LECTURE 1 : 2^{nd} Jan. 2025

EE61012: Convex Optimization for Control and Signal Processing
Instructor: Prof. Ashish R. Hota

- Class Hours: G Slot + S3(2) Slot. Wednesday: 11am - 11:55pm, Thursday: 12pm - 12:55pm, Thursday: 5pm - 5:55pm, Friday: 8am-8:55am
- Venue: NR 413
- Grading Scheme: 50 % Endsem, 30 % Midsem, 20 % Tutorial and Class Tests
- Preferred Mode of Contact: Send email to ahota@ee.iitkgp.ac.in with subject containing [EE61012]. Do not forget to write your name and roll no.
- Any email with a blank subject and without name and roll no. will be ignored.

Content

Theory:

- Formal definition of an optimization problem
- Basic topology of sets and existence of optimal solutions
- Gradient, Hessian, and optimality conditions for unconstrained problems
- Convex sets and properties
- Convex functions and properties
- Convex optimization problems and their classifications
- Separating Hyperplane Theorems, Theorems of the Alternative, LP Duality
- Lagrangian duality and KKT optimality conditions

Algorithms:

- First order gradient based algorithms under smoothness, strong convexity
- Accelerated, stochastic and distributed gradient descent

Applications:

- Regression, support vector machines, ML estimation, hypothesis testing
- Stability analysis and controller synthesis for linear dynamical systems
- Robust optimization

References

Primary References:

- Optimization Models by G.C. Calafiore and L. El Ghaoui, Cambridge University Press, 2014. Link: https://people.eecs.berkeley.edu/~elghaoui/optmodbook.html
- Convex Optimization by Stephen Boyd and L. Vandenberghe, Cambridge University Press. Available online at: https://web.stanford.edu/~boyd/cvxbook/
- Algorithms for Convex Optimization by Nisheeth K. Vishnoi, Cambridge University Press. Available online at: https://convex-optimization.github.io

Advanced References on Theory:

- Lectures on Modern Convex Optimization, Aharon Ben-Tal and Arkadi Nemirovski, SIAM. Available online at: https://epubs.siam.org/doi/book/10.1137/1.9780898718829
- Convex Analysis and Optimization, Bertsekas, Athena Scientific. More information at: http://www.athenasc.com/convexity.html
- Convex Analysis and Minimization Algorithms, Jean-Baptiste Hiriart-Urruty, Claude Lemarechal, Springer. Available online at: https://link.springer.com/book/10.1007/978-3-662-02796-7

Advanced References on Algorithms:

- Optimization for Modern Data Analysis, Benjamin Recht and Stephen J. Wright, Available online at: https://people.eecs.berkeley.edu/~brecht/opt4ml_book/
- Numerical Optimization by Jorge Nocedal, Stephen J. Wright, Springer. Available online at: https://link.springer.com/book/10.1007/978-0-387-40065-5

- Introductory Lectures on Convex Optimization A Basic Course, by Yurii Nesterov. Available online at: https://link.springer.com/book/10.1007/978-1-4419-8853-9

- First-order Methods in Optimization, by Amir Beck, SIAM. For more information: https://epubs.siam.org/doi/10.1137/1.9781611974997.

Advanced References on Applications in Control:

- Linear Matrix Inequalities in System and Control Theory, by Stephen Boyd, Laurent El Ghaoui, E. Feron, and V. Balakrishnan, Society for Industrial and Applied Mathematics (SIAM), 1994. Available online at: https://web.stanford.edu/~boyd/lmibook/

- A Course in Robust Control Theory: A Convex Approach, Springer. Available online at: https://link.springer.com/book/10.1007/978-1-4757-3290-0

- Predictive Control for Linear and Hybrid Systems, Cambridge University Press. More information at: http://www.mpc.berkeley.edu/mpc-course-material

Advanced References on Applications in Signal Processing and Machine Learning:

- Convex Optimization in Signal Processing and Communications, Cambridge University Press. More information at: https://www.cambridge.org/in/academic/subjects/engineering/communications-and-signal-processing/convex-optimization-signal-processing-and-communications?format=HB&isbn=9780521762229

- Optimization for Machine Learning, by Suvrit Sra, Stephen J. Wright, Sebastian Nowozin, MIT Press. More information at: https://mitpress.mit.edu/9780262537766/optimization-for-machine-learning/

- Recent Special Issue of Proceedings of the IEEE: https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=9241485&punumber=5

Computing Resources

MATLAB Toolbox

- YALMIP: https://yalmip.github.io/
- CVX: http://cvxr.com/cvx/

Python Toolbox

- CVXOPT: https://cvxopt.org/index.html
- CVXPY: https://www.cvxpy.org/
- PYOMO: http://www.pyomo.org/

Solvers

- MOSEK: https://www.mosek.com/
- Gurobi: https://www.gurobi.com/
- IPOPT: https://github.com/coin-or/Ipopt
- COIN-OR: https://github.com/coin-or/
- For optimal control, Casadi: https://web.casadi.org/

Preliminaries

See https://www.stat.cmu.edu/~ryantibs/convexopt/prerequisite_topics.pdf for refresher.

Please also see the Appendices of Boyd's Book and Chapter 2 of ACO Book.

$x \in \mathbb{R}^2, \quad f(x) = x^T Q x, \qquad Q = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$

$X = \left\{ x \in \mathbb{R}^2 \,\middle|\, \begin{array}{l} x_1 \geq 0 \\ x_2 \geq 1 \\ x_1 + x_2 = 5 \end{array} \right\}$

Optimization in Abstract Form

$$\min_{x} \; ax^2 + b \quad \text{s.t. } 1 \leq x \leq 2$$

An optimization problem can be stated as

$$\min_{x \in X} f(x), \tag{1}$$

where

- x decision variable, often a vector in $\mathbb{R}^n$
- X set of feasible solutions, often a subset of $\mathbb{R}^n$
  - often specified in terms of equality and inequality constraints $X := \{x \in \mathbb{R}^n | g_i(x) \leq 0, h_j(x) = 0, i \in \{1, 2, \ldots, m\}, j \in \{1, 2, \ldots, p\}\}$.
- $f : \mathbb{R}^n \to \mathbb{R}$ cost function

$x \in \mathbb{R}$

$X = \{x \in \mathbb{R} \mid 1 \leq x \leq 2\} = [1, 2]$

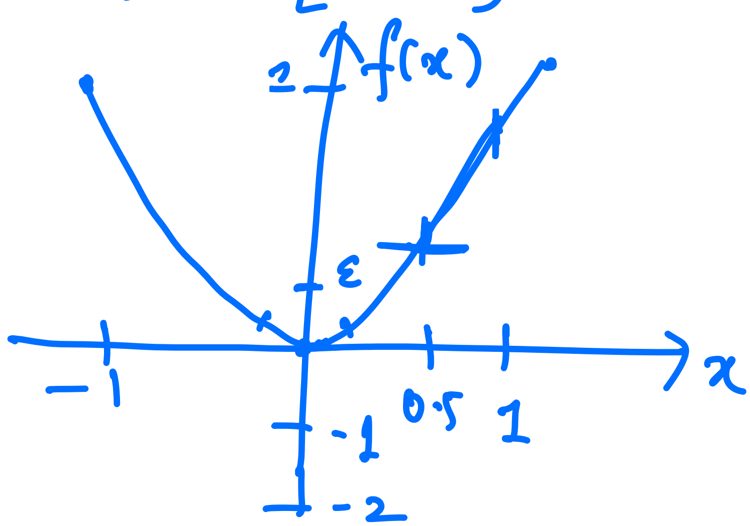
Goal:

- Find $x^* \in X$ that minimizes the cost function, i.e., $f(x^*) \leq f(x)$ for every $x \in X$.
- Optimal value: $f^* := \inf_{x \in X} f(x)$
- Optimal solution: $x^* \in X$ if $f(x^*) = f^*$.

What is $\inf_{x \in X} f(x)$?

$f(x) = x^2$

$X = [-1, 1)$



Infimum vs. Minimum

$f^* := \inf_{x \in X} f(x)$ if f^* is the **greatest lower bound** on the value of the function $f(x)$ over $x \in X$.

- For any $\epsilon > 0$, there exists some $\bar{x} \in X$ such that $f^* \leq f(\bar{x}) < f^* + \epsilon$.

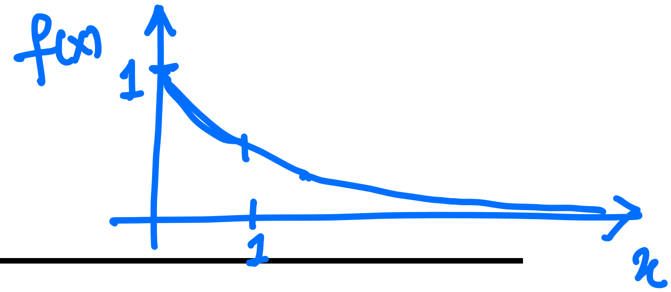
There are two possibilities:

- There exists $x^* \in X$ for which $f(x^*) = f^*$. Then, we say that x^* is the optimal solution and $f^* := \min_{x \in X} f(x)$ is the optimal value.
- $f(x) \neq f^*$ for any $x \in X$. We then say that the infimum is not attained for this problem.
- If $|X|$ is finite, then infimum is always attained.
- The set of optimal solutions is denoted by argmin, and we say

$$x^* \in \text{argmin}_{x \in X} f(x) = \{y \in X | f(y) = f^*\}.$$

- Note that $[\text{argmin}_{x \in X} f(x)] \subseteq X$.

$$\inf_{x \in [0,\infty)} e^{-x} = 0 = f^*$$



Examples

- Let $f(x) = e^{-x}$ and $X = [0, \infty)$. Find f^* and x^*.

$$\left[\operatorname{argmin}_{x \in X} f(x)\right] = \emptyset$$

does not exist

- What if $X = [0, 1]$?

$f^* = e^{-1}$, $f(1) = f^*$

$\Rightarrow x^* = 1$

$$\left[\operatorname{argmin}_{x \in X} f(x)\right] = \{1\}$$

- What if $X = [0, 1)$?

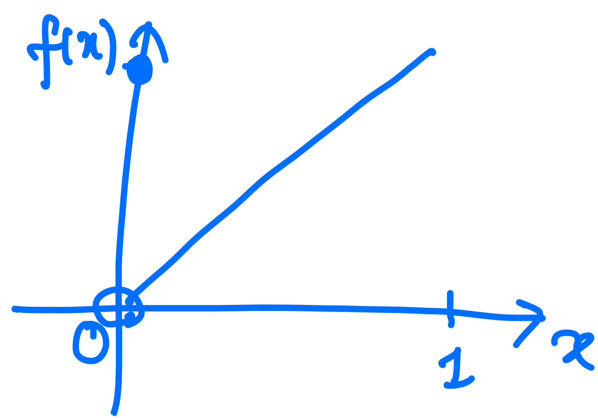
$f^* = e^{-1}$, optimal solution does not exist.

Moral of the story: Properties of feasibility set X is critical in existence of optimal solution.

Now suppose $X = [0, 1]$ and $f(x) = x$ for $x > 0$ and $f(x) = 1$ for $x = 0$.

$$f^* = \inf_{x \in X} f(x) = 0$$

optimal solution does not exist.



Moral of the story:

Ex: $f(x) = x^2$, $X = \mathbb{R} = (-\infty, \infty)$

$x^* = 0$ is a (global) optimum, despite the set X being unbounded.

Infeasible optimization problem

- The problem is infeasible when X is an empty set.
- In this case, $f^* := +\infty$.
- Example: $X = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} x_1 \geq 0 \\ x_2 \geq 1, \\ x_1 + x_2 \leq -1 \end{array} \right\}$

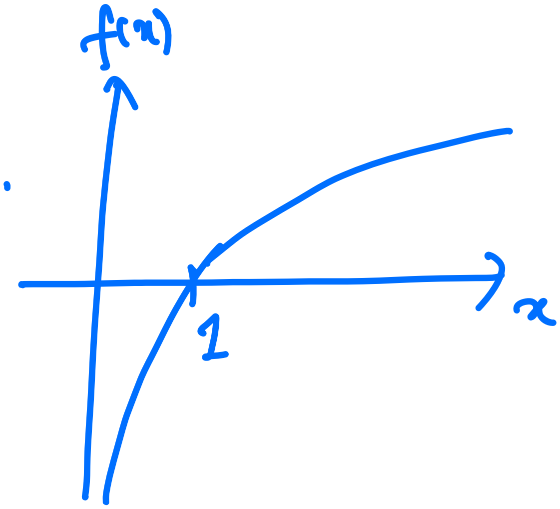
Unbounded optimization problem

- The problem is unbounded when $f^* = -\infty$ over the feasibility set X.
- Example:

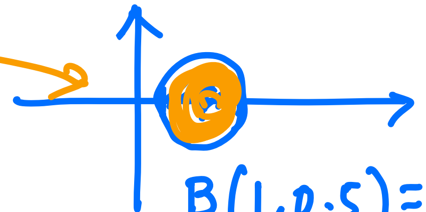
$f(x) = \log x$

$X = [1,5)$: not unbounded.

$X = [0,5)$: problem is unbounded.



$B\left(\begin{bmatrix}1\\0\end{bmatrix}, 0.5\right) =$



$B(1,0.5) = [0.5, 1.5]$

Basic Topology of Sets

Let $B(x_0, r) := \{x \in \mathbb{R}^n | ||x - x_0||_2 \leq r\}$ denote the ball around point $x_0 \in \mathbb{R}^n$ with radius $r > 0$.

- Interior of the set X $\subseteq \mathbb{R}^n$, denoted $\texttt{int}(X)$ $= \{x \in X \mid \exists\, r > 0$ for which $B(x, r) \subseteq X\}$

Example: $(0,1)$

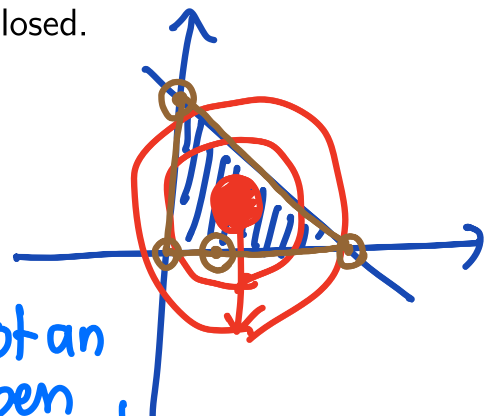
- Set X is called an **open** set if $X = \texttt{int}(X)$.

$(-\infty, 0] \cup [1, \infty)$

- Set X is called **closed** if and only if its complement is open.

- Intersection of arbitrary number of closed sets is closed.

Examples of Open and Closed Sets':



$X = \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$ $\rightarrow$ not an open set

$\text{int}\, X = \{x \in X \mid x_1 > 0, x_2 > 0, x_1 + x_2 < 1\}$

Example: ϕ : both open and closed

$\Rightarrow \mathbb{R}^n$: both closed and open.

Example: $[0,1]$: not an open set since $\text{int}[0,1] = (0,1) \neq [0,1]$.

$[0,1]^c = \underbrace{[-\infty, 0] \cup (1, \infty)}$: not an open set since 0 is not an interior point.

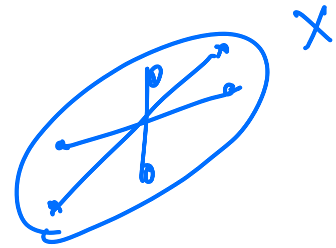
$\Rightarrow [0,1]$ is not a closed set.

LECTURE 2: 2^{nd} Jan 2025

Bounded and Compact Set

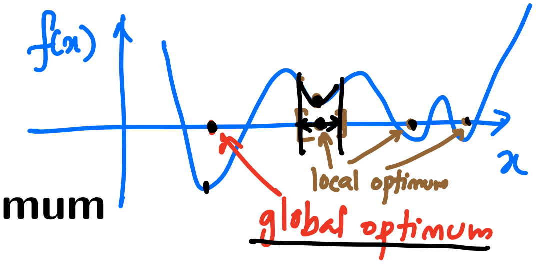
- A set X is bounded if there exists $B \in (0, \infty)$ such that for any $x_1, x_2 \in X$, $||x_1 - x_2||_2 \leq B$.

Ex: $X = [-50, 7000]$



- A set X is compact if it is closed and bounded.

$$\min_{x \in X} f(x)$$



Global and Local Optimum

Definition 1 (Global Optimum). *A feasible solution $x^* \in X$ is a global optimum if $f(x^*) \leq f(x)$ for all $x \in X$. In this case, $f^* = f(x^*)$. The set of global optima is denoted by*

$$\text{argmin}_{x \in X} f(x) := \{z \in X | f(z) = f^*\}.$$

Definition 2 (Local Optimum). *A feasible solution $x^* \in X$ is a local optimum if $f(x^*) \leq f(x)$ for all $x \in B(x^*, r)$ for some $r > 0$.*

Existence of Optimal Solution:

Theorem 1: Weierstrass Theorem

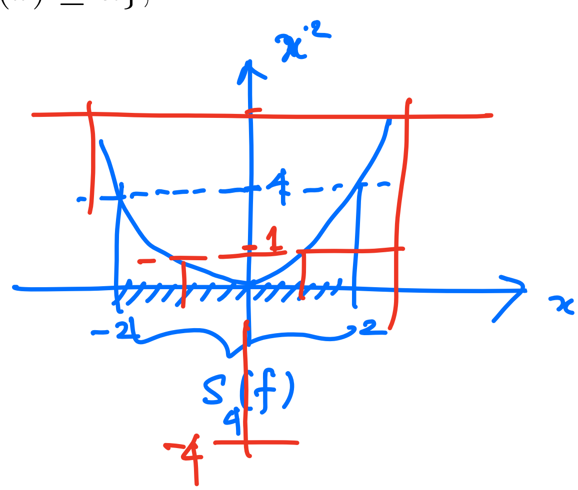
If the cost function f is continuous and the feasible region X is compact (closed and bounded), then (at least one global) optimal solution x^* exists.

Example: $f(x) = x^2$, $X = (-\infty, \infty) = \mathbb{R}$ $\qquad S_4(f) = \{x \in \mathbb{R} \mid x^2 \leq 4\} = [-2, 2]$

When X is not bounded, then the above theorem still holds when an α-sublevel set of f, defined as

$$S_\alpha(f) := \{x \in X | f(x) \leq \alpha\},$$

is non-empty and bounded for some $\alpha \in \mathbb{R}$.



$\min_{x \in X} f(x)$ is equivalent to $\min_{x \in S_\alpha(f)} f(x)$.

Notes

Given an optimization problem, first determine

- the decision variable x and the space in which it resides
- feasibility set X
- cost function $f : X \to \mathbb{R}$

Before attempting to solve the problem, check whether

- f is continuous
- X is non empty, or the problem is unbounded
- X is closed, and bounded (or any sub-level set of X is bounded)

How to verify whether some x^* is indeed an optimal solution?

We are going to derive necessary & sufficient conditions for optimality when the cost function f is differentiable.

For a function $f : \mathbb{R}^n \to \mathbb{R}$, its derivative at point $x_0 \in \mathbb{R}^n$ is denoted by $Df(x_0) \in \mathbb{R}^{1\times n}$, satisfies

$$f(x_0+\Delta x) \simeq f(x_0) + Df(x_0)\Delta x.$$

Gradient of function f at x_0 is denoted by

$$\nabla f(x_0) = Df(x_0)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \frac{\partial f}{\partial x_2}(x_0) \\ \vdots \end{bmatrix}_{n\times 1}.$$

Gradient ($\nabla f(x)$)

For a function $f : \mathbb{R}^n \to \mathbb{R}$, its gradient is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

Compute gradient of

- $f(x) = x^\top a$ $= \sum_{i=1}^{n} a_i x_i$

  $= a^T x$
- $f(x) = x^\top A x$ $= x^T \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \vdots \\ \sum \end{bmatrix} = \sum_{i=1}^{n} \left[a_{ii} x_i^2 + \sum_{j \neq i} a_{ij} x_i x_j \right]$
- $f(x) = ||Ax - b||_2^2$

$$\frac{\partial f}{\partial x_K} = 2a_{KK} x_K + \sum_{i \neq K} a_{iK} x_i + \sum_{j \neq K} a_{Kj} x_j = \sum_{i=1}^{n} a_{iK} x_i + \underbrace{\sum_{j=1}^{n} a_{Kj} x_j}$$

$$= (A^T x)_K + (Ax)_K$$

$$\boxed{\nabla f(x) = A^T x + Ax},$$

$$f(x) = ||Ax - b||_2^2$$

$$= (Ax - b)^T (Ax - b)$$

$$= x^T A^T A x - 2b^T A x + b^T b.$$

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

$$= 2A^T A x - 2A^T b.$$

$$D\nabla f(x) = \begin{bmatrix} D\frac{\partial f}{\partial x_1} \\ D\frac{\partial f}{\partial x_2} \\ \vdots \\ D\frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} [\quad] \\ \vdots \\ [\quad] \end{bmatrix}$$

$$D\left(\frac{\partial f}{\partial x_1}\right) = \left[\frac{\partial^2 f}{\partial x_1^2} \quad \frac{\partial f}{\partial x_1 \partial x_2} \quad \frac{\partial f}{\partial x_1 \partial x_3} \quad \cdots \quad \frac{\partial f}{\partial x_1 \partial x_n}\right] \in \mathbb{R}^{1\times n}$$

Hessian ($H(x)$)

For a function $f : \mathbb{R}^n \to \mathbb{R}$, its Hessian is defined as: $H(x) = D\,\nabla f(x)$

$H(x) \in \mathbb{R}^{n\times n} \qquad [H(x)]_{ij} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}$

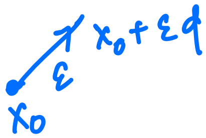
Compute Hessian of

- $f(x) = x^\top a$, $\nabla f(x) = a$, $H(x) \in \mathbb{R}^{n\times n}$, $(H(x))_{ij} = 0 \;\; \forall_{1\le i,j\le n}$
  $\to Df(x) = a^T$
- $f(x) = x^\top A x$, $\nabla f(x) = A^T x + Ax$, $H(x) = A + A^T$
- $f(x) = ||Ax - b||_2^2$, $\nabla f(x) = 2A^TAx - 2A^Tb$, $H(x) = 2A^TA$

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

$\underbrace{Dh(x)}_{\in \mathbb{R}^{p\times n}} = \underbrace{Dg(y)}_{\in \mathbb{R}^{p\times m}} \cdot \underbrace{Df(x)}_{\in \mathbb{R}^{m\times n}}$, where $y = f(x)$

$$\underbrace{f(x_0 + \Delta x)}_{\in \mathbb{R}^m} \cong \underbrace{f(x_0)}_{\in \mathbb{R}^m} + \underbrace{Df(x_0)}_{\mathbb{R}^{m\times n}} \cdot \underbrace{\Delta x}_{\in \mathbb{R}^n}$$



Directional Derivative and Descent Direction

Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Let $d \in \mathbb{R}^n$ be the direction of interest.

Definition: The directional derivative of f at point $x_0 \in \mathbb{R}^n$ along direction $d \in \mathbb{R}^n$ is defined as

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon d) - f(x_0)}{\varepsilon} = \underline{\nabla f(x_0)^T d}$$

Define $\phi(t) := f(x_0 + td)$.

Compute $\phi'(0)$:

$$\underbrace{\phi'(t)}_{\in \mathbb{R}} = \underbrace{Df(x_0 + td)}_{\in \mathbb{R}^{1\times n}} \cdot \underbrace{\frac{d}{dt}(x_0 + td)}_{\in \mathbb{R}^{n \times 1}}$$

$$= Df(x_0 + td) \cdot d$$

$$\phi'(0) = Df(x_0) \cdot d = \nabla f(x_0)^T d$$

If the directional derivative is negative along direction d, then d is called a descent direction of the function at point x_0.

$d = -\nabla f(x_0)$ is always a direction of descent at x_0.

LECTURE 3: 3rd Jan 2025

Necessary Condition of Optimality for Unconstrained Problems

$\Rightarrow X = \mathbb{R}^n$

Theorem 2

If x^* is a local optimum for the problem $\min_{x \in \mathbb{R}^n} f(x)$, then $\nabla f(x^*) = 0$.

Proof by contradiction: Suppose $\nabla f(x^*) \neq 0$. we need to show that x^* is not a local optimum. In other words, there exist points arbitrarily close to x^* at which $f(x) < f(x^*)$.

$$f(x^* + \varepsilon d) \simeq f(x^*) + \varepsilon \nabla f(x^*)^T d + (\text{higher order terms})$$

let $d = -\nabla f(x^*)$

$$= f(x^*) \underbrace{- \varepsilon \|\nabla f(x^*)\|_2^2}_{-ve} + (\text{hot})$$

$< f(x^*)$ when ε is sufficiently small.

Thus, for x^* to be a local optimum, we need to have $\nabla f(x^*) = 0$.

Sufficient Condition of Optimality for Unconstrained Problems

$v^T H(x^*) v > 0 \quad \forall v \neq 0.$

Let f be twice continuously differentiable over $\mathbb{R}^n$.

Theorem 3

If for $x^* \in \mathbb{R}^n$, we have $\nabla f(x^*) = 0$ and the Hessian of the cost function f at x^* is a positive definite matrix, then x^* is a local optimum for the problem $\min_{x \in \mathbb{R}^n} f(x)$.

Using the Taylor series expansion, we obtain

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T H(x^*)(x - x^*) + (hot)$$

$$= f(x^*) + \underbrace{\frac{1}{2}(x - x^*) H(x^*) (x - x^*)}_{> 0} + (hot)$$

thus, $\exists$ a neighborhood around x^* s.t.

$$f(x) > f(x^*) \quad \forall x \in B(x^*, \varepsilon).$$

$\Rightarrow x^*$ is a local optimum.

Least Sqaures Problem

Consider the following optimization problem:

$$\min_{x \in \mathbb{R}^n} ||Ax - b||_2^2.$$

$$f(x) = ||Ax - b||_2^2$$
$$= (Ax-b)^T (Ax-b)$$
$$= x^T A^T A x - 2b^T A x + b^T b.$$

$$H(x) = 2A^T A$$

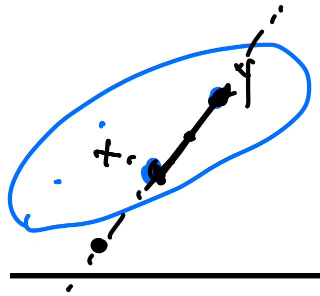
Let x^* be a local optimum.

then we must have $\nabla f(x^*) = 0$

$$\Rightarrow 2A^T A x^* - 2A^T b = 0$$
$$\Rightarrow \boxed{A^T A x^* = A^T b}$$

If $A^T A$ is a positive definite matrix, then any x^* that satisfies $A^T A x^* = A^T b$ is a local optimum.

$$\Rightarrow x^* = (A^T A)^{-1} A^T b \text{ is the local optimum.}$$



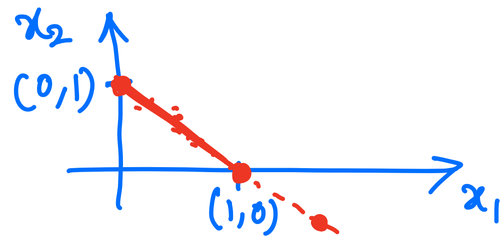
Convex Sets

$\lambda_1 \geq 0$, $1-\lambda_1 \geq 0 \Rightarrow \underline{0 \leq \lambda_1 \leq 1}$

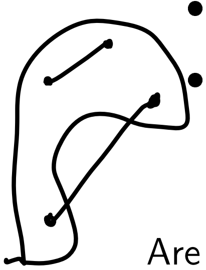
Definition 1. *Given a collection of points $x_1, x_2, \ldots, x_k$, the combination $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k$ is called* **Convex combination** *if $\lambda_i \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$.*
A set X is a **convex set** *if all convex combinations of its elements are in the set.*

$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $Z = \lambda_1 x + \lambda_2 y$

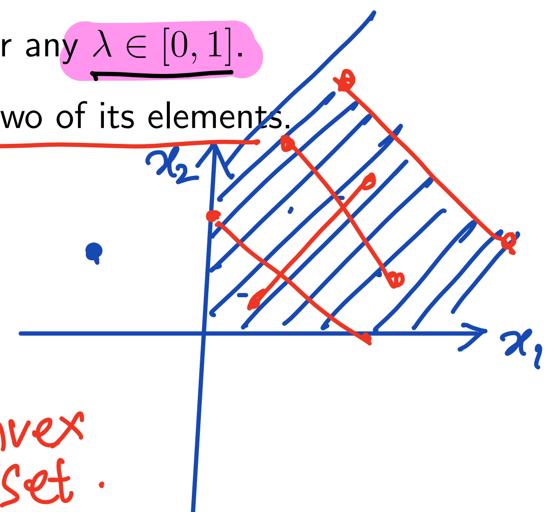
$Z = \lambda_1 x + (1-\lambda_1) y$



Equivalently, X is a convex set if



- **for every** $x, y \in X$, $\lambda x + (1-\lambda) y \in X$ for any $\lambda \in [0, 1]$.
- it contains **all** convex combinations of any two of its elements.



Are the following sets convex:

- $X_1 = \{x \in \mathbb{R}^2 | x_1 \geq 0, x_2 \geq 0\}$.
- $X_2 = \{x \in \mathbb{R}^2 | x_1 x_2 \geq 0\}$. $\Rightarrow$ not a convex set.

let $\bar{x}, \bar{y} \in X_1$

Let $\bar{z} = \lambda \bar{x} + (1-\lambda)\bar{y}$, $\lambda \in [0,1]$

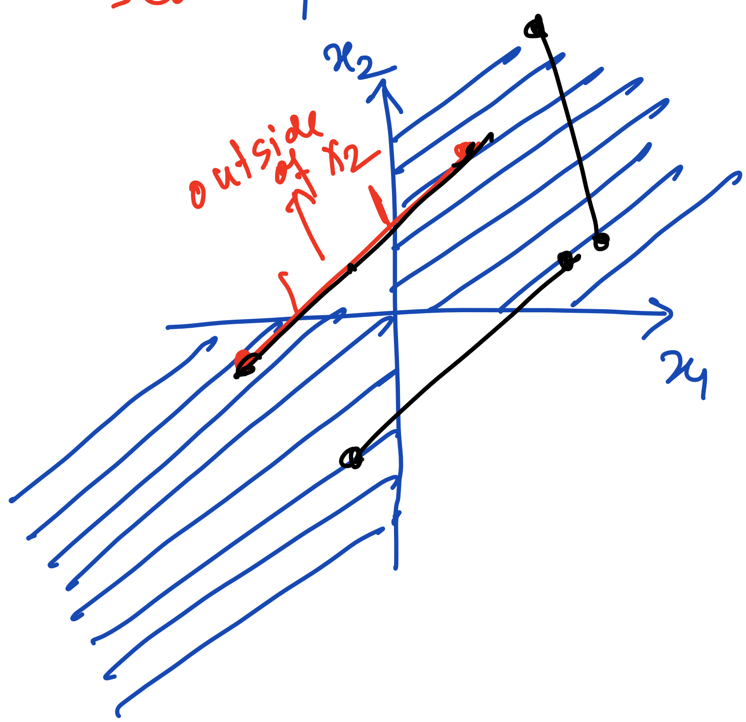
$= \lambda \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} + (1-\lambda) \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}$

$= \begin{bmatrix} \lambda \bar{x}_1 + (1-\lambda)\bar{y}_1 \\ \lambda \bar{x}_2 + (1-\lambda)\bar{y}_2 \end{bmatrix}$

$\geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

$\Rightarrow \bar{z} \in X_1$

Hence X_1 is a convex set.



Basic Examples of Convex Sets

Sets Defined by Linear Inequalities:

- Hyperplane: $H = \{x \in \mathbb{R}^n | a^\top x = b\}$ for some $a \in \mathbb{R}^n, b \in \mathbb{R}$.

Let $\bar{x}, \bar{y} \in H \Rightarrow a^T\bar{x} = b,\ a^T\bar{y} = b$

$\lambda \in [0,1]$. $\bar{z} = \lambda\bar{x} + (1-\lambda)\bar{y}$. To show that $\bar{z} \in H$,

we compute

$$a^T\bar{z} = a^T(\lambda\bar{x} + (1-\lambda)\bar{y}) = \lambda a^T\bar{x} + (1-\lambda)a^T\bar{y}$$
$$= \lambda b + (1-\lambda)b = b$$

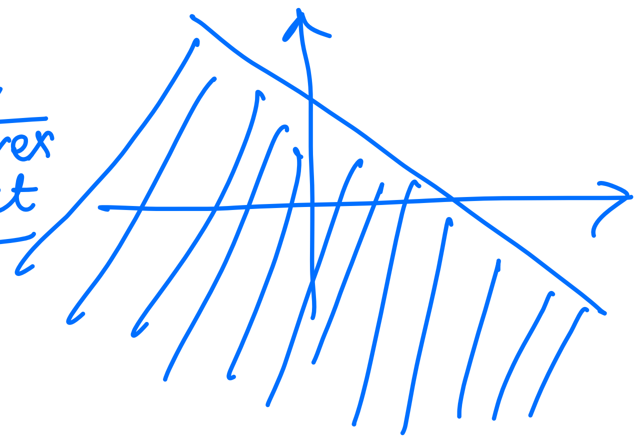
$\Rightarrow \bar{z} \in H$. $\Rightarrow$ H is a convex set.

- Halfspaces: $\{x \in \mathbb{R}^n | a^\top x \le b\}$ for some $a \in \mathbb{R}^n, b \in \mathbb{R}$.

convex set

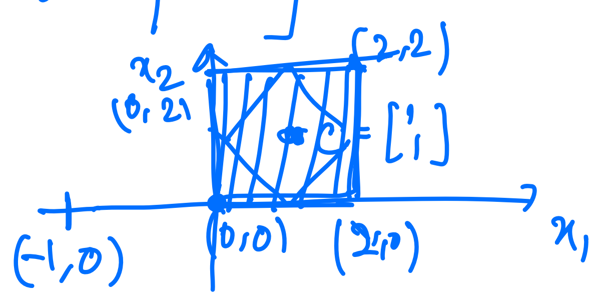
$$a^T\bar{z} = \lambda a^T\bar{x} + (1-\lambda)a^T\bar{y}$$
$$\le \lambda b + (1-\lambda)b$$
$$= b$$

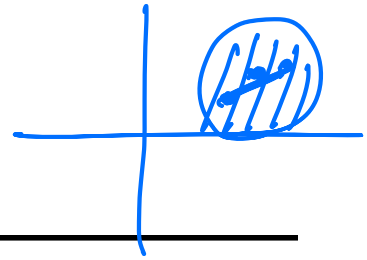
$\Rightarrow \bar{z} \in$ (halfspace)



$$B_\infty\left(\begin{bmatrix}1\\1\end{bmatrix}, 1\right) = \left\{x \in \mathbb{R}^2 \,\middle|\, \max_{i=1,2} |x_i - 1| \le 1\right\}$$

$\max(2,1) \le 1$: not true





Sets Defined by Norms

Consider the Ball $B_p(c, R) := \{x \in \mathbb{R}^n | \quad ||x - c||_p \leq R\}$ where

$$||x||_p := \begin{cases} \left(\sum_{i \in [n]} |x_i|^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \max_{i \in [n]} |x_i|, & p = \infty. \end{cases}$$

Recall that norm satisfies triangle inequality and positive homogeneity. We define $[n] := \{1, 2, \ldots, n\}$.

$g(\alpha x) = \alpha g(x) \quad \forall \alpha \geq 0$

Proposition 1. *$B_p(c, R)$ is a convex set.*

Let $\bar{x}, \bar{y} \in B_p(c,R)$. let $\lambda \in [0,1]$.

need to show: $\bar{z} = \lambda \bar{x} + (1-\lambda)\bar{y} \in B_p(c,R)$

$$\begin{aligned}
||\bar{z} - c||_p &= ||\lambda \bar{x} + (1-\lambda)\bar{y} - c||_p \\
&= ||\lambda \bar{x} + (1-\lambda)\bar{y} - \lambda c - (1-\lambda) c||_p \\
&= ||\lambda(\bar{x} - c) + (1-\lambda)(\bar{y} - c)||_p \\
\text{(triangle ineq.)} &\leq ||\lambda(\bar{x} - c)||_p + ||(1-\lambda)(\bar{y} - c)||_p \\
\text{(positive homogeneity)} &\leq \lambda \underbrace{||\bar{x} - c||_p}_{\leq R} + (1-\lambda) \underbrace{||\bar{y} - c||_p}_{\leq R} \\
&= \lambda R + (1-\lambda) R = R
\end{aligned}$$

$\Rightarrow \bar{z} \in B_p(c,R)$ & $B_p(c,R)$ is a convex set.

LECTURE - 4 , 8th Jan 2025

S^n : the set of all symmetric matrices of size $n \times n$

Positive Semidefinite Matrices

Proposition 2. *Set of symmetric positive semidefinite matrices, denoted by* $\mathcal{S}_n^+ := \{X \in S^n | X \succeq 0_{n\times n}\}$, *is a convex set.*

$A \succeq B \iff (A-B)$ is positive semi-definite.

Let X_1 & $X_2 \in S_n^+$.

Let $\lambda \in [0,1]$.

we need to show $Z := \lambda X_1 + (1-\lambda) X_2 \in S_n^+$.

Let $v \in \mathbb{R}^n$.

we evaluate $v^T Z v = v^T (\lambda X_1 + (1-\lambda) X_2) v$

$$= \lambda \underbrace{v^T X_1 v}_{\geq 0} + (1-\lambda) \underbrace{v^T X_2 v}_{\geq 0}$$

$$\geq 0.$$

thus, $Z \in S_n^+$.

Hence S_n^+ is a convex set.

Operations that preserve convexity of sets

$[m] = \{1, 2, \dots m\}$

Proposition 3 (Intersection). *If $X_1, X_2, \dots, X_m$ are convex sets, then $\cap_{i\in[m]} X_i$ is a convex set.*

Let $Z := \bigcap_{i\in[m]} X_i$, let $z_1, z_2 \in Z$, let $\lambda \in [0,1]$.

We need to show that $\bar{z} = \underbrace{\lambda z_1 + (1-\lambda) z_2}_{} \in Z$.

$z_1 \in Z \Leftrightarrow z_1 \in \bigcap_{i\in[m]} X_i$

$\Leftrightarrow$. $z_1 \in X_1, z_1 \in X_2, \dots, z_1 \in X_m$

. $z_2 \in X_1, z_2 \in X_2, \dots, z_2 \in X_m$

Since X_i is a convex set, $z_1 \in X_i, z_2 \in X_i, \lambda \in [0,1]$

$\Rightarrow \bar{z} \in X_i \quad \forall i \in [m]$.

$\Rightarrow \bar{z} \in \bigcap_{i\in[m]} X_i = Z$.

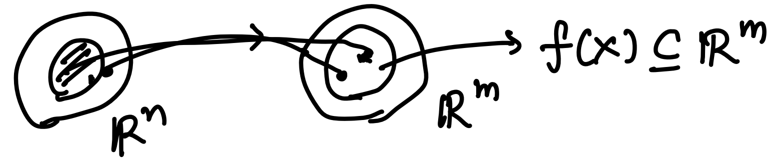
Therefore, Z is a convex set.

Example: Polyhedron $\{x \in \mathbb{R}^n | Ax \le b\}$ for some $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ which is an intersection of half-spaces.

$$A = \begin{bmatrix} - a_1^T - \\ - a_2^T - \\ \vdots \\ - a_m^T - \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$X_i = \{x \in \mathbb{R}^n \mid a_i^T x \le b_i\}$

Then, $\{x \in \mathbb{R}^n \mid Ax \le b\} = \bigcap_{i\in[m]} X_i$.



Operations that preserve convexity of sets

Proposition 4 (Affine Image)**.** *If X is a convex set, $f(x) = Ax + b$ with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, then the set $f(X) := \{y | y = Ax + b \quad \text{for some} \quad x \in X\}$ is a convex set.*

Let $y_1, y_2 \in f(X)$, and let $\lambda \in [0,1]$.

We need to show $\bar{y} = \lambda y_1 + (1-\lambda) y_2 \in f(X)$.

i.e., $\exists\, \bar{x} \in X$ s.t $\bar{y} = A\bar{x} + b$.

$y_1 = Ax_1 + b$ for some $x_1 \in X$

$y_2 = Ax_2 + b$ for some $x_2 \in X$.

$\underbrace{\lambda x_1 + (1-\lambda) x_2}_{\bar{x}} \in X$ due to convexity of X.

$$
\begin{aligned}
A\bar{x} + b &= A(\lambda x_1 + (1-\lambda)x_2) + b = \lambda A x_1 + (1-\lambda) A x_2 + b \\
&= \lambda A x_1 + (1-\lambda) A x_2 + \lambda b + (1-\lambda) b \\
&= \lambda (Ax_1 + b) + (1-\lambda)(Ax_2 + b) \\
&= \lambda y_1 + (1-\lambda) y_2 = \bar{y}
\end{aligned}
$$

Consequently $\bar{y} \in f(X)$. Hence $f(X)$ is a convex set.

Operations that preserve convexity of sets

$\subseteq \mathbb{R}^{n_1}$ $\subseteq \mathbb{R}^{n_2}$

Proposition 5 (Product). *If $X_1, X_2, \ldots, X_m$ are convex sets, then*

$$X = X_1 \times X_2 \times \ldots \times X_m := \{(x_1, x_2, \ldots, x_m) \mid x_i \in X_i, i \in [m]\}$$

is a convex set. $\subseteq \mathbb{R}^{(n_1+n_2+\ldots n_m)}$

$$x \in X, \quad x = \begin{bmatrix} x_1 \\ \hline x_2 \\ \hline \vdots \\ \hline x_m \end{bmatrix}$$

Proposition 6 (Weighted Sum). *If $X_1, X_2, \ldots, X_m$ are convex sets, then* $\bar{X} :=$ $\sum_{i\in[m]} \alpha_i X_i := \{y \mid y = \sum_{i\in[m]} \alpha_i x_i, \quad x_i \in X_i\}$ *is a convex set for $\alpha_i \in \mathbb{R}$.*

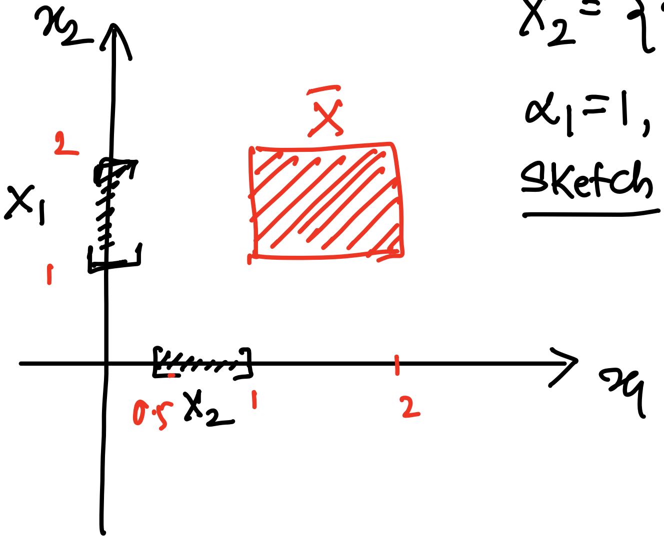
Example:

$X_1 = \{x \in \mathbb{R}^2 \mid x_1 = 0, x_2 \in [1,2]\}$

$X_2 = \{x \in \mathbb{R}^2 \mid x_1 \in [0.5, 1], x_2 = 0\}$.

$\alpha_1 = 1, \quad \alpha_2 = 2$

Sketch $\quad \bar{X} = \alpha_1 X_1 + \alpha_2 X_2$



Let $\bar{x} \in \bar{X}$,

$\bar{x} = \bar{x}_1 + 2\bar{x}_2$, where $\bar{x}_1 \in X_1$, $\bar{x}_2 \in X_2$

$= \begin{bmatrix} 0 \\ \bar{x}_{12} \end{bmatrix} + 2\begin{bmatrix} \bar{x}_{21} \\ 0 \end{bmatrix}$, $\bar{x}_{21} \to [0.5, 1]$, $\bar{x}_{12} \in [1,2]$

$= \begin{bmatrix} [1,2] \\ [1,2] \end{bmatrix}$

proof: Homework.

Operations that preserve convexity of sets

Proposition 7 (Inverse Affine Image). *Let $X \in \mathbb{R}^n$ be a convex set and $\mathcal{A} : \mathbb{R}^m \to \mathbb{R}^n$ be an affine map with $\mathcal{A}(y) = Ay + b$ for matrix A and vector b of suitable dimension. Then, the set $\mathcal{A}^{-1}(X) := \{y \in \mathbb{R}^m \mid Ay + b \in X\}$ is a convex set.*

Let $y_1, y_2 \in A^{-1}(X)$, let $\lambda \in [0,1]$

$\Rightarrow$ $Ay_1 + b \in X$

$Ay_2 + b \in X$.

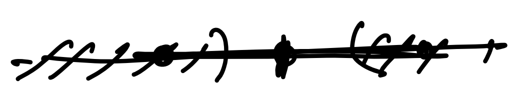
we need to show $\bar{y} = \lambda y_1 + (1-\lambda) y_2 \in A^{-1}(X)$

or equiv. $A\bar{y} + b \in X$.

we evaluate
$$
\begin{aligned}
A\bar{y} + b &= A(\lambda y_1 + (1-\lambda) y_2) + b \\
&= \lambda A y_1 + (1-\lambda) A y_2 + \lambda b + (1-\lambda) b \\
&= \lambda [\underbrace{A y_1 + b}_{\in X}] + (1-\lambda) [\underbrace{A y_2 + b}_{\in X}] \\
&\in X
\end{aligned}
$$
due to convexity of X.

Problem: Let X_1 and X_2 be convex sets. Determine if $X_1 \setminus X_2$ is convex.

$X_1 = \mathbb{R}$, $X_2 = [0,1]$



$\underbrace{X_1 \setminus X_2}_{\Downarrow\ \text{not a convex set.}} = (-\infty, 0) \cup (1, \infty)$.

$\hookrightarrow$ Let us pick points $-1, 2 \in X_1 \setminus X_2$

Let us pick $\lambda = 0.5$

$\lambda x_1 + (1-\lambda) x_2 = \frac{-1+2}{2} = 0.5 \notin X_1 \setminus X_2$.

Lecture 5 : 9th Jan 2025

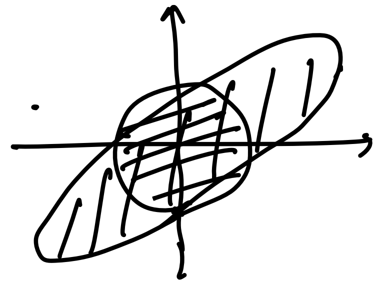
Ellipsoid

Proposition 8. *Let A be a symmetric positive definite matrix. Then, the set $\mathcal{E} := \{x \in \mathbb{R}^n | (x-c)^\top A^{-1}(x-c) \leq 1\}$ is convex.*

Let us try to show

$$\mathcal{E} = f(B_2(c, r)),$$

$f(x)$ -



$A^{-1} = \Sigma^\top \Sigma$,

Find $f(x) = Gx + h$

such that when $x \in B_2(c, r)$, then $f(x) \in \mathcal{E}$.

$G = \Sigma^{-1}$, $h = c$

We need to see if $\Sigma^{-1}x + c$ belongs to $\mathcal{E}$.

i.e., $(\Sigma^{-1}x + \cancel{c} - \cancel{c})^\top A^{-1}(\Sigma^{-1}x + \cancel{c} - \cancel{c})$

$$= x^\top (\Sigma^{-1})^\top A^{-1} \Sigma^{-1} x$$

$$= x^\top \underbrace{(\Sigma^{-1})^\top \Sigma^\top}_{I} \underbrace{\Sigma \Sigma^{-1}}_{I} x = x^\top x.$$

Let us choose $c = 0$ & $r = 1$.

Then, we have shown that $x \in B_2(0, 1) \Rightarrow \Sigma^{-1}x + c \in \mathcal{E}$.

Thus, $\mathcal{E}$ is a convex set.

Convex Combination

Given a collection of points $x_1, x_2, \ldots, x_k$, the combination $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k$ is called **Convex** if $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$.

Equivalent Definition:

Definition 4 (Convex Set)**.** *A set is convex if it contains all convex combinations of its points.*

Definition 5 (Convex Hull)**.** *The convex hull of a set $X \in \mathbb{R}^n$ is the set of all convex combinations of its elements, i.e.,*

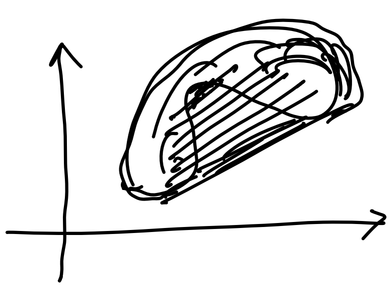
$$\texttt{conv}(X) := \left\{ y \in \mathbb{R}^n \mid y = \sum_{i \in [k]} \lambda_i x_i, \mathit{where} \lambda_i \geq 0, \sum_{i \in [k]} \lambda_i = 1, x_i \in X \forall i \in [k], k \in \mathbb{N} \right\}.$$

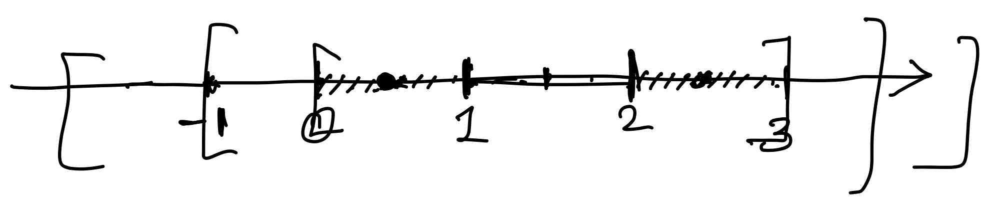
Proposition 9 (Convex Hull)**.** *The following are true.*

- $\texttt{conv}(X)$ *is a convex set (even when X is not).*
- *If X is convex, then* $\texttt{conv}(X) = X$.
- *For any set X,* $\texttt{conv}(X)$ *is the smallest convex set containing X.*

Example: Determine the convex hull of $X = [0, 1] \cup [2, 3]$.

$\text{conv}(X) = [0,3]$

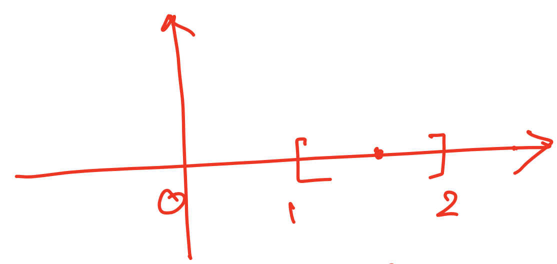




Combination of points

Given a collection of points $x_1, x_2, \ldots, x_k$, the combination $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_k x_k$ is called

- **Convex** if $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$.
- **Conic** if $\lambda_i \geq 0$,
- **Affine** if $\sum_{i=1}^{n} \lambda_i = 1$,
- **Linear** if $\lambda_i \in \mathbb{R}$.



$X = [1, 2]$ is a convex set, but not a cone.

A set is convex/ convex cone/ affine subspace/linear subspace if it contains all convex/conic/affine/linear combinations of its elements.

A set which is convex and a cone, then it is called a convex cone.

Definition 6. *A set X is a cone if for any $x \in X, \alpha \geq 0$, we have $\alpha x \in X$.*

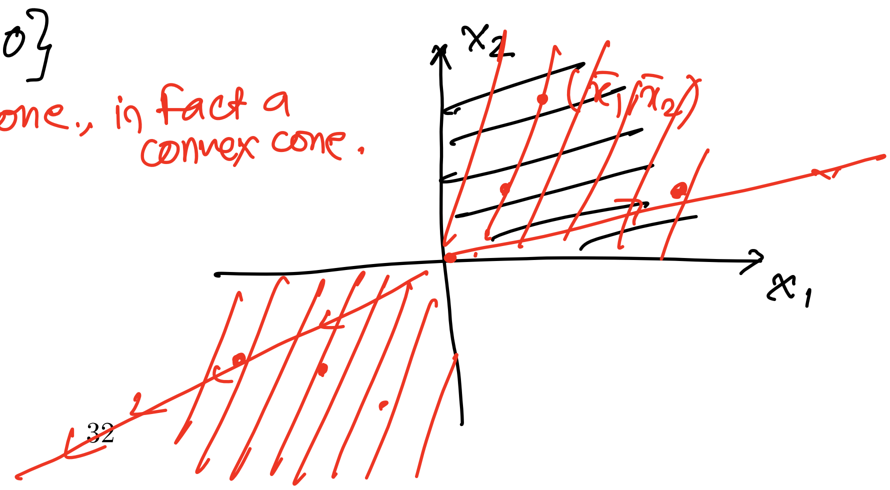
Note: Every cone must include the origin. Union of two cones is a cone.

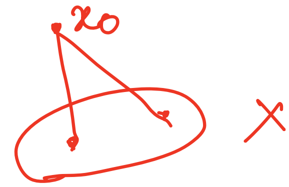
$X_1 = \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0\}$

$\Rightarrow$ is a cone, in fact a convex cone.

$X_2 = \{x \in \mathbb{R}^2 \mid x_1 x_2 \geq 0\}$

$\hookrightarrow$ is a cone, but not a convex set.





Projection

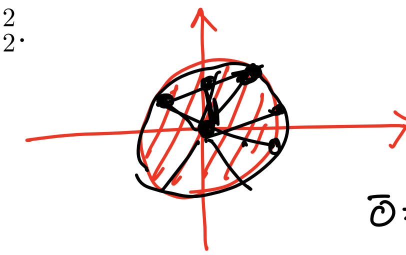
If $x_0 \in X$, $\texttt{proj}_X(x_0) = x_0$

Definition 7 (Projection). *The projection of a point x_0 on a set X, denoted $\texttt{proj}_X(x_0)$ is defined as*

$$\texttt{proj}_X(x_0) := \text{argmin}_{x \in X} \, ||x - x_0||_2^2.$$

$X_1 = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \le 1\}$ $\quad \texttt{proj}_{X_1}(\bar{0}) = \{\bar{0}\}$

$X_2 = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$ $\quad \texttt{proj}_{X_2}(\bar{0}) = X_2$



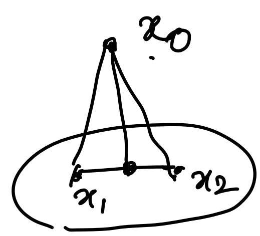
$\bar{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Theorem 4: Projection Theorem

If X is closed and convex, then $\texttt{proj}_X(x_0)$ exists and is unique.

Main idea:

- Existence due to Weierstrass Theorem
- Uniqueness via contradiction exploiting convexity



Suppose $x_1, x_2 \in \texttt{proj}_X(x_0)$, and $x_1 \neq x_2$.

$\|x_1 - x_0\|_2 = \|x_2 - x_0\|_2 = d_{\min}$, $\quad \|x - x_0\|_2 > d_{\min}$ for $x \notin \texttt{proj}_X(x_0)$.

Let us find distance between x_0 & $\frac{x_1 + x_2}{2}$.

$$x_0 - \frac{x_1 + x_2}{2} = \frac{1}{2}(2x_0 - x_1 - x_2) = \frac{1}{2}(x_0 - x_1 + x_0 - x_2)$$

Let us evaluate the following.

$$\|x_0 - x_1 + x_0 - x_2\|_2^2 = \|x_0 - x_1\|_2^2 + \|x_0 - x_2\|_2^2 + 2(x_0 - x_1)^T(x_0 - x_2)$$

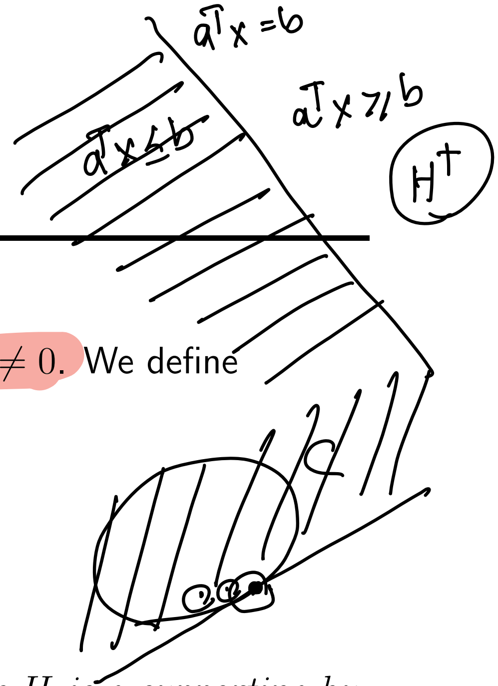
$$\|x_0 - x_1 - (x_0 - x_2)\|_2^2 = \|x_0 - x_1\|_2^2 + \|x_0 - x_2\|_2^2 - 2(x_0 - x_1)^T(x_0 - x_2)$$

(+)

$$\|x_0 - x_1 + x_0 - x_2\|_2^2 + \|x_2 - x_1\|_2^2 = 2\left[\|x_0 - x_1\|_2^2 + \|x_0 - x_2\|_2^2\right]$$

$$\Rightarrow 4\left\|x_0 - \frac{x_1 + x_2}{2}\right\|_2^2 + \|x_2 - x_1\|_2^2 = 4 d_{\min}^2$$

$$\Rightarrow \left\|x_0 - \underbrace{\frac{x_1 + x_2}{2}}_{\in X}\right\|_2^2 \underbrace{<}_{\text{due to convexity}} d_{\min}^2 \Rightarrow$$ x_1 and x_2 cannot be projections.



Supporting Hyperplane

Consider a hyperplane $H = \{x \in \mathbb{R}^n \mid a^\top x = b\}$ with $a \neq 0$. We define

$$H^- := \{x \in \mathbb{R}^n \mid a^\top x \leq b\}.$$

Definition 8 (Supporting Hyperplane). *A hyperplane H is a supporting hyperplane for a convex set C at a boundary point $z \in \delta C$ if $z \in C$ and $C \subseteq H^-$.*

$z \in H$.

set of all boundary points of C.

Theorem 5: Supporting Hyperplane Theorem

If C is a convex set and $z \in \delta C$ is a boundary point, then there exists a supporting hyperplane for C at z.

Example: $X = \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$

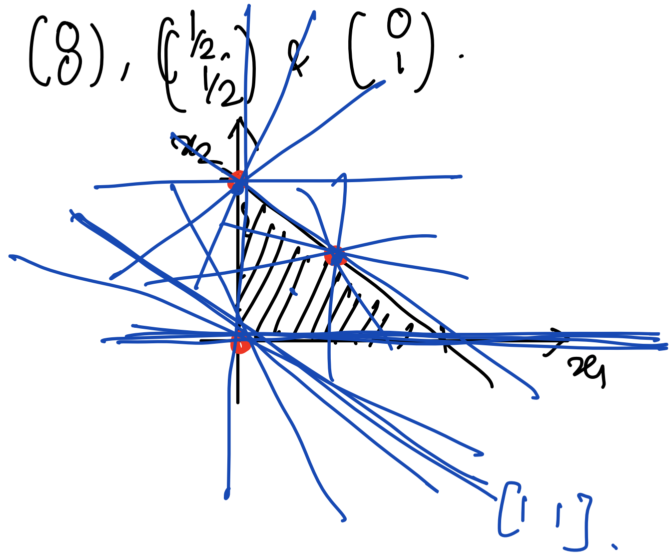
find supporting hyperplances at $\begin{pmatrix}0\\0\end{pmatrix}$, $\begin{pmatrix}1/2\\1/2\end{pmatrix}$ & $\begin{pmatrix}0\\1\end{pmatrix}$.

$H = \{x \in \mathbb{R}^2 \mid a^T x = b\}$

at $\begin{pmatrix}0\\0\end{pmatrix}$: $H = \{x \in \mathbb{R}^2 \mid [-1\ 0]^T x = 0\}$

at $\begin{pmatrix}0.5\\0.5\end{pmatrix}$: $H = \{x \in \mathbb{R}^2 \mid x_1 + x_2 = 1\}$

at $\begin{pmatrix}0\\1\end{pmatrix}$:

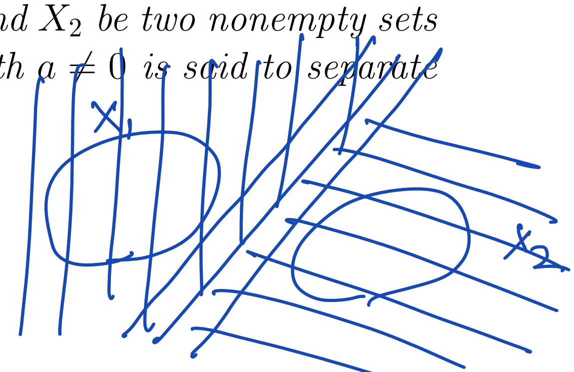


LECTURE 6 : 9th Jan 2025

Separating Hyperplane

Definition 9 (Separating Hyperplane). *Let X_1 and X_2 be two nonempty sets in $\mathbb{R}^n$. A hyperplane $H = \{x \in \mathbb{R}^n \mid a^\top x = b\}$ with $a \neq 0$ is said to separate X_1 and X_2 if*

- $X_1 \subseteq H^- := \{x \in \mathbb{R}^n \mid a^\top x \leq b\}$,
- $X_2 \subseteq H^+ := \{x \in \mathbb{R}^n \mid a^\top x \geq b\}$.



Separation is said to be **strict** *if $X_1 \subset \{x \in \mathbb{R}^n \mid a^\top x \leq b'\}$, $X_2 \subset \{x \in \mathbb{R}^n \mid a^\top x \geq b''\}$ with $b' < b''$.*

$X_1 \subseteq H^-$, $X_2 \subseteq H^{++} = \{x \in \mathbb{R}^2 \mid a^\top x > b\}$

Equivalently

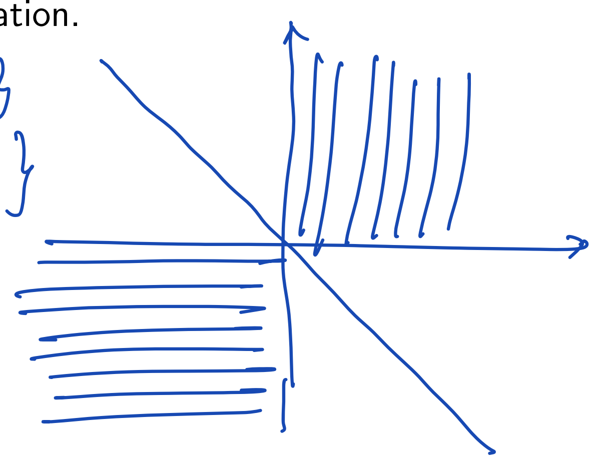
$$\sup_{x \in X_1} a^\top x \leq \inf_{x \in X_2} a^\top x$$

with the inequality being strict for strict separation.

Examples:

$X_1 = \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0\}$

$X_2 = \{x \in \mathbb{R}^2 \mid x_1 \leq 0, x_2 \leq 0\}$

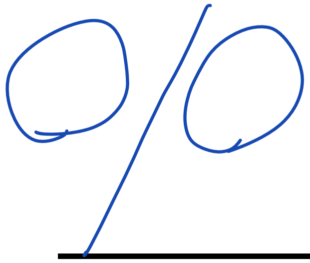


$H = \{x \in \mathbb{R}^2 \mid [1, 1]x = 0\}$

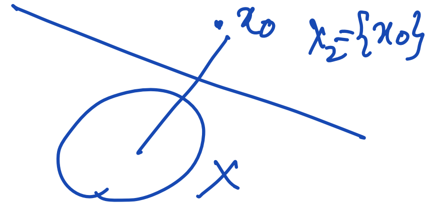
$x \in X_1$, then $[1, 1]x \geq 0$

$x \in X_2$, then $[1, 1]x \leq 0$.

The separation is not strict.

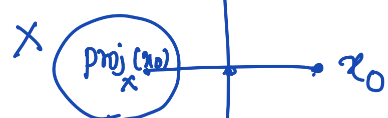


Separating Hyperplane Theorem



Theorem 6: Separating Hyperplane Theorem

- Let X_1 and X_2 be convex sets with $X_1 \cap X_2 = \phi$. Then, there exists a separating hyperplane for X_1 and X_2. If X_1 is closed and bounded, and X_2 is closed, then X_1 and X_2 can be strictly separated.
- Let X be a closed convex set and $x_0 \notin X$. Then, there exists a hyperplane that strictly separates x_0 and X.



We will prove the second statement. Main Idea:

1. Let $H = \{x \in \mathbb{R}^n \mid a^\top x = b\}$ with $a = x_0 - \texttt{proj}_X(x_0)$ and $b = \boxed{a^\top x_0 - \frac{\|a\|_2^2}{2}}$.
2. Use properties of projection and convexity of X to verify that H is indeed the separating hyperplane.

mid point : $\left(\frac{x_0 + \text{proj}_X(x_0)}{2}\right)$

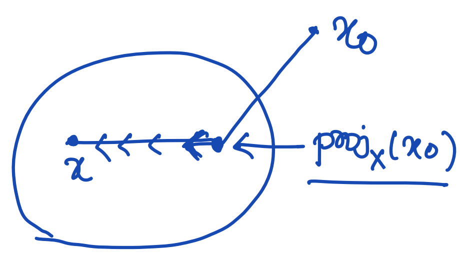
we can show that

$$a^T\left(\frac{x_0 + \text{proj}_X(x_0)}{2}\right) = b$$

To prove the result, we need to show $a^T x_0 > b$, and $a^T x < b$ for all $x \in X$.

It is easy to see that,

$$a^T x_0 > b$$



Since X is a convex set, any point

$$\lambda x + (1-\lambda)\text{proj}_X(x_0) \in X$$

$$\phi(\lambda) = \|\lambda x + (1-\lambda)\text{proj}_X(x_0) - x_0\|_2^2$$

Since $\text{proj}_X(x_0)$ minimizes the distance between x_0 & elements in X, we must have $\phi'(\lambda) \geq 0$.

$$\phi(\lambda) = \|\lambda(x - \text{proj}_X(x_0)) + \text{proj}_X(x_0) - x_0\|_2^2$$

$$= \lambda^2\|x - \text{proj}_X(x_0)\|_2^2 + \|\text{proj}_X(x_0) - x_0\|_2^2 + 2\lambda(x - \text{proj}_X(x_0))^T(\text{proj}_X(x_0) - x_0)$$

$\phi'(\lambda)\big|_{\lambda=0} = 2(x - \text{proj}_X(x_0))^T (\underbrace{\text{proj}_X(x_0) - x_0}_{-a}) \geq 0$

$= 2a^T \text{proj}_X(x_0) - 2a^T x \geq 0. \Rightarrow a^T x \leq \underbrace{a^T \text{proj}_X(x_0)}_{} = a^T(x_0 - a)$

Theorem of the Alternative (Farkas' Lemma)

$= a^T x_0 - a^T a$

$< b.$

Suppose $S_1 \neq \emptyset \Rightarrow \exists\ \bar{x} \in \mathbb{R}^n$ satisfying $A\bar{x} = b$, $\bar{x} \geq 0$.

Suppose S_2 is also non-empty, let $\bar{y} \in S_2$.

$$(\bar{y})^T A\bar{x} = (\bar{y})^T b > 0$$

$$\Rightarrow (\bar{y})^T A\bar{x} > 0$$

If $A^T\bar{y} = \bar{y}^T A \leq 0$, and $\bar{x} \geq 0$, then $(\bar{y})^T A\bar{x} \leq 0$

Thus, we have a contradiction, and hence we must have $S_2 = \emptyset$.

Lemma 1 (Farkas' Lemma)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then, exactly one of the following sets must be empty:*

1. $\{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ $= S_1$
2. $\{y \in \mathbb{R}^m \mid A^\top y \leq 0, b^\top y > 0\}$. $= S_2$

Insight: If unable to show a system of linear inequalities does not have a solution, try to show that its alternative system does.

Main Idea:

1. Easy to show that if (2) is feasible, (1) is infeasible.
   or $S_1 = \emptyset$.
2. For the converse, suppose (1) is infeasible. Then, $b \notin \text{cone}(a_1, a_2, \ldots, a_n)$ where a_i is the i-th column of A. Find a hyperplane separating b from $\text{cone}(a_1, a_2, \ldots, a_n)$ and show that (2) is feasible.

there is no $\bar{x} \geq 0$ satisfying $A\bar{x} = b$ $= \begin{bmatrix} a_1 \\ \vdots \end{bmatrix} x_1 + \begin{bmatrix} a_2 \\ \vdots \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_n \\ \ \end{bmatrix} x_n$

cone $C = \{y \in \mathbb{R}^m \mid y = \sum_{i=1}^{n} a_i x_i,\ x_i \geq 0\}$

$= \text{cone}(a_1, a_2 \ldots a_n)$

$S_1 = \emptyset \Leftrightarrow b \notin \underbrace{\text{cone}(a_1, a_2 \ldots a_n)}$ $\to$ convex set, can be shown to be a closed set.

then, there exists a hyperplane $H = \{x \in \mathbb{R}^n \mid g^T x = h\}$

that strictly separates b from $\text{cone}(a_1, a_2 \ldots a_n)$

$\Rightarrow \quad g^T b > h$, and $g^T x \le h$ for all $x \in cone(a_1, a_2 \; a_n)$.

Proof

It remains to show that $h = 0$.

Since origin $\bar{0} \in cone(a_1, .. a_n)$, we always have $g^T(0) \le h$

$\Rightarrow h \geq 0$

Suppose $h > 0$.

$\left(g^T b \geq 5, \; g^T x \le 5 \text{ for all } x \in cone(a_1 .. \; a_n) \right)$

suppose $\bar{x} \in cone(a_1, a_2 .. a_n)$, and $g^T \bar{x} > 0$

Since X is a cone, $\alpha \bar{x} \in cone(a_1, ... a_n) \; \forall \alpha \geq 0$

we will be able to find α large enough s.t.

$g^T(\alpha \bar{x}) > h$ which will violate strict separation condition of the hyperplane. $\Rightarrow$ we can always choose $h = 0$

Thus $\forall \bar{x} \in cone(a_1 .. a_n)$, $g^T \bar{x} \le 0$.

$\Rightarrow g^T a_1 \le 0, \; .. \; g^T a_n \le 0$

$\Rightarrow g^T A \le 0, \; g^T b > 0$

$\Rightarrow g \in S_2$.

LECTURE 7: 10th Jan 2025

Domain of a Function

- We consider *extended real-valued* functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} =: \bar{\mathbb{R}}$.

- The (effective) domain of f, denoted $\mathtt{dom}(f)$, is the set $\{x \in \mathbb{R}^n \mid |f(x)| < +\infty\}$.
  i.e., $-\infty < f(x) < +\infty$

- Example: $f(x) = \frac{1}{x}$. What is $\mathtt{dom}(f)$? $f : \mathbb{R} \to \mathbb{R}$ $\quad \mathrm{dom}(f) = \mathbb{R} \setminus \{0\}$

- $f(x) = \sum_{i=1}^n x_i \log(x_i)$. What is $\mathtt{dom}(f)$? $\quad \lim_{x \to 0} x \log x = 0$
  $x \in \mathbb{R}^n$

- When $\mathtt{dom}(f) \neq \phi$, we say that the function f is *proper*.

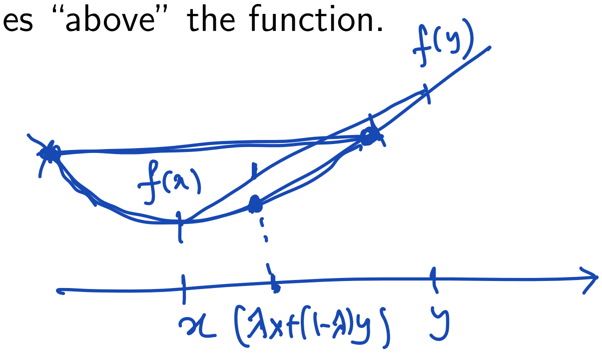
$\mathrm{dom}(f) = \{x \in \mathbb{R}^n \mid x_i \geq 0\}$

Convex Functions

Definition 10 (Convex Function). *A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex if*

1. $\mathtt{dom}(f) \subseteq \mathbb{R}^n$ *is a convex set, and*
2. **for every** $x, y \in \mathtt{dom}(f), \lambda \in [0, 1]$, *we have* $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$.

The Line segment joining $(x, f(x))$ and $(y, f(y))$ lies "above" the function.



Examples:

- $f_1(x) = x^2$, $\mathrm{dom}(f_1) = \mathbb{R}$
- $f_2(x) = e^x$: Homework
- $f_3(x) = a^\top x + b$ for $x \in \mathbb{R}^n$

* . $f_1(\lambda x + (1-\lambda)y) = (\lambda x + (1-\lambda)y)^2 = \lambda^2 x^2 + (1-\lambda)^2 y^2 + 2\lambda(1-\lambda)xy$

$\leq \lambda x^2 + (1-\lambda)y^2$ (we want to show)

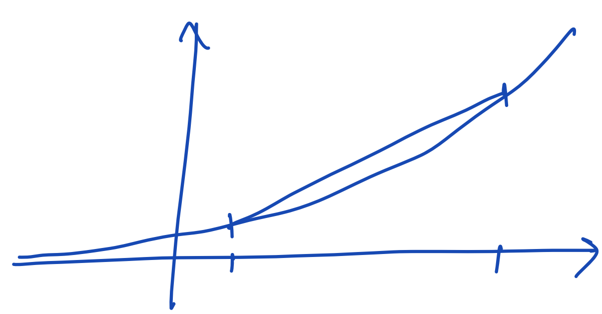
$\lambda^2 \leq \lambda$, $(1-\lambda)^2 \leq (1-\lambda)$ since $\lambda \in [0,1]$. (Homework)

* $\mathrm{dom}(f_3) = \mathbb{R}^n$

$f_3(\lambda x + (1-\lambda)y) = a^T(\lambda x + (1-\lambda)y) + b$

$= \lambda(a^T x + b) + (1-\lambda)(a^T y + b) = \lambda f_3(x) + (1-\lambda) f_3(y)$.

$\Rightarrow f_3$ is a convex function.



Example: Norms

Definition 11 (Norms)**.** *A function* $\pi : \mathbb{R}^n \to \bar{\mathbb{R}}$ *is a norm if*

- $\pi(x) \geq 0, \quad \forall x$ *and* $\pi(x) = 0$ *if and only if* $x = 0$*,*
- $\pi(\alpha x) = |\alpha| \pi(x)$ *for all* $\alpha \in \mathbb{R}$*,* : positive homogeneity
- $\pi(x + y) \leq \pi(x) + \pi(y)$. : triangle inequality.

Examples:

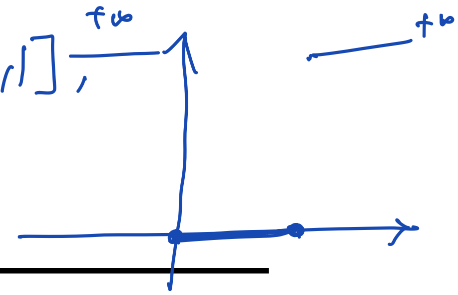
- $||x||_p := (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$ for $p \geq 1$.
- $||x||_Q := \sqrt{x^\top Q x}$ where Q is a positive definite matrix.
- $||A||_F := (\sum_{i=1}^m \sum_{j=1}^n |A_{i,j}|^2)^{1/2}$ Frobenius norm on $\mathbb{R}^{m \times n}$.

Proposition 10. *A Norm is a convex function.*

Let $x, y \in dom(\pi)$., $\lambda \in [0,1]$

$$\pi(\lambda x + (1-\lambda) y) \leq \pi(\lambda x) + \pi[(1-\lambda) y]$$
$$\leq \lambda \pi(x) + (1-\lambda) \pi(y).$$

$C = [0,1]$,



Example: Indicator Function

Definition 12. *Indicator function $I_C(x)$ of a set C is defined as*

$$I_C(x) := \begin{cases} 0, & x \in C, \\ \infty, & x \notin C. \end{cases}$$

Proposition 11. *Indicator function $I_C(x)$ is convex* function *if the set C is a convex set.*

$\mathrm{dom}(I_C(x)) = C$ which is a convex set.

Let $x, y \in C$, $\lambda \in [0,1]$.

$$I_C(\underbrace{\lambda x + (1-\lambda)y}_{\in C}) = 0$$

$$\|$$

$$\lambda I_C(x) + (1-\lambda) I_C(y) = 0$$

Hence $I_C(x)$ is a convex function.

Example: Support Function

show that this set is convex. (HW.)

Proposition 12. *Support function of a set C is defined as $I_C^*(x) := \sup_{y\in C} x^\top y$. Support function of a set is always a convex function.*

Let $x_1, x_2 \in \boxed{\mathrm{dom}\,(I_C^*(x))}$, $\lambda \in [0,1]$

$$I_C^*\left(\lambda x_1 + (1-\lambda) x_2\right) = \sup_{y\in C} (\lambda x_1 + (1-\lambda) x_2)^T y$$

$$= \sup_{y\in C} \left[\lambda x_1^T y + (1-\lambda) x_2^T y\right]$$

we want to show

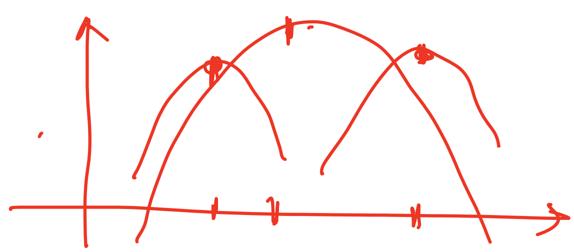
$$\leq \lambda I_C^*(x_1) + (1-\lambda) I_C^*(x_2)$$

$$= \lambda \sup_{y\in C} x_1^T y + (1-\lambda) \sup_{y\in C} x_2^T y$$

$\lambda x_1^T y \equiv f_1(y)$, $(1-\lambda) x_2^T y \equiv f_2(y)$

$$\sup_{y\in C} \left[f_1(y) + f_2(y)\right] \leq \sup_{y\in C} \left[f_1(y)\right] + \sup_{y\in C} \left[f_2(y)\right]$$

(always true)



$$f_1(\bar y) + f_2(\bar y) \leq \sup_{y\in C} \left[f_1(y) + f_2(y)\right] \quad \forall \bar y.$$

$$\sup_{y\in C} f_1(y) + \sup_{y\in C} f_2(y) \qquad f_1(\bar y_1) + f_2(\bar y_2) \quad \forall \bar y_1, \bar y_2$$

$$\geq \underbrace{f_1(\bar y) + f_2(\bar y) \quad \forall \bar y}$$

$$\Rightarrow \sup_{y\in C} f_1(y) + \sup_{y\in C} f_2(y) \geq \sup_{y\in C} \left[f_1(y) + f_2(y)\right].$$

Special Types of Convex Functions

Definition 13. *A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is*

- **strictly convex** *if property (2) above holds with strict inequality for $\lambda \in (0, 1)$,* $\Rightarrow f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$, $x \neq y$, $\lambda \in (0,1)$
- μ-**strongly convex** *if $f(x) - \mu\frac{||x||_2^2}{2}$ is convex, and*, $\mu > 0$.
- **concave** *if $-f(x)$ is convex.*

equivalently, a function g is concave if dom g is a convex set

$\forall x, y \in \text{dom } g$, $\lambda \in [0,1]$,

$$g(\lambda x + (1-\lambda)y) \geq \lambda g(x) + (1-\lambda) g(y)$$

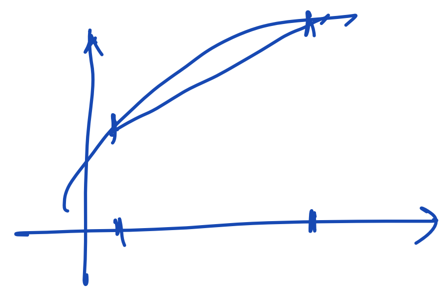
If a function is affine, it is both convex and concave.

Examples:

$g_1(x) = -x^2$

$g_2(x) = \sqrt{x}$

$g_3(x) = \log x$



Jensen's Inequality

Proposition 13. *For a convex function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$, for any collection of points $\{x_1, x_2, \ldots, x_k\}$, we have $f(\sum_{i=1}^k \lambda_i x_i) \leq \sum_{i=1}^k \lambda_i f(x_i)$ when $\lambda_i \geq 0$ and $\sum_{i=1}^k \lambda_i = 1$.*

If g is concave, $\sum_{i=1}^{k} \lambda_i g(x_i) \leq g\left(\sum_{i=1}^{k} \lambda_i x_i\right)$

Proof is straightforward via induction.

Ex:- For a collection $(x_1, x_2 \ldots x_n)$,

$$AM = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$GM = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

$$\log\left(\prod_{i=1}^{n} x_i\right)^{1/n} = \frac{1}{n}\left(\log\left(\prod_{i=1}^{n} x_i\right)\right)$$

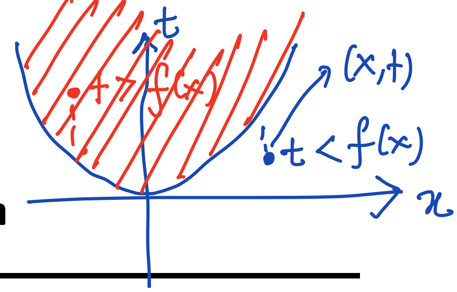
$$= \frac{1}{n} \sum_{i=1}^{n} \log(x_i) \qquad \left(\text{see: } \lambda_i = 1/n,\ \underset{\text{concave}}{f} : \log\right)$$

$$\leq \log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \qquad (\text{using Jensen's inequality}).$$

$$\Rightarrow \left(\prod_{i=1}^{n} x_i\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (\text{as } \log(\cdot) \text{ is monotonically increasing})$$

$f(x) = x^2$, $f : \mathbb{R} \to \mathbb{R}$.

$\text{epi}(f) \subseteq \mathbb{R}^2$



Epigraph Characterization

Definition 14. *A epigraph of a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is defined as the set*

$$\texttt{epi}(f) := \{(x,t) \in \mathbb{R}^{n+1} | f(x) \le t\}. \subseteq \mathbb{R}^{n+1}$$

$(x, f(x)) \in \text{epi}(f)$

Example: Norm cone: $\{(x,t) | ||x|| \le t\}$ is a convex set. since $\|\cdot\|$ is a convex function.

Proposition 14. *Function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex in $\mathbb{R}^n$ if and only if its epigraph is a convex set in $\mathbb{R}^{n+1}$.*

proof: Let f be a convex function.

Let $\begin{pmatrix} x_1 \\ t_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ t_2 \end{pmatrix} \in \text{epi}(f)$. $\Rightarrow f(x_1) \le t_1$, $f(x_2) \le t_2$

we need to show $\lambda \begin{pmatrix} x_1 \\ t_1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} x_2 \\ t_2 \end{pmatrix} \in \text{epi}(f)$, $\lambda \in [0,1]$.

$$\Leftrightarrow \begin{bmatrix} \lambda x_1 + (1-\lambda) x_2 \\ \lambda t_1 + (1-\lambda) t_2 \end{bmatrix} \in \text{epi}(f)$$

$$\Leftrightarrow f(\lambda x_1 + (1-\lambda) x_2) \le \lambda t_1 + (1-\lambda) t_2 . \text{(to show)}.$$

$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) \le \lambda t_1 + (1-\lambda) t_2$$ which is what we wanted to show.

$(\Leftarrow)$ suppose epi f is a convex set.

$(x_1, f(x_1))$, $(x_2, f(x_2))$

then for $\lambda \in [0,1]$, $\begin{bmatrix} \lambda x_1 + (1-\lambda) x_2 \\ \lambda f(x_1) + (1-\lambda) f(x_2) \end{bmatrix} \in \text{epi} f$.

$$\Rightarrow f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$