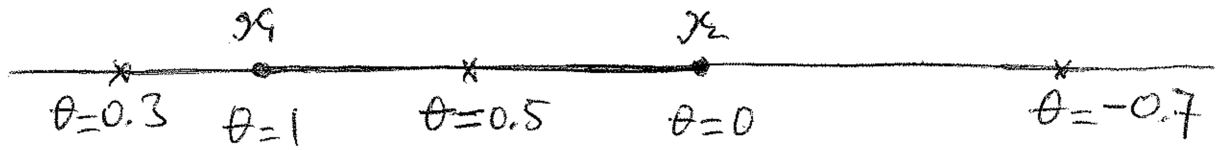
This topic provides definition of convex sets, operations on sets that preserve convexity, and examples of important sets.

1. Definitions.

+) Affine sets:

- Line: For 2 points $x_1, x_2 \in \mathbb{R}^n$, the line through these 2 points contains all points of the form

$$x = \theta x_1 + (1-\theta) x_2 \quad , \quad \theta \in \mathbb{R}.$$



- Affine set: A set $C \subseteq \mathbb{R}^n$ is affine if it contains all lines through any two distinct points in C.

$$\forall\, x_1, x_2 \in C \text{ and } \theta \in \mathbb{R} \rightarrow \theta x_1 + (1-\theta) x_2 \in C.$$

This idea can be generalized to an affine combination of more than two points: $\sum_{i=1}^{k} \theta_i x_i \in C$, $\sum_{i=1}^{k} \theta_i = 1$, $x_i \in C$.

- Example: The solution set of linear equations is an affine set. Conversely, every affine set can be expressed as the solution set of a system of linear equation.

$$C = \{ x \mid Ax = b \} \,, \quad A \in \mathbb{R}^{m \times n} \,, \; b \in \mathbb{R}^n$$

+) Convex sets:

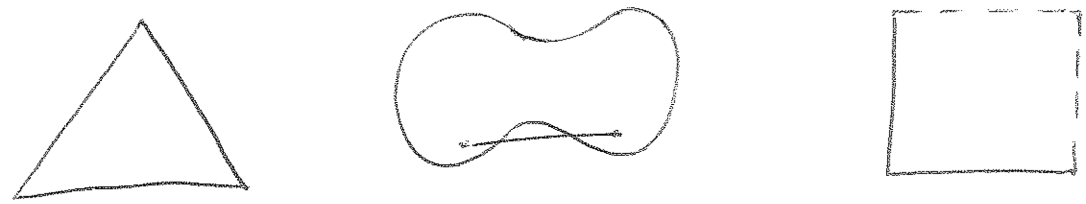
- Line segment between x_1 and x_2 contains all points

$$x = \theta x_1 + (1-\theta) x_2 \quad \text{with} \quad 0 \leq \theta \leq 1.$$

x_1 x_2

- Convex set: A set $C \subseteq \mathbb{R}^n$ is convex if it contains the line segment between any two points in the set.

$$x_1, x_2 \in C, \quad 0 \le \theta \le 1 \longrightarrow \theta x_1 + (1-\theta) x_2 \in C.$$



convex non-convex non-convex (boundary)

- Convex combination:
A convex combination of k points $x_1, \dots, x_k \in \mathbb{R}^n$ is any point in $\mathbb{R}^n$ of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k$$

where $\theta_1 + \theta_2 + \dots + \theta_k = 1, \ \forall \theta_i \ge 0.$

This is like an weighted average (the weights are non-negative)

+ This idea can be generalized to include infinite sums and integrals. In the most general form it is like a probability distribution.

If C is a convex set and

$$\left.\begin{array}{l} x_1, x_2 \dots \in C, \\ \theta_i \ge 0, \ \sum_{i=1}^{\infty} \theta_i = 1 \end{array}\right. \longrightarrow \sum_{i=1}^{\infty} \theta_i x_i \in C. \text{ (provided the sum converges).}$$

If $f(x)$ is a pdf: $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) \ge 0 \ \forall x \in C$ and $\int_C f(x)\,dx = 1$

then $\int_C f(x)\, x\, dx \in C$. for convex set C.

Most generally: If C is convex and x is a random vector in C then $Ex \in C$.

- Convex hull: For a general set S, the convex hull of S is the set of all convex combinations of points in S.

$$\operatorname{conv} S = \{\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k \mid x_i \in S, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1\}$$

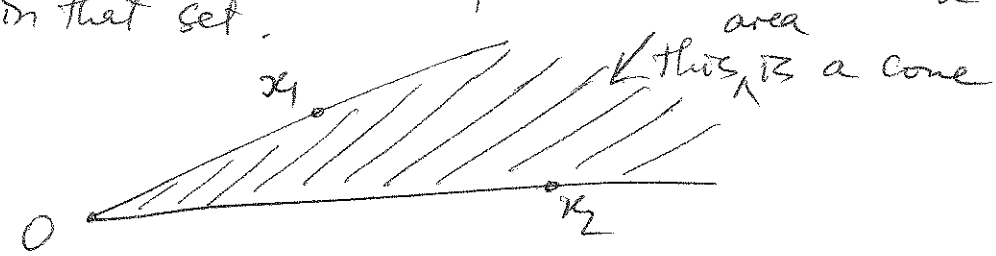
$\operatorname{conv} S$ is the smallest convex set that contains S.

+) Cones:

- Conic combination of x_1 and x_2 is any point of the form

$$\theta_1 x_1 + (1-\theta_1) x_2 \quad \text{with } \theta_1 \geq 0, \theta_2 \geq 0.$$

- A set C is a convex cone if it contains all conic combinations of points in that set.



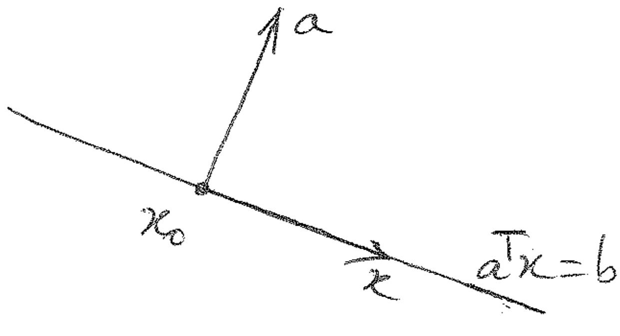
2. Some important examples:

+) Hyperplanes and halfspaces:

- A hyperplane is a set of the form $\{x \mid a^T x = b\}$ ($a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$)

A hyperplane can also be written as

$$\{x \mid a^T(x - x_0) = 0\} \quad \text{for an } x_0 \in \text{hyperplane} \ (a^T x_0 = b).$$

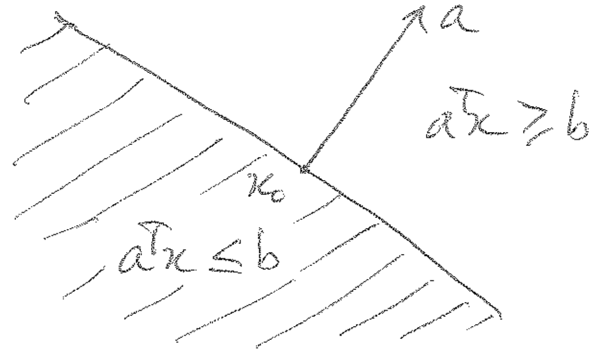


Thus a hyperplane is orthogonal to the vector a (a is the normal vector)
The constant b defines the offset of the hyperplane from the origin.

- A hyperplane divides $\mathbb{R}^n$ into two halfspaces.

  A closed halfspace $= \{x \mid a^T x \leq b\}$ $(a \in \mathbb{R}^n, a \neq 0)$.

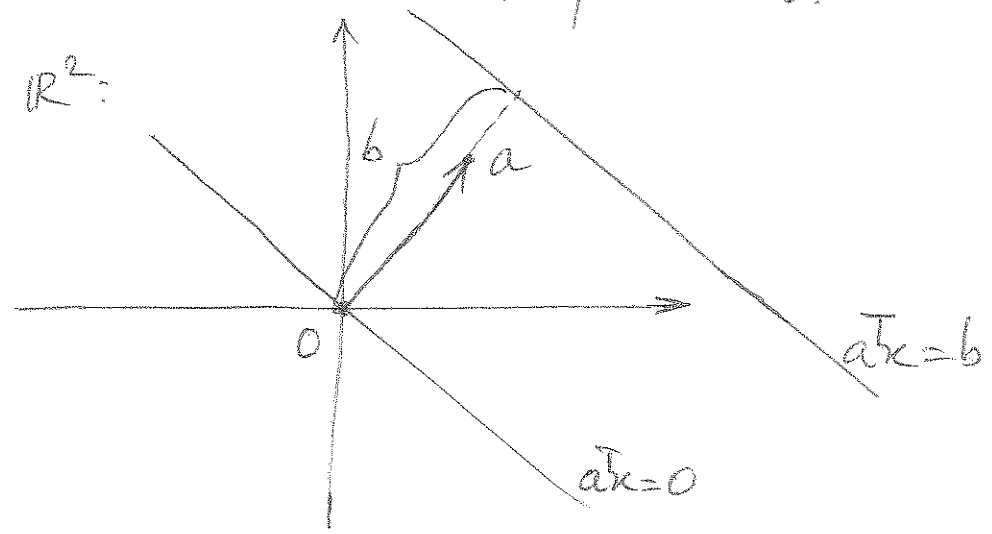
  A open halfspace $= \{x \mid a^T x < b\}$.



- A hyperplane is affine and convex.
  A halfspace is convex but not affine.
  Neither hyperplane nor halfspace is a subspace.

[Side notes: A vector space is close under addition & scalar multiplication. A subspace contains all linear combination of its points and is also closed under addition and scalar multiplication.]

- Example in $\mathbb{R}^2$:



+) Euclidean balls and ellipsoids:

- An Euclidean ball has the form:

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\} \qquad x, x_c \in \mathbb{R}^n, r \in \mathbb{R}$$
$$= \{x \mid (x-x_c)^T (x-x_c) \le r^2\}.$$

x_c : center
r : radius $\rightarrow B(x_c, r) = \{x = x_c + ru \mid \|u\|_2 \le 1\}$.

An Euclidean ball is convex (proved based on definition)

- An ellipsoid is a generalization of a ball as

$$\mathcal{E} = \{x \mid (x-x_c)^T P^{-1} (x - x_c) \le 1\}.$$

where $P = P^T \succ 0$ is a positive definite, symmetric matrix.

A ball is where $P = r^2 I$.

An ellipsoid can also be written as

$$\mathcal{E} = \{x = x_c + Au \mid \|u\|_2 \le 1\}.$$

where $A \in \mathbb{R}^{n \times n}$ (square), $A^T A = P^{-1}$

Ellipsoids are convex.

+) Norm balls and norm cones:

A norm on $\mathbb{R}^n$ is a function $\|.\| : \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies

i) $\|x\| \ge 0$, "=" iff $x = 0$

ii) $\|tx\| = |t| . \|x\| \quad \forall t \in \mathbb{R}$

iii) $\|x + y\| \le \|x\| + \|y\|$

A norm generalizes the concept of distance in n dimensions.

ℓ_p-norm:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Some common norms:

ℓ_1-norm: $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$

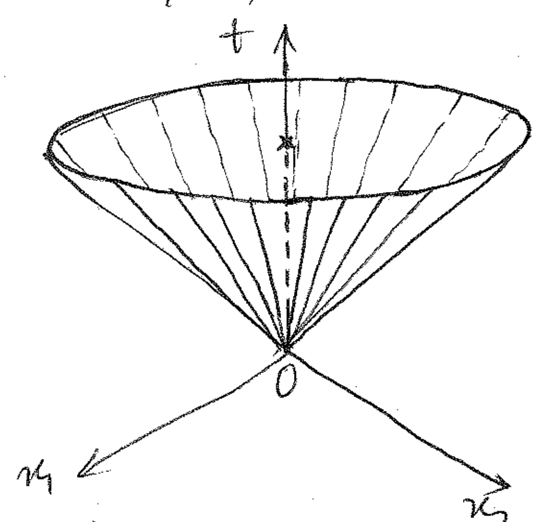
ℓ_∞-norm: $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$.

ℓ_2-norm: $\|x\|_2 = \left(|x_1|^2 + |x_2|^2 + \dots + |x_n|^2\right)^{1/2}$ Euclidean norm

- A norm ball with center x_c and radius r is a set of the form
$$\{x \mid \|x - x_c\| \le r\}.$$
This is a set in $\mathbb{R}^n$.

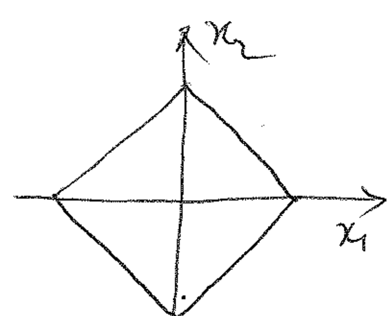
- A norm cone is a set of the form:
$$C = \{(x,t) \mid \|x\| \le t\}.$$
Note that a cone is a set in $\mathbb{R}^{n+1}$ ($n+1$ dimensions)

<u>Example</u>: The second-order cone or ice-cream cone
$$C = \{(x,t) \in \mathbb{R}^{n+1} \mid \|x\|_2 \le t\}$$

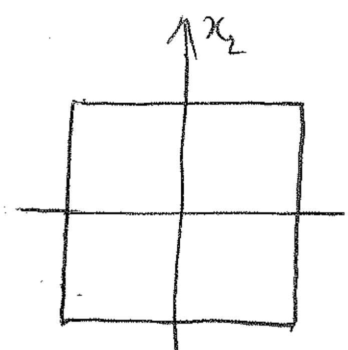


quadratic cone in $\mathbb{R}^3$.

Cross-section at a fixed t is a norm ball.



ℓ_1-ball in $\mathbb{R}^2$



ℓ_∞ ball in $\mathbb{R}^2$

+) Polyhedra:

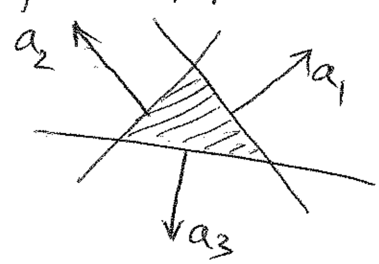
- A polyhedron is the solution set of a finite number of linear inequalities and equalities.

$$P = \{x \mid Ax \preceq b, \ Cx = d\} \qquad A \in \mathbb{R}^{m \times n}, \ C \in \mathbb{R}^{p \times n}$$

component-wise inequality

P is the intersection of a finite number of halfspaces and hyperplanes.

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}, \quad C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix}$$



- Simplex is a special case of polyhedra:
Let $v_0, \dots v_k \in \mathbb{R}^n$ be $k+1$ points that are affinely independent, which means $v_1 - v_0, v_2 - v_0, \dots, v_k - v_0$ are linearly independent. Then the simplex determined by these points are given as

$$C = \mathrm{conv}\{v_0, \dots, v_k\} = \{x = \sum_{i=0}^{k} \theta_i x_i \mid \sum_{i=0}^{k} \theta_i = 1, \theta \succeq 0\}$$

→ this is a k-dimensional simplex in $\mathbb{R}^n$.

- The convex hull of a finite set $\{x_1, \dots, x_m\}$ is a polyhedron, but it is not simple to express it in the form of linear inequalities and equalities.

The representations, however, can be very different. For example: Consider the unit ball in l_∞ norm:

$$C = \{x \mid |x_i| \le 1, \ i = 1 \dots n\}$$ polyhedron with $2n$ linear inequalities.

$$C = \mathrm{conv}\{v_1, \dots, v_{2^n}\}$$ Convex hull of 2^n points whose components are ± 1.

+) <u>The positive semidefinite cone:</u>

Denote S^n as the set of symmetric $n \times n$ matrices.

$$S^n = \{ X \in \mathbb{R}^{n \times n} \mid X = X^T \}$$

S^n_+ : set of symmetric positive semidefinite matrices

S^n_{++} : " " definite matrices.

$$S^n_+ = \{ X \in S^n \mid X \succeq 0 \} \qquad (z^T X z \geq 0 \ \forall z \in \mathbb{R}$$

$$S^n_{++} = \{ X \in S^n \mid X \succ 0 \}$$

The set S^n_+ (S^n_{++}) is a convex cone.

If $\theta_1, \theta_2 \geq 0$, $A, B \subseteq S^n_+ \rightarrow \theta_1 A + \theta_2 B \subseteq S^n_+$

$$z^T(\theta_1 A + \theta_2 B) z = \theta_1 z^T A z + \theta_2 z^T B z \geq 0 \ \forall z \in \mathbb{R}^n$$

A symmetric matrix in S^n is a vector of size $\frac{1}{2} n(n+1)$.

3. <u>Operations that preserve convexity of sets</u>

+) <u>Intersections</u>: If S_1, S_2 convex $\rightarrow S_1 \cap S_2$ convex.

This property applies to an infinite number of sets.

$$S_\alpha \text{ convex } \forall \alpha \in A \longrightarrow \bigcap_{\alpha \in A} S_\alpha \text{ is convex.}$$

+) <u>Affine functions</u>: $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is affine if

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

If S is convex then $f(S)$ is also convex

$$f(S) = \{ f(x) \mid x \in S \}, \quad f: \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

The inverse image of S under an affine function is also convex
$$f^{-1}(S) = \{x \mid f(x) \in S\}, \quad f: \mathbb{R}^k \to \mathbb{R}^n$$

Examples: Scaling and translation.

$S \subseteq \mathbb{R}^n$ is convex $\to$ $\alpha S + a$ is convex $\forall \alpha \in \mathbb{R}, a \in \mathbb{R}^n$.

Projection: $S \subseteq \mathbb{R}^n \times \mathbb{R}^m$ is convex

$\to T = \{x_1 \in \mathbb{R}^n \mid (x_1, x_2) \in S \text{ for some } x_2 \in \mathbb{R}^m\}$

is convex.

Example: Linear matrix inequality (LMI)
$$A(x) = x_1 A_1 + x_2 A_2 + \dots + x_n A_n \preccurlyeq B, \quad A_i, B \in S^m$$
The solution set of an LMI is a convex set, since it is the inverse image of the positive semidefinite cone under the affine function $f: \mathbb{R}^n \to S^m$ as $f(x) = B - A(x) \succcurlyeq 0$.

+) Linear-fractional and perspective functions:

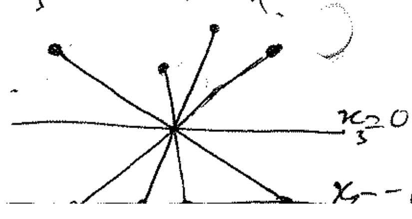
. Perspective function: $P: \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$
$$P(x,t) = \frac{x}{t}, \quad x \in \mathbb{R}^n, \ t \in \mathbb{R}_{++} \ (t > 0)$$
$P(x,t)$ normalizes vectors so the last component is 1, then drops this component.

The perspective image of a convex set is convex.

If C is convex $\to$ $P(C) = \{P(x) \mid x \in C\}$ is convex.

For the proof of this result see the textbook.

(pinhole camera)



- Linear-fractional functions: Composing the perspective function with an affine function.

Let $g : \mathbb{R}^n \to \mathbb{R}^{m+1}$ be affine

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix} \qquad A \in \mathbb{R}^{m\times n},\ c \in \mathbb{R}^n \quad b \in \mathbb{R}^m,\ d \in \mathbb{R}.$$

then
$$f(x) = \frac{Ax+b}{c^Tx+d}, \quad \mathrm{dom}(f) = \{x \mid c^Tx + d > 0\}.$$

f is a linear-fractional function or a projective function.

A linear fractional function preserves convexity. That is, if $C \in \mathrm{dom}(f)$ and C is convex $\to f(C)$ is convex.

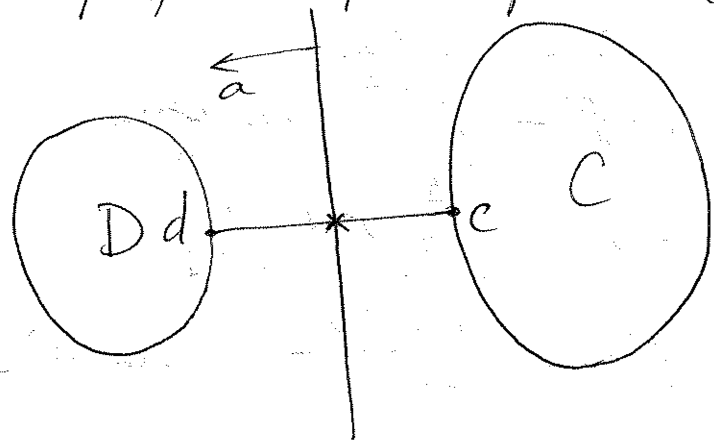
This result is straightforward by combining the properties of affine and perspective functions.

4. Seperating and supporting hyperplane theorems.

+) Separating hyperplane theorem:
Suppose two sets C and D are convex sets that do not intersect, $C \cap D = \emptyset$. Then there exists a hyperplane that separate them.
That is, $\exists\ a \neq 0$ and b such that
$$a^Tx \leq b \ \forall x \in C \quad \text{and} \quad a^Tx \geq b \ \forall x \in D$$
The hyperplane $\{x \mid a^Tx = b\}$ separates sets C and D. It is called a separating hyperplane.

+) The rigorous proof is technical and involves looking at cases when sets C and D have the "difference" that is open or closed and empty or not empty.
"Difference" set $S = \{x - y \mid x \in C,\ y \in D\}$ – see prob. 2.22

Here we show the proof ideas for a special case as follows:



Suppose $\exists$ points $c \in C$ and $d \in D$ such that

$$\|c-d\|_2 = \inf\{\|u-v\|_2 \mid u \in C, v \in D\}$$

c and d are two closest points on sets C & D according to the Euclidean distance measure (exist when, eg, C, D are closed and one set is bounded)

Then define

$$a = d-c, \quad b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$$

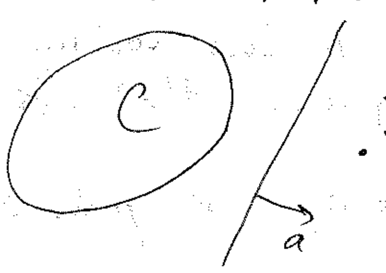
$\rightarrow$ a separating hyperplane is given by $\{x \mid a^T x = b\}$.

It comes down to showing that $f(x) = a^T x - b$ is nonpositive on C and nonnegative on D. This can be shown after some straightforward malnipuletion (using contradiction and convexity).

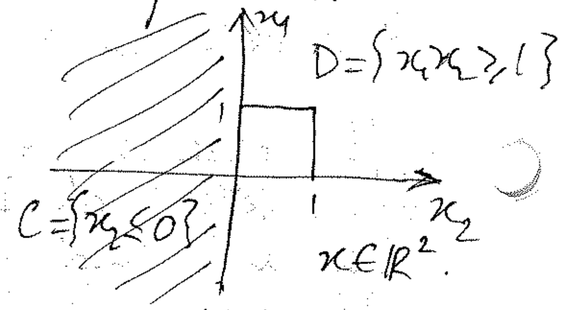
+) Strict separation:
Strict separation is a stronger condition that says

$$a^T x < b \;\; \forall\, x \in C \quad \text{and} \quad a^T x > b \;\; \forall\, x \in D.$$

Disjoint sets C & D donot guarantee strict separation.
Need additional conditions such as
C is closed and D is a single point.



strict separation



no strict separation.

+) Converse: The converse that if there exists a separating hyperplane then C and D are disjoint donot generally hold. Need additional conditions beyond convexity. For example:

If C and D convex sets, at least one is open then C and D are disjoint <u>iff</u> there exists a separating hyperplane.

<u>Example:</u> Feasibility of a system of strict linear inequality

$$Ax \prec b$$

This system is infeasible iff the following convex sets do not intersect:

$$C = \{b - Ax \mid x \in \mathbb{R}^n\}, \quad D = \mathbb{R}^m_{++} = \{y \in \mathbb{R}^m \mid y \succ 0\}$$

Here D is open.

This condition is equivalent to there exists a separating hyperplane between C and D, that is

$$\exists\, \lambda \in \mathbb{R}^m, \lambda \neq 0 \text{ and } \mu \in \mathbb{R} : \quad \begin{array}{l} \lambda^T y \leq \mu \text{ on } C \\ \lambda^T y \geq \mu \text{ on } D. \end{array}$$

i) $\lambda^T y \leq \mu$ on $C \iff \lambda^T(b - Ax) \leq \mu \quad \forall x \in \mathbb{R}^n$

$\iff \lambda^T b \leq \mu$ and $\lambda^T A = 0$.

ii) $\lambda^T y \geq \mu$ on $D \iff \lambda^T y \geq \mu \quad \forall\, y \succ 0$

$\iff \mu \leq 0$ and $\lambda^T \succeq 0$, $\lambda \neq 0$.

Putting together, $Ax \prec b$ is infeasible <u>iff</u> $\exists\, \lambda \in \mathbb{R}^m$ s.t.

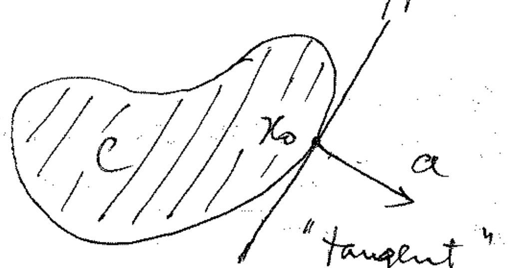
$$\lambda \neq 0, \ \lambda \succeq 0, \ A^T\lambda = 0 \text{ and } \lambda^T b \leq 0.$$

This is also a system of linear inequalities and equalities. These two systems form a pair of alternatives: for any given (A, b), exactly one of them is feasible.

+) Supporting hyperplane: This concept applies to any set, not necessarily convex sets.

Defn: Suppose $C \subseteq \mathbb{R}^n$ and x_0 is a point on the boundary of C. If there exists $a \neq 0$ such that $a^T x \leq a^T x_0 \ \forall x \in C$ then the hyperplane $\{x \mid a^T x = a^T x_0\}$ is called a supporting hyperplane to C at point x_0.

+ The supporting hyperplane theorem:
For any nonempty convex set C and any x_0 on its boundary, there exists a supporting hyperplane to C at x_0.



The idea is straightforward from seeing this hyperplane as separating int C and x_0.
↖ (the interior of C).

+ Partial converse:
If a set is closed, has nonempty interior and has a supporting hyperplane at every point on its boundary, then it is convex.

5. <u>A few words about minimum and minimal elements.</u>

In a vector space, not all elements can be ordered. For example, we cannot always directly compare two vectors or two matrices.
$x \prec y$ or $A \prec B$ only holds for certain pairs.

Because of this, we need to distinguish between minimum and minimal elements.

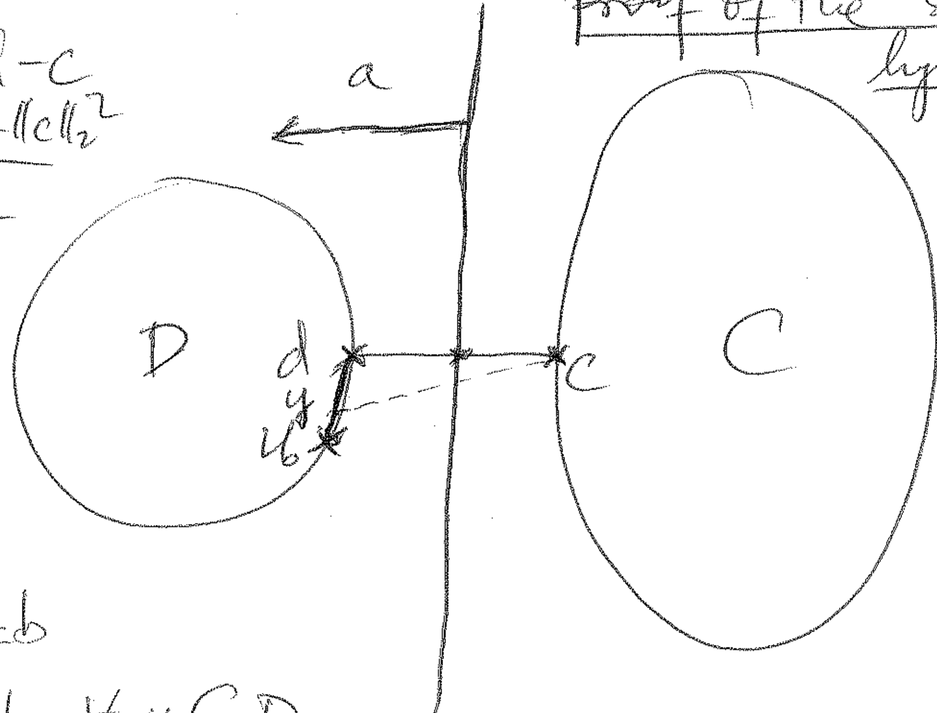
+) Minimum element of a set S:
$x \in S$ is the minimum element if $x \preceq y \ \forall y \in S$.
The minimum element is usually unique (if exists).

+) Minimal elements:
$x \in S$ is a minimal element if $y \in S$, $y \preceq x$ only if $y = x$.
There can be multiple minimal elements.

Proof of the separating hyperplane theorem

$$a = d - c$$

$$b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$$



$$a^T x = b$$

$$\rightarrow \begin{cases} a^T u \geq b & \forall u \in D \\ a^T v \leq b & \forall v \in C \end{cases}$$

$$u_0 \in D: \quad a^T u_0 < b$$

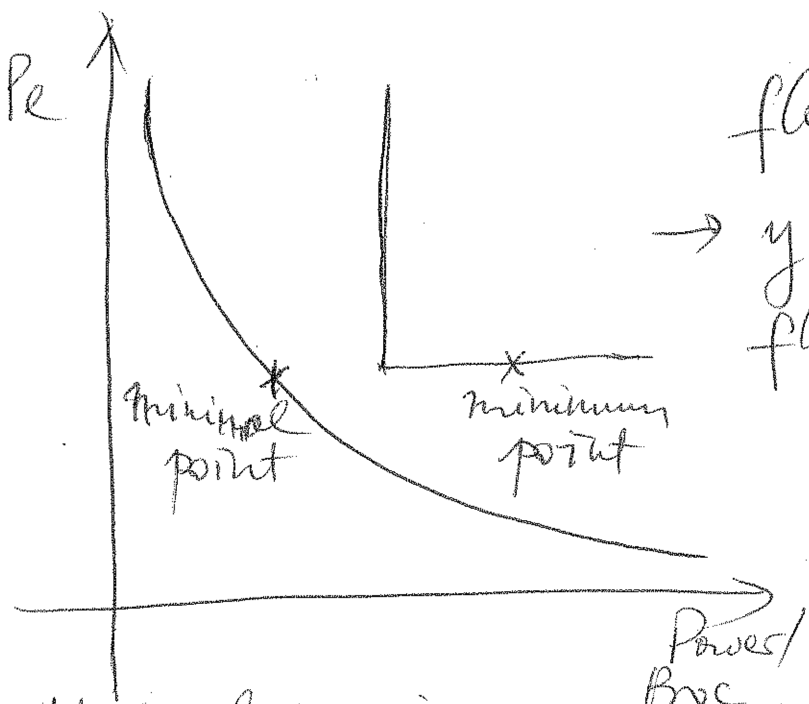
$$f(x) = (d-c)^T\left(x - \frac{d+c}{2}\right)$$

$$f(d) = (d-c)^T(d-c)\frac{1}{2} = \frac{1}{2}\|d-c\|_2^2 \geq 0$$

$$f(u_0) < 0$$

$$\rightarrow y = tu + (1-t)d$$

$$f(y) \leq t f(d) + (1-t) f(u_0) < t f(d) \leq f(d)$$



Minimal vs. minimum

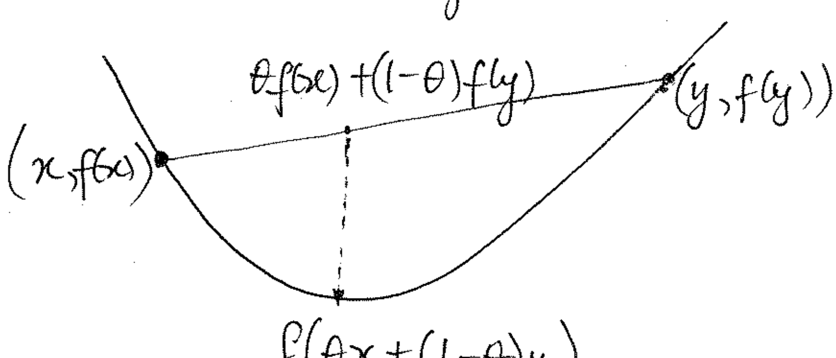
Lecture 4: Convex Functions

In this topic, we will focus on functions that are convex. We will cover definition and common examples, operations that preverse convexity of a function, conjugate functions which are very useful in analysis of a convex optimization problem later. We will also cover several extensions including quasiconvex, log-concave and log-convex functions.

1. Convex functions - basic properties and examples:

- Defn: A function $f: \mathbb{R}^n \to \mathbb{R}$ is convex if $\text{dom} f$ is a convex set [$\text{dom} f$ is the set of all variables of f].
and $\forall x, y \in \text{dom} f$: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$.

Note that $\theta f(x) + (1-\theta) f(y)$ is the line segment between $(x, f(x))$ and $(y, f(y))$ (in $\mathbb{R}^{n+1}$). Geometrically this means this line segment lies above the function f.



+) f is strictly convex if strict inequality holds $\forall x \neq y$.

+) f is concave when $-f$ is convex.

Example: An affine function $f(x) = a^T x + b$ always satisfies the definition with equality $\to$ an affine function is both convex and concave.

+) Properties: i) A function is convex iff it is convex when restricted to any line that intersects its domain.

$$f \text{ convex} \Leftrightarrow \left.\begin{array}{l}\forall x \in \text{dom} f \\ \forall v \in \mathbb{R}^n\end{array}\right\} g(t) = f(x+tv) \text{ is convex}$$

where $\text{dom}\, g = \{t \mid x + tv \in \text{dom} f\}$.

This property is very useful by allowing us to check if a function is convex by restricting it to a line.

ii) A convex function is continuous (but not necessarily differentiable) on the interior of its domain. It can have discontinuity only on its boundary.

+) Examples: i) Functions on $\mathbb{R}$: $f: \mathbb{R} \to \mathbb{R}$

- exponential $f(x) = e^{ax}$, $a \in \mathbb{R}$, $x \in \mathbb{R}$ convex
- powers: $f(x) = x^\alpha$, $x \in \mathbb{R}_{++}$ convex if $\alpha \geq 1$ or $\alpha \leq 0$
  concave if $0 \leq \alpha \leq 1$.
- powers of abosolute value:
  $$f(x) = |x|^p, \; p \geq 1, \; x \in \mathbb{R} \quad \text{convex}$$
- logarithm $f(x) = \lg x$, $x \in \mathbb{R}_{++}$ concave.
- negative entropy
  $$f(x) = x \lg x, \; x \in \mathbb{R}_{++} \quad \text{convex.}$$

ii) Functions on $\mathbb{R}^n$ or $\mathbb{R}^{m\times n}$:

- all norms are convex: $f(x) = \|x\|_p = \left(\sum |x_i|^p\right)^{1/p}$, $p \geq 1$
- affine function on $\mathbb{R}^{m\times n}$: $f(X) = tr(A^T X) + b = \sum_{ij} A_{ij} X_{ij} + b$.
- spectral norm: $f(X) = \|X\|_2 = \sigma_{max}(X) = \left(\lambda_{max}(X^T X)\right)^{1/2}$.

+) <u>Extended-value extension:</u>

Define $\tilde{f}$ as

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

This extension simplifies notation as we do not need to explicitly describe the domain everytime.

$\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $\text{dom}\, \tilde{f} = \mathbb{R}^n$.

We can recover the domain of f as: $\text{dom} f = \{x \mid \tilde{f}(x) < \infty\}$

The extension $\tilde{f}$ also satisfies the basic inequality as

$$\tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y) \quad \forall x, y \in \mathbb{R}$$

Subsequently we drop the tilde (when there is no confusion for using the same symbol). That is, we implicitly extend all convex functions to ∞ outside their domains.

+) <u>First and second order conditions:</u>

- Suppose that f is differentiable, that is, its gradient ∇ exists at each point $x \in \text{dom} f$ (open set).

Then f is convex <u>iff</u> $\text{dom} f$ is convex and

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

- Gradient of a real function of a vector: $f : \mathbb{R}^n \to \mathbb{R}$.

∇f is a vector in $\mathbb{R}^n$ with elements as

$$\nabla f(x)_i = \frac{\partial f(x)}{\partial x_i}, \quad i = 1, \dots, n.$$

<u>Ex</u>: $f(x) = \frac{1}{2} x^T P x + q^T x + r \to \nabla f(x) = P^T x + q$. $(P \in S^n)$

- Chain rule for gradient:

let $h(z) = g(f(z))$ where $f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R} \to \mathbb{R}$

then

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

Eg: $h(x) = \log(a^T x + b) \to \nabla h(x) = \frac{1}{a^T x + b} \cdot a$.

- Specifically, composition with affine function can be written as follows.

Let $f: \mathbb{R}^n \to \mathbb{R}$ be differentiable

define

$$g(x) = f(Ax + b), \quad g: \mathbb{R}^p \to \mathbb{R}$$

$$A \in \mathbb{R}^{p \times n}, \quad b \in \mathbb{R}^p$$

then

$$\nabla g(x) = A^T \nabla f(Ax + b).$$

- For example, when restricting $f(x)$ to a line in its domain, we define

$$g(t) = f(x + tv), \quad x, v \in \mathbb{R}^n. \quad (g: \mathbb{R} \to \mathbb{R})$$

$$\to g'(t) = \nabla g(t) = v^T \nabla f(x + tv)$$

- More generally, we can define the gradient and derivative of a vector function of a vector.

$$f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

The derivative of f at x is a matrix (the Jacobian)

$$Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j} \qquad \begin{array}{l} i = 1 \dots m \\ j = 1 \dots n. \end{array}$$

$Df(x) \in \mathbb{R}^{m \times n}$. The gradient is $\nabla f(x) = Df(x)^T$.

Eg: $y = Ax + b \longrightarrow Dy = A$, $\nabla y = A^T$.
$A \in \mathbb{R}^{m \times n}$

Chain rule: $h(x) = g(f(x))$, $f: \mathbb{R}^n \to \mathbb{R}^m$
$g: \mathbb{R}^m \to \mathbb{R}^p$

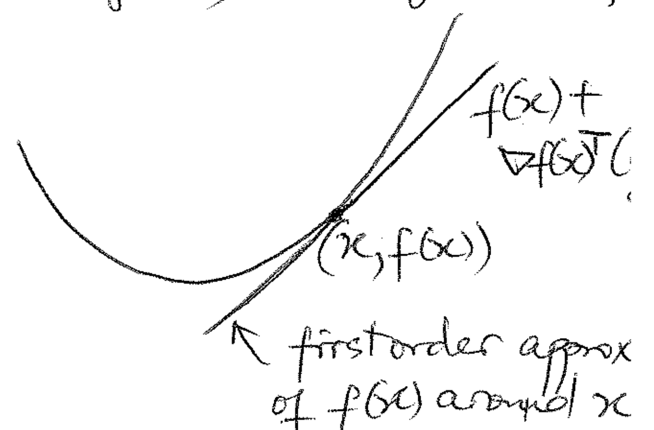
then $Dh(x) = Dg(f) \cdot Df(x)$.

or $\nabla h(x) = \nabla f(x) \cdot \nabla g(f)$.

- Now back to the first order condition on convex fun

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

The line
$\nabla f(x)^T (y-x) + f(x)$
is the first-order Taylor series approximation of $f(x)$ around point x.



A function is convex <u>iff</u> its first-order Taylor approx is a global underestimator of the function ($\leq f(y)$) $\forall$

This condition shows that if $\nabla f(x^*) = 0$ then x^* i global minimizer of the function $f(x)$.

- Strict convexity: first order condition holds with inequality.

- For the proof of the first order condition, see the b The ideas involve:
  - proving the condition for the special case of function of a scalar, $f: \mathbb{R} \to \mathbb{R}$.
  - then generalizing to $f: \mathbb{R}^n \to \mathbb{R}$ by restricting to the line passing through x and y on its d therefore going back to a function of a sc

- <u>Second order condition:</u>

Suppose that f is twice differentiable. Then f is convex <u>iff</u> $\mathrm{dom} f$ is convex and

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

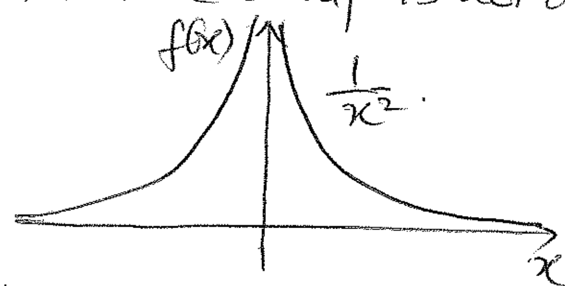
For a function on $\mathbb{R}$, this reduces to $f''(x) \geq 0$.

+) Note that the requirement that $\mathrm{dom} f$ is convex cannot be dropped from the first- and second-order conditions.

<u>Eg</u>: $f(x) = \frac{1}{x^2}$, $\mathrm{dom} f = \{x \in \mathbb{R}, x \neq 0\}$

$f''(x) = \frac{6}{x^4} > 0 \ \forall x \in \mathrm{dom} f$

but $f(x)$ is not convex since $\mathrm{dom} f$ is not convex.



- Second derivative and the Hessian matrix:

For a real value function $f: \mathbb{R}^n \to \mathbb{R}$, the second derivative or Hessian matrix is given by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \qquad i, j = 1 \ldots n$$

$\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ is a square matrix and symmetric.

. $\nabla^2 f(x)$ can also be interpreted as the derivative of the first derivative.

For $f: \mathbb{R}^n \to \mathbb{R}$, the gradient mapping $\nabla f: \mathbb{R}^n \to \mathbb{R}^n$, with $\mathrm{dom}\, \nabla f = \mathrm{dom} f$. Then

$$\nabla^2 f(x) = D \nabla f(x).$$

- Chain rule for second derivatives:
  - Composition with scalar function
    $h(x) = g(f(x))$ where $f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R} \to \mathbb{R}$
    then $\nabla^2 h(x) = g'(f) \nabla^2 f(x) + g''(f) \nabla f(x) \nabla f(x)^T$.
  - Composition with affine function
    Let $f: \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$
    and $g(x) = f(Ax+b)$, $g: \mathbb{R}^m \to \mathbb{R}$
    then $\nabla^2 g(x) = A^T \nabla^2 f(Ax+b) A$

- Example: $f(x) = \frac{1}{2} x^T P x + q^T x + r$, $P \in S^n$, $q \in \mathbb{R}^n$, $r \in$
  $\nabla f(x) = Px + q$
  $\nabla^2 f(x) = P$
  Thus $f(x)$ is convex iff $P \succeq 0$ (eg. see least_square).

+) <u>Examples:</u>
  i) Norms $\|x\|$ are convex (proof based on definition)
  ii) Max function: $f(x) = \max_i x_i$ is convex (l_∞ norm
  iii) Log_sum_exp: $\log(e^{x_1} + e^{x_2} + \ldots + e^{x_n})$ is convex
  based on the Hessian.
  iv) Geometric mean: $f(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ is concave, $x \in \mathbb{R}^n_{++}$
  v) Log determinant: $f(X) = \log\det(X)$ is concave, $X \in S^n_+$

We can show the concavity of $\log\det(X)$ as follows:

Restrict f to a line in its domain $X = Z + tV$ where $Z, V \in S^n$, $Z \succ 0$.

Consider $g(t) = f(Z + tV)$, $Z + tV \succ 0$, $Z \succ 0$.

$$\rightarrow g(t) = \log\det(Z + tV)$$
$$= \log\det\left(Z^{1/2}(I + tZ^{-1/2} V Z^{-1/2}) Z^{1/2}\right)$$
$$= \sum_{i=1}^{n} \log(1 + t\lambda_i) + \log\det Z$$

where $\lambda_i = \lambda_i(Z^{-1/2} V Z^{-1/2})$ are the eigenvalues.

Thus $g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1 + t\lambda_i}$, $g''(t) = -\sum \frac{\lambda_i^2}{(1 + t\lambda_i)^2} < 0$

$\rightarrow g(t)$ is concave $\rightarrow f(X)$ is concave.

The gradient and Hessian:

$$\nabla f(X) = X^{-1} \quad \text{(see Appendix A for derivation)}$$

The Hessian is more complicated, will cover later.

- Epigraph and sublevel sets:

+) α- sublevel set of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as:

$$C_\alpha = \{x \in \operatorname{dom} f \mid f(x) \leq \alpha\}.$$

Sublevel sets of a convex function are convex. This is straightforward to show:

$$x, y \in C_\alpha \rightarrow f(x) \leq \alpha,\ f(y) \leq \alpha$$
$$\rightarrow f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y) \leq \alpha$$
$$\rightarrow \theta x + (1-\theta) y \in C_\alpha.$$

We can show the concavity of $\log\det(X)$ as follows:

Restrict f to a line in its domain $X = Z + tV$ where $Z, V \in S^n$, $Z \succ 0$.

Consider $g(t) = f(Z+tV)$, $Z+tV \succ 0$, $Z \succ 0$.

$$\rightarrow g(t) = \log\det(Z+tV)$$
$$= \log\det\left(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}\right)$$
$$= \sum_{i=1}^{n} \log(1+t\lambda_i) + \log\det Z$$

where $\lambda_i = \lambda_i(Z^{-1/2}VZ^{-1/2})$ are the eigenvalues.

Thus $g'(t) = \sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i}$, $g''(t) = -\sum \frac{\lambda_i^2}{(1+t\lambda_i)^2} < 0$

$\rightarrow g(t)$ is concave $\rightarrow f(X)$ is concave.

The gradient and Hessian:

$$\nabla f(X) = X^{-1} \quad \text{(see Appendix A for derivation)}$$

The Hessian is more complicated, will cover later.

Lecture 5:

- Epigraph and sublevel sets:

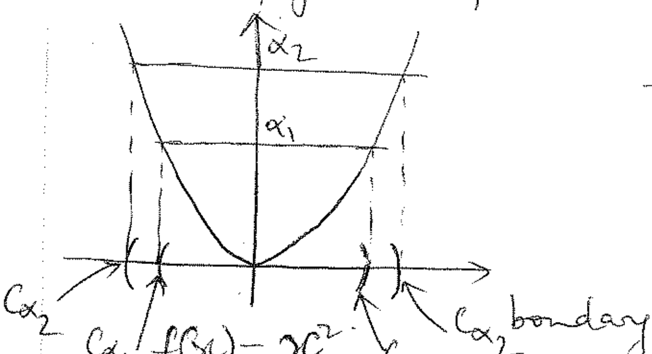
+) α-sublevel set of $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as:

$$C_\alpha = \{x \in \operatorname{dom} f \mid f(x) \le \alpha\}.$$

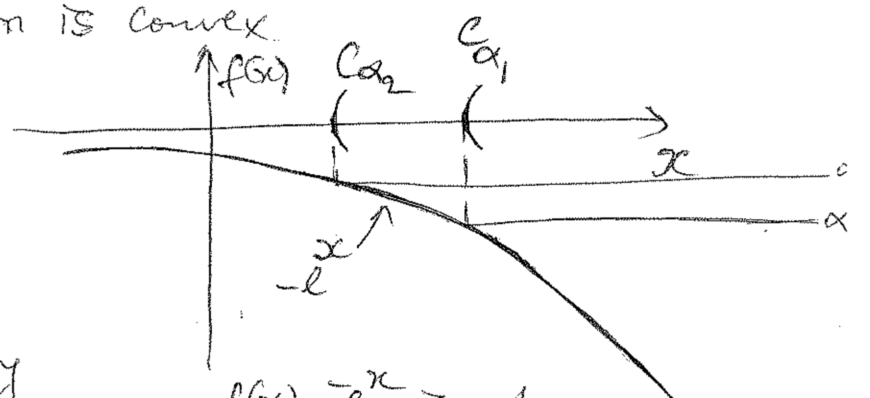
Sublevel sets of a convex function are convex. This is straightforward to show:

$$x, y \in C_\alpha \rightarrow f(x) \le \alpha,\ f(y) \le \alpha$$
$$\rightarrow f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) \le \alpha$$
$$\rightarrow \theta x + (1-\theta) y \in C_\alpha.$$

The converse is not true: all sublevel sets are convex does not imply the function is convex.



$f(x) = x^2$
convex

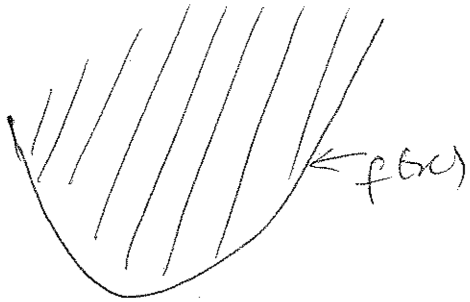


$f(x) = -e^{x}$ is not convex even though all level sets are convex

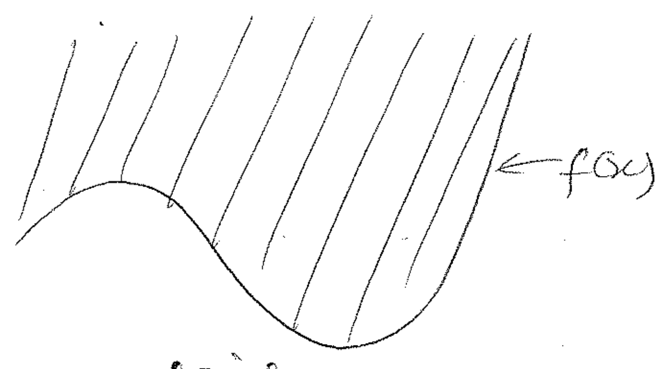
+) The epigraph of a function $f: \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\text{epi } f = \{(x,t) \mid x \in \text{dom} f,\ f(x) \le t\}.$$

Note that $\text{epi} f \subseteq \mathbb{R}^{n+1}$



epi f



epi f

A function $f(x)$ is convex iff its epigraph is a convex set.

+ Using epigraph, many results on convex functions can be obtained from results on convex sets.

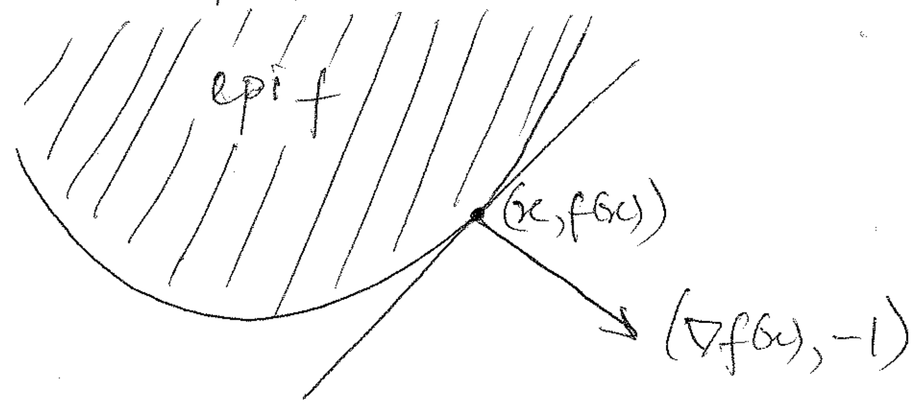
For example, the first derivative condition can be interpreted as $\forall (y,t) \in \text{epi} f \longrightarrow t \ge f(y) \ge f(x) + \nabla f(x)^T (y-x)$

or expressed as $(y,t) \in \text{epi} f \to \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \le 0$

This inequality is that of a halfspace:

let $z = \begin{bmatrix} y \\ t \end{bmatrix}$, $a = \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}$, $b = \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \begin{bmatrix} x \\ f(x) \end{bmatrix}$

then the hyperplane $a^T z - b$ supports epi f at the boundary point $(x, f(x))$.



- Jensen's inequality:

The basic convexity inequality is sometimes called Jensen's inequality.

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

It can be extended to convex combination of more than two points, integrals and expectations.

If $f(x)$ is convex then the following inequalities hold:

$$f\left(\sum_{i=1}^{k} \theta_i x_i\right) \le \sum_{i=1}^{k} \theta_i f(x_i), \quad x_i \in \mathbb{R}^n.$$

$$f\left(\int_S p(x) x \, dx\right) \le \int_S p(x) f(x) \, dx, \quad S \subseteq \mathbf{dom}\, f,\ S \text{ convex},\ p(x) \ge 0,\ \int_S p(x)\,dx = 1$$

$$f(Ex) \le E f(x)$$

This inequality is powerful and can be used to prove many other inequalities (see the textbook for example on getting Holder's inequality).

2. Operations that preserve convexity

How do you check that a function is convex?

- by applying definition
- check $\nabla^2 f(x) \succeq 0$ for twice differentiable functions
- show that $f(x)$ can be obtained from simple convex functions by operations that preserve convexity.

+) Note on checking the second derivative:

You must check the Hessian matrix as the whole. Checking only the second derivatives wrt each scalar variable is not enough!

Eg. $f(x,y) = x^2 + \frac{x}{y}$, $x \in \mathbb{R}_+$, $y \in \mathbb{R}_{++}$.

Then $\frac{\partial^2 f}{\partial x^2} = 2$, $\frac{\partial^2 f}{\partial y^2} = +\frac{2x}{y^3} > 0$ for $y \in \mathbb{R}_{++}$, $x \in \mathbb{R}_+$

But the Hessian is

$$\nabla^2 f = \begin{bmatrix} 2 & -1/y^2 \\ -1/y^2 & \frac{2x}{y^3} \end{bmatrix}$$ is not a positive semidefinite matrix.

[Check: $|\nabla^2 f| = \frac{4x}{y^3} - \frac{1}{y^4}$ is not ≥ 0 $\forall$ $y \in \mathbb{R}_{++}$, $x \in \mathbb{R}_+$].

Thus $f(x,y)$ is not convex.

+) Positive weighted sum:

If $f_1, f_2 \dots f_m$ are convex functions, then

$$f = \omega_1 f_1 + \omega_2 f_2 + \dots \omega_m f_m \quad \text{is convex}$$

provided $w_i \geq 0$.

For strictly convex, need f_i strictly convex and $w_i > 0$.

This property extends to integrals and infinite sums as well.

If $f(x,y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ $\forall y \in \mathcal{A}$ then

$$g(x) = \int_{\mathcal{A}} w(y) f(x,y)\, dy \quad \text{is convex.}$$

This property can be verified easily from the definition.

+) <u>Composition with an affine function:</u>
If $f : \mathbb{R}^n \to \mathbb{R}$ is convex then

$$g(x) = f(Ax+b) \quad \text{is convex}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}$.

+) <u>Point-wise maximum and supremum:</u>
If f_1, f_2 are convex functions then

$$f(x) = \max\{f_1(x), f_2(x)\}, \quad \mathrm{dom}\, f = \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2$$

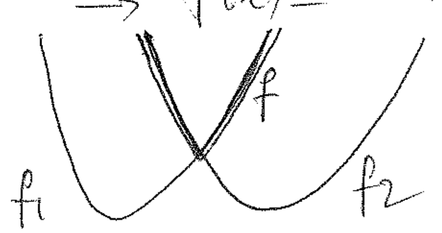
is also convex.

This property is easy to show:

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

This property also generalizes to m functions: If f_i convex

$$\to f(x) = \max\{f_1(x), f_2(x), \dots, f_m(x)\} \text{ is convex}$$



Also extends to point-wise supremum

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y) \text{ is convex if } f(x,y) \text{ convex } \forall y \in \mathcal{A}.$$

In terms of epigraph, point wise supremum is the intersection of epigraphs.

$$\mathrm{epi}\, g = \bigcap_{y \in A} \mathrm{epi} f(\cdot, y)$$

Then intersection of convex sets is a convex set.

<u>Examples:</u> i) Piece-wise linear

$$f(x) = \max\{a_1^T x + b_1, a_2^T x + b_2, \ldots, a_L^T x + b_L\}$$

ii) Sum of r largest components:
let $x \in \mathbb{R}^n$ and order the elements of x as:

$$x_{[1]} \geq x_{[2]} \geq \ldots \geq x_{[n]}$$

then $f(x) = \sum_{i=1}^{r} x_{[i]}$ is convex.

Since $f(x) = \sum_{i=1}^{r} x_{[i]} = \max\{x_{i_1} + x_{i_2} + \ldots + x_{i_r}, i_j \neq i_k\}$

iii) Maximum eigenvalue of a symmetric matrix:
For $X \in S^n$, the maximum eigenvalue is

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

$\lambda_{max}(X)$ is the point-wise supremum of a family of linear functions of X (that is, $y^T X y$), indexed by y.

+) <u>Composition:</u>

let $h : \mathbb{R}^k \to \mathbb{R}$
$g : \mathbb{R}^n \to \mathbb{R}^k$ and $f = h(g(x)) : \mathbb{R}^n \to \mathbb{R}$.
$\text{dom} f = \{x \in \text{dom}\, g \mid g(x) \in \text{dom}\, h\}$.

∘ Scalar composition: $k=1$, thus $h : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$.

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

f is convex if g is convex, $\tilde{h}$ is convex & nondecreasing
or g is concave, $\tilde{h}$ is convex & nonincreasing.

Here $\tilde{h}$ is the extended-value extension. What these conditions mean is that if dom h is not $\mathbb{R}$, then it should be of the form $(-\infty, a)$ or $(-\infty, a]$ for nondecreasing functions. Similarly dom h should extend to $+\infty$ for nonincreasing h.

<u>Example</u>: +) $h(x) = x^2$ with dom $h = \mathbb{R}_+$ is convex but does <u>not</u> satisfy the condition $\tilde{h}$ nondecreasing.

+) $h(x) = \begin{cases} x^2 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$ is convex and <u>does</u> satisfy $\tilde{h}$ nondecreasing.

This is a technical condition (on $\tilde{h}$) to make the composition rule general such that it can be applied to all functions without assuming differentiability of h and g or that dom $g = \mathbb{R}^n$ and dom $h = \mathbb{R}$.

<u>Proof</u>: For the simple case of $n=1$ and both f, g are twice differentiable, then

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

Alternatively, composition rules can be proved directly without making any of the above assumptions on n and differentiability. This proof is simply based on convexity definition (see textbook).

<u>Examples:</u> $f(x) = e^{g(x)}$ is convex if $g(x)$ is convex
$f(x) = \log(g(x))$ is concave if $g(x)$ is concave and positive

∘ Vector composition: $k \geq 1$: $h : \mathbb{R}^k \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^k$.

$$f(x) = h(g(x)) = h(g_1(x), \dots, g_k(x)) \; : \; g_i : \mathbb{R}^n \to \mathbb{R}.$$

Without loss of generality, we can consider $n = 1$ since convexity can be determined by restricting the function to a line in its domain.

For twice differentiable functions h, g and $\mathrm{dom}\, h = \mathbb{R}^k$, $\mathrm{dom}\, g = \mathbb{R}$, then

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

f is convex if h is convex, h is nondecreasing in each argument and g is convex.
or h is convex, h is nonincreasing in each argument and g is concave.

<u>Example</u>: i) $h(z) = \log\left(\sum_{i=1}^{k} e^{z_i}\right)$ is convex and nondecreasing in each argument

$\to h(g(x)) = \log\left(\sum_{i=1}^{k} e^{g_i(x)}\right)$ is convex if g_i convex.

ii) $\sum_{i=1}^{k} \log g_i(x)$ is concave if $g_i(x)$ are concave & positive.

This rule can also hold for the general case $n > 1$, no assumption on the differentiability of h or g, and general domains using extended-value extension $\tilde{h}$.

Lecture 6:

+) **Minimization**: Some forms of minimization also preserves convexity.

If f is convex in (x,y) and C is a convex nonempty set

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

provided that $g(x) > -\infty \ \forall x$

Here $\operatorname{dom} g = \{x \mid (x,y) \in \operatorname{dom} f \text{ for some } y \in C\}$.

(can also prove using Jensen's inequ.)

Proof: Using epigraph, assume that the infimum over $y \in C$ is attained for each x, then

$$\operatorname{epi} g = \{(x,t) \mid (x,y,t) \in \operatorname{epi} f \text{ for some } y \in C\}$$

Since $\operatorname{epi} f$ is convex, $\operatorname{epi} g$ is the projection of $\operatorname{epi} f$ on some of its components and hence is convex.

Example: i) Given $f(x,y) = x^T A x + 2x^T B y + y^T C y$ convex in (x,y), where A, C are symmetric.

This convexity is equivalent to

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0.$$

Now minimize f over y to get

$$g(x) = \inf_y f(x,y) = x^T (A - BC^{\dagger}B^T) x$$

where $C^{\dagger} = C^T (CC^T)^{-1}$ is the pseudoinverse of C. (If C is invertible $\rightarrow C^{\dagger} = C^{-1}$).

By the minimization rule, $g(x)$ is convex, hence

$$A - BC^{\dagger}B^T \succeq 0$$

This expression is called the Schur complement of C.

ii) Distance to a set

$$\operatorname{dist}(x,S) = \inf_{y \in S} \|x - y\| \quad \text{is convex if } S \text{ is convex}$$

7) Perspective of a function:

$f: \mathbb{R}^n \to \mathbb{R}$, then perspective of f is $g: \mathbb{R}^{n+1} \to \mathbb{R}$ s.t.

$$g(x,t) = t f(x/t).$$

where $\operatorname{dom} g = \{(x,t) \mid x/t \in \operatorname{dom} f, t>0\}$.

If f is convex then g is convex. Also preserves concavity

Example: $f(x) = -\log x$ is convex on R_{++}

$$\to g(x,t) = -t \log \frac{x}{t} \text{ is convex on } \mathbb{R}^2_{++}.$$

From this we get the relative entropy between 2 vectors $u, v \in$

$$D(u,v) = \sum_{i=1}^{n} u_i \log \frac{u_i}{v_i} \quad \text{is convex in } (u,v).$$

3. The conjugate function:

- Defn: Let $f: \mathbb{R}^n \to \mathbb{R}$. The conjugate function of f is defined as:

$$f^*(y) = \sup_{x \in \operatorname{dom} f} \left(y^T x - f(x)\right)$$

Since $f^*(y)$ is the supremum of a family of affine function of y, then $f^*(y)$ is convex, even if $f(x)$ is not convex.

This concept will be useful in duality (Chapter 5).

Example: $f(x) = \frac{1}{2} x^T Q x$, $Q \in S^n_{++}$ (strictly convex

Then $f^*(y) = \sup_{x \in \mathbb{R}^n} \left(y^T x - \frac{1}{2} x^T Q x\right)$

The RHS is a quadratic function of x, so it has a maximum at $Qx - y = 0 \to x = Q^{-1} y$.

Then the conjugate function is

$$f^*(y) = \frac{1}{2} y^T Q^{-1} y.$$

What this means is:

$$\frac{1}{2} y^T Q^{-1} y \geq y^T x - \frac{1}{2} x^T Q x \quad \forall\, x, y \in \mathbb{R}^n$$

or simply $\frac{1}{2}(y^T Q^{-1} y + x^T Q x) \geq y^T x$

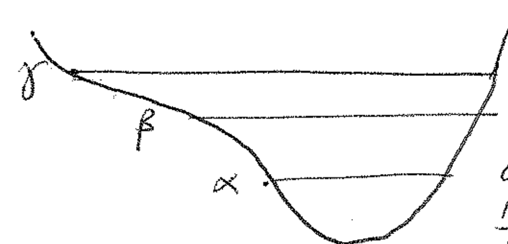
This is an example of Fenchel's inequality.

4. Quasiconvex functions:

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

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

$f(x)$ is quasilinear if it is both quasiconvex and quasiconcave



quasiconvex function.

Example: i) $f : \mathbb{R}^2 \to \mathbb{R}$, $\text{dom} f = \mathbb{R}^2_+$.
$f(x_1, x_2) = x_1 x_2$ is not convex since the Hessian is

$$\nabla^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

But this function is quasiconcave on $\mathbb{R}^2_+$ since all its superlevel sets are convex

$$\{x \in \mathbb{R}^2_+ \mid x_1 x_2 \geq \alpha\}.$$

ii) $\sqrt{|x|}$ is quasiconvex on $\mathbb{R}$

iii) Distance ratio $f(x) = \frac{\|x-a\|_2}{\|x-b\|_2}$, $\text{dom} f = \{x \mid \|x-a\|_2 < \|x-b\|_2\}$

is quasiconvex on $\mathbb{R}$.

Basic properties:

Quasiconvex functions generalizes convex functions and has some similar properties:

+) Modified Jensen's inequality: for f quasiconvex,

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

Example:

For $X \in S^n_+$ (PSD matrices), rank(X) is quasiconcave since

$$\text{rank}(X+Y) \geq \min\{\text{rank}(X), \text{rank}(Y)\}.$$

+) Similar to convexity, quasiconvex function can be verified by restricting to a line in its domain and this restriction is quasiconvex.

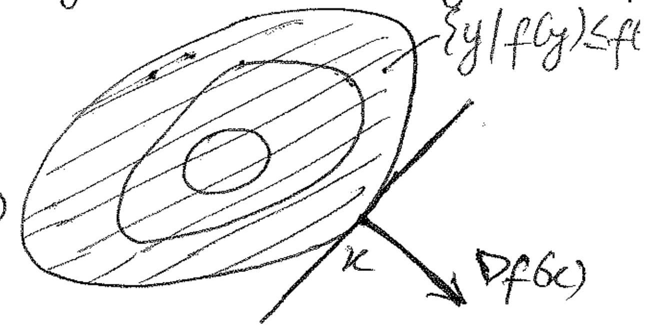
+) Quasiconvex functions on $\mathbb{R}$: For a continuous function $f: \mathbb{R} \to \mathbb{R}$, it is quasiconvex iff at least one of the following conditions holds:

- f is non-decreasing
- f is non-increasing
- $\exists$ a point $c \in$ dom f: $t \leq c$, $f(t)$ is non increasing; $t \geq c$, $f(t)$ is non decreasing.

→ point c is the global minimizer of $f(x)$.

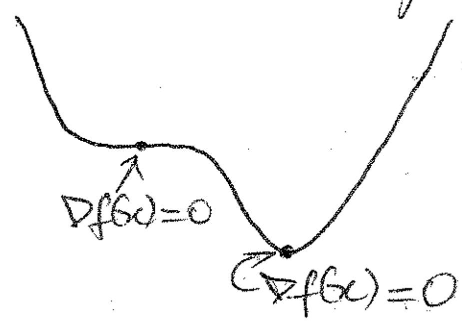
+) First order condition: $f: \mathbb{R}^n \to \mathbb{R}$ is differentiable, then f is quasiconvex iff dom f is convex and

$$f(y) \leq f(x) \rightarrow \nabla f(x)^T (y - x) \leq 0 \quad \forall x, y \in \text{dom} f$$



Note that for convex function $f(x)$, x^* is the global minimizer iff $\nabla f(x^*) = 0$, but for quasiconvex $f(x)$ only one direction holds.

f quasiconvex, then
if x^* is the global minimizer $\rightarrow \nabla f(x^*) = 0$
but $\nabla f(x^*) = 0$ does not guarantee x^* to be the global minimizer.



+) Second order condition:
$f: \mathbb{R}^n \rightarrow \mathbb{R}$ twice differentiable, f is quasiconvex, then
if $y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0 \;\; \forall x \in \mathrm{dom} f$
$y \in \mathbb{R}^n$.
For quasiconvex function on $\mathbb{R}$, this condition reduces to
$$f'(x) = 0 \Rightarrow f''(x) \geq 0.$$
The converse also holds. (If $y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0 \; \forall x \in \mathrm{dom} f$ and $y \in \mathbb{R}^n$, $y \neq 0$ then f is quasiconvex).

5. Log concave and log-convex functions:
- Defn: A positive function $f(x)$ is log-concave if $\log f(x)$ is concave.
$$f(\theta x + (1-\theta) y) \geq f(x)^\theta f(y)^{1-\theta} \quad \forall \; 0 \leq \theta \leq 1.$$

f is log-convex if $\log f$ is convex.

Note: Since e^h is convex if h is convex $\rightarrow$ a log-convex function is also convex. (So there is not much difference for convex functions).

For log-concave: a nonnegative concave function is also log-concave

Examples: + $f(x) = x^a$, $x \in \mathbb{R}_{++}$ is log convex for $a \leq 0$
log concave for $a \geq 0$.

+) many probability densities are log-concave

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2} x^T \Sigma^{-1} x}$$ is log-concave $x \in \mathbb{R}^n$.

+) $\det X$ is log-concave on S^n_{++}.

+) Properties:

∘ For twice differentiable f with convex domain, f is log-concave iff

$$f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T. \quad \forall x \in \operatorname{dom} f.$$

This is because

$$\nabla^2 \log f(x) = \frac{+1}{f(x)} \nabla^2 f(x) - \frac{1}{f(x)^2} \nabla f(x) \nabla f(x)^T \succeq 0$$

∘ Product of log-concave functions is log-concave

∘ Sum of log-concave functions is not always log-concave.

∘ Integration: If $f(x,y)$ is log-concave in x for each $y \in C$ then

$g(x) = \int_C f(x,y)\, dy$ is log-concave ($x \in \mathbb{R}^n, y \in \mathbb{R}^m$).

(this result is not simple to show)

Examples:

+ Marginal distributions of log-concave pdf are log-concave
+ Convolution of 2 log-concave function preserves log concavity
+ Volume of a polyhedron:

$$P_u = \{x \in \mathbb{R}^n \mid Ax \preceq u\}, \quad A \in \mathbb{R}^{m \times n}$$

then $\operatorname{vol} P_u$ is a log-concave function of u.

To see this, note that the indicator function

$$\varphi(x,u) = \begin{cases} 1 & Ax \leq u \\ 0 & \text{otherwise} \end{cases}$$

is log-concave. By the integration property, then

$$\int \varphi(x,u)\,dx = \operatorname{Vol} P_u$$

is log-concave.