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

- Class Hours: D Slot. Monday: 10am - 10:55pm, Wednesday: 8am - 9:55am, Thursday: 10am - 10:55am
- Venue: NR 313
- Grading Scheme: 50 % Endsem, 30 % Midsem, 20 % Tutorial, 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.

Weekly Plan

Week 1: 4th - 11th January

- Formal Definition of an Optimization Problem
- Constraints, Feasible solutions, Optimal solution, Optimal value
- Infeasible and unbounded optimization problems
- Local vs. global optimal solutions
- Compact Sets, Continuous Functions, Weierstrass Theorem on existence of global optima
- Gradient, Hessian, Optimality conditions for unconstrained problems

Week 2: 15th - 18th January

- Convex Sets
- Examples
- Operations that preserve convexity of sets
- Convex functions: Definition, Level set Characterization, First order characterization

Weekly Plan

Week 3: 22nd - 25th January

- Convex functions: Second order characterization
- Operations that preserve convexity of functions
- Examples
- Formulate and solve simple convex optimization problems (such as constrained least squares problem) using suitable solvers

Week 4: 29th January - 1st February

- Examples of Convex Optimization Problem Classes
- Equivalent Forms
- Separating Hyperplane Theorems, Theorems of the Alternative, LP Duality

Weekly Plan

Week 5 & 6: 5th - 14th February

- Convex Theorems of the Alternative, Constraint Qualification
- Lagrangian Duality: weak and strong versions
- Saddle Point Formulations
- KKT Optimality Conditions
- Examples
- Properties of Convex Optimization Problems: Global Optimality, Strong Duality, Necessary Conditions being Sufficient
- Regression Problems and applications
- Practice Problems

Mid-semester Examination

Week 7: 26th - 29th February

- Classification via Support Vector Machines
- ML Estimation
- Hypothesis Testing and Optimal Detection

Weekly Plan

Week 8 & 9: 4th - 14th March

- First order algorithms, Accelerated Methods
- Stochastic Gradient Descent
- Distributed Optimization

Week 10 & 11: 18th - 28th March

- Linear Matrix Inequality
- Conic Duality
- Semidefinite Programming
- Applications of SDP in Control: Stability, State Feedback Synthesis, Robust Synthesis

Week 12 & 13: 1st - 11th April

- Constrained Optimal Control, Model Predictive Control
- Applications in System Identification
- Robust Optimization via Duality

References

Primary Reference:

- 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.

References

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

- 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.

Optimization in Abstract Form

$|X|$ is finite.

An optimization problem can be stated as

$$\boxed{\min_{x \in X} f(x),} \qquad \min \; ax^2 \tag{1}$$

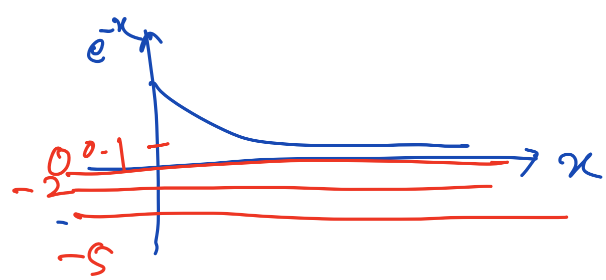
where

- 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

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 $\inf_{x \in X} f(x)$? : Greatest lower bound.



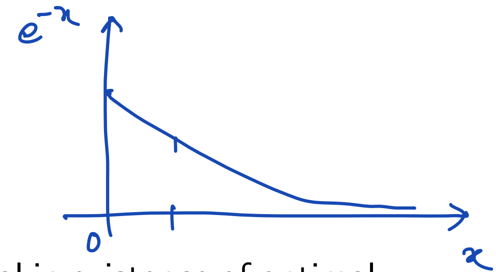
Examples

$\{x \in \mathbb{R} \mid x \geq 0,\ x < \infty\}$

- Let $f(x) = e^{-x}$ and $X = [0, \infty)$. Find f^* and x^*. $f^* = 0$, x^* is not defined

$\min_{x \in X} e^{-x} = ,\ x \in \mathbb{R}.$

- What if $X = [0, 1]$? minimum possible value e^{-1} and $x^* = 1$.



- What if $X = [0, 1)$? $\{x \in \mathbb{R} \mid x \leq 1,\ x \geq 0\}$

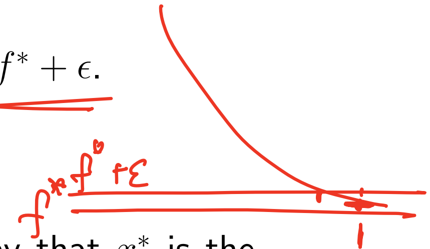
Moral of the story: Properties of feasibility set X is critical in existence of optimal solution.

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 $x \in X$.

$f^* <$

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

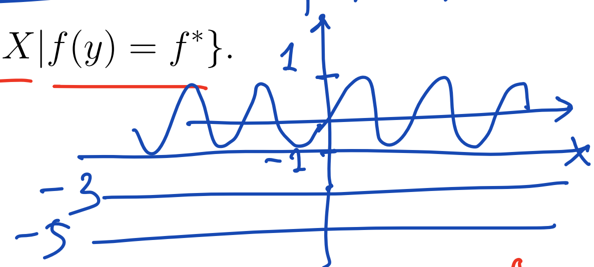


There are two possibilities:

- 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

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

$f(x) = \sin x$



$\subseteq X$

$\min_{x \in [0,1]} \quad 0$

$f(x) = 0$

$X = [0, 10\pi]$

$\inf_{x \in X} f(x) = \sin x$

$-1+\varepsilon \quad = -1$

$\max_{x \in X} f(x) \equiv \min_{x \in X} -f(x)$

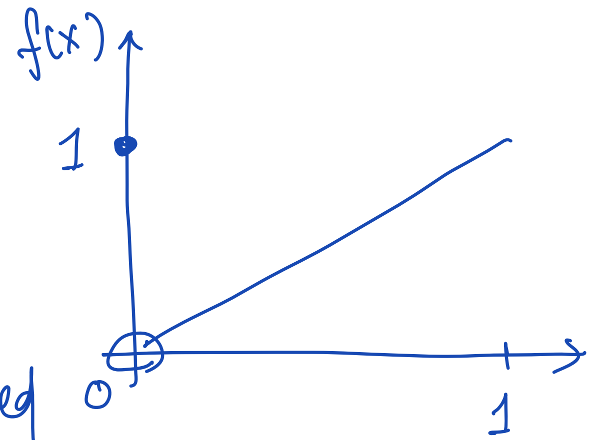
$g(x) = -f(x)$

$\operatorname{argmin}_{x \in X} f(x) = \left\{\frac{3\pi}{2}, \frac{7\pi}{2}, \ldots\right\}$

Example

$\min_{x \in X} f(x)$, $f(x) = \begin{cases} 1 , & \text{when } x = 0 \\ x , & \text{when } x > 0 \end{cases}$

$X = [0, 1]$



Moral of the story:

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

optimal solution is not defined

optimal value is not attained

$\operatorname{argmin}_{x \in X} f(x) = \emptyset$

Infeasible optimization problem

- The problem is infeasible when X is an empty set.
- In this case, $f^* := +\infty$.
- Example:

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

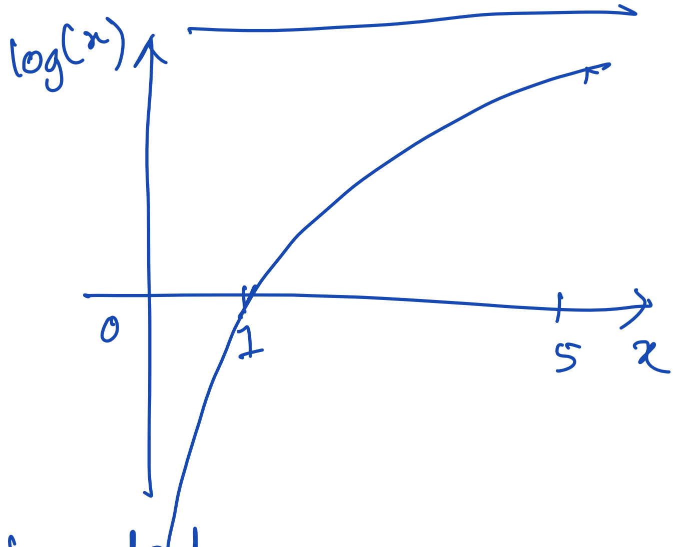
$$X_1 = \{x_1 \geq 0\}$$
$$X_2 = \{x_2 \geq 1\}$$
$$X_3 = \{x_1 + x_2 = -1\}$$

$$X = X_1 \cap X_2 \cap X_3 = \emptyset$$

Unbounded optimization problem

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

- Example:

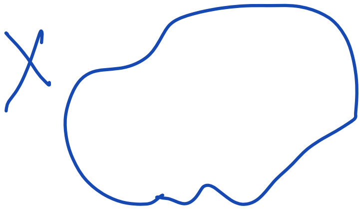


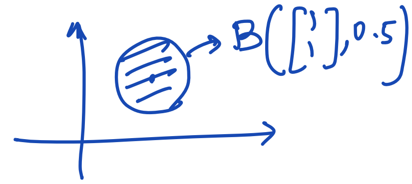
unbounded $\begin{cases} f(x) = \log x \\ X = [0, 5] \end{cases}$

$\begin{cases} f(x) = \log x \\ X_1 = [1, 5] \\ \min\limits_{x \in X_1} f(x) \text{ is not unbounded.} \end{cases}$

$f^* = 0$, and optimal solution $x^* = 1$

$X_2 = (0, 5]$: also unbounded





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, denoted $\mathtt{int}(X)$ $= \{x_0 \in X \mid \exists$ a radius r for which $B(x_0, r) \subset X\}$
- Set X is called an **open** set if $X = \mathtt{int}(X)$.
- 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':

Example of a set which is neither open nor closed:

$X = (0, 1]$.

$\mathrm{int}(X) = (0, 1)$

X is not open

$X^c = (-\infty, 0] \cup (1, \infty)$

$\mathrm{int}(X^c) = (-\infty, 0) \cup (1, \infty)$

$X^c \neq \mathrm{int}(X^c)$

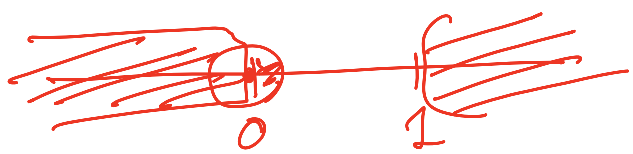
$\Rightarrow X^c$ is not open $\Rightarrow X$ is not closed.

$X = \{1\}$ $\Rightarrow$ X is not open.

$\mathrm{int}(X) = \emptyset$

$X^c = (-\infty, 1) \cup (1, \infty)$ is open

$\Rightarrow X$ is closed.



Bounded and Compact Set

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

  Is $[0, 10]$ bounded? yes

  is $\mathbb{R}^2$ bounded? No.

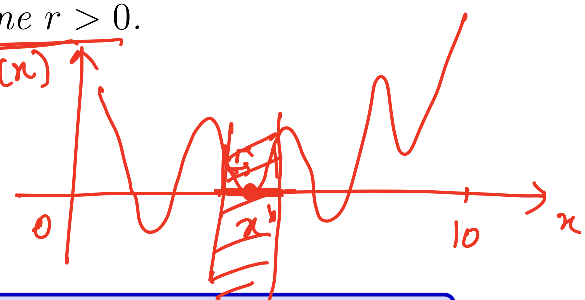
- A set X is compact if it is closed and bounded.

Global and Local Optimum

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

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

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



Existence of Optimal Solution:

Theorem 1: Weierstrass Theorem

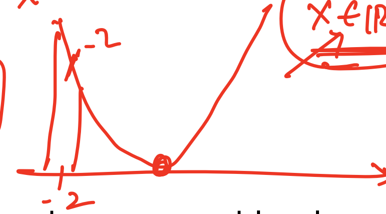
If the cost function f is continuous and the feasible region X is compact (closed and bounded), then (at least one global) optimal solution x^* exists.

$f(x) = x^2$

$\min\limits_{x\in\mathbb{R}} x^2$

Example:

$B(-2, \varepsilon) = [-2-\varepsilon, -2+\varepsilon]$



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

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

(for at least one $\alpha \in \mathbb{R}$)

is non-empty and bounded.

If $f(x) = x^2$, $\alpha = 36$,

$S_\alpha(f) = \{x \in \mathbb{R} | f(x) \le 36\}$

$= [-6, 6]$

The story so far

Given an optimization problem, $\min_{x \in X} f(x)$

first determine

i) decision variable $x \in$?

ii) feasibility set X

iii) cost function $f: X \to \mathbb{R}$

To check whether a globally optimal solution exists, check whether

i) f is continuous

ii) X is closed

iii) X is bounded or any sub-level set of X is bounded.

convince yourself that $\min_{x \in X} f(x) \equiv \boxed{\min_{x \in S_\alpha(f)} f(x)}$ for any $\alpha \in \mathbb{R}$ at which $S_\alpha(f) \neq \phi$

Today's lecture:

How do we verify if $x^* \in X$ is indeed an optimal solution?

Derivative of a function $f: \mathbb{R}^n \to \mathbb{R}^m$, denoted $Df(x)$ is such that $f(x+\Delta x) \simeq f(x) + Df(x) \cdot \Delta x$

$\Rightarrow Df(x) \in \mathbb{R}^{m \times n}$

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

$$\underbrace{Dh(x_0)}_{\in \mathbb{R}^{p \times n}} = \underbrace{Dg(y_0)}_{\mathbb{R}^{p \times m}} \underbrace{Df(x_0)}_{\mathbb{R}^{m \times n}} \quad \text{where } y_0 = f(x_0)$$

Gradient ($\nabla f(x)$)

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

$$\nabla f(y) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_y \\ \left.\frac{\partial f}{\partial x_2}\right|_y \\ \vdots \\ \left.\frac{\partial f}{\partial x_n}\right|_y \end{bmatrix} = \left(Df(y)\right)^T$$

Compute gradient of

- $f(x) = x^\top a$, $\nabla f(x) = a$
- $f(x) = x^\top A x$
- $f(x) = ||Ax - b||_2^2$ $\Rightarrow$ $\boxed{\nabla f(x) = 2A^TAx - 2A^Tb.}$

$$f(x) = [x_1 \; x_2 \; \cdots \; x_n] \begin{bmatrix} \sum_{i=1}^n a_{1i} x_i \\ \sum_{i=1}^n a_{2i} x_i \\ \vdots \\ \sum_{i=1}^n a_{ni} x_i \end{bmatrix} = \sum_{j=1}^n \sum_{i=1}^n x_i a_{ij} x_j$$

$$= \sum_{j=1}^n \left[a_{jj} x_j^2 + \sum_{i \neq j} x_i a_{ij} x_j \right]$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\left[a_{kk} x_k^2 + \underbrace{\sum_{i \neq k} x_i a_{ik} x_k} + \sum_{j \neq k} \left[a_{jj} x_j^2 + \sum_{i \neq j} x_i a_{ij} x_j \right] \right]$$

$$= 2a_{kk} x_k + \sum_{i \neq k} x_i a_{ik} + \sum_{j \neq k} a_{kj} x_j$$

$$= \sum_{i=1}^n x_i a_{ik} + \sum_{j=1}^n a_{kj} x_j = [Ax]_k + [A^T x]_k$$

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

Hessian ($H(x)$)

$\nabla f(x) : \mathbb{R}^n \to \mathbb{R}^n$

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

$$\begin{bmatrix} D\frac{\partial f}{\partial x_1} \\ D\frac{\partial f}{\partial x_2} \\ \vdots \\ D\frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \\ \vdots & & & \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Compute Hessian of

- $f(x) = x^\top a$ $\Rightarrow Hf(x) = Da = 0_{n\times n}$
- $f(x) = x^\top Ax$
- $f(x) = ||Ax - b||_2^2$

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

$H(x) = D[(A + A^T)x]$

$= A + A^T$

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

$\nabla f(x) = 2A^TAx - 2A^Tb$

$H(x) = 2A^TA$

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

$$\lim_{\varepsilon \to 0} \frac{f(x_0 + \varepsilon d) - f(x_0)}{\varepsilon} = \underline{\phi'(0)}$$

'dash' will only be used for derivatives when the function is from $\mathbb{R}$ to $\mathbb{R}$.

Define $\underline{\phi(t) := f(x + td).}$, $t \in \mathbb{R}$,

Compute $\phi'(0)$:

$$\frac{d\phi}{dt} = \phi'(t) = Df(y) \frac{d}{dt}(x + td) \quad , \quad y = x + td$$

$$= \nabla f(x+td)^T \underline{\underline{\frac{d}{dt}(x + td)}}$$

$$= \nabla f(x+td)^T d$$

$$\underline{\underline{\phi'(0) = \nabla f(x)^T d}}$$

$g(t) = x + td$

$g : \mathbb{R} \to \mathbb{R}^n$

$Dg(t) \in \mathbb{R}^{n \times 1}$

$= d$

feasibility space

$\Rightarrow \quad \mathcal{X} = \mathbb{R}^n.$

Necessary Condition of Optimality for Unconstrained Problems

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 x^* is a local optimum, yet $\nabla f(x^*) \neq 0$.

Recall: directional derivative along d: $\underline{\nabla f(x^*)^T d}$

If $d = -\nabla f(x^*)$, then directional derivative

$$-\nabla f(x^*)^T \nabla f(x^*) = -\|\nabla f(x^*)\|_2^2$$

$$\underline{f(x^* + td)} \simeq f(x^*) + t\nabla f(x^*)^T d + (\text{HOT})$$

$$= f(x^*) - t\|\nabla f(x^*)\|_2^2 + (\text{HOT})$$

There always exists $\bar{t}$ sufficiently small such that

$$-\bar{t}\|\nabla f(x^*)\|_2^2 + (\text{HOT}) < 0$$

$\Rightarrow f(x^* + \bar{t}d) < f(x^*)$

$\Rightarrow x^*$ is not a local optimum $\Rightarrow$ contradiction!

Sufficient Condition of Optimality for Unconstrained Problems

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 $\min_{x \in \mathbb{R}^n} f(x)$.

Recall: Taylor series expansion

$$f(x) = f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0) + \cdots$$

If at x^*, $\nabla f(x^*) = 0$, then

$$f(x^* + \varepsilon d) = f(x^*) + \underbrace{(\varepsilon d)^T H(x^*)\, \varepsilon d}_{+\text{ve } \forall d.} + \text{HOT}$$

Therefore $f(x^* + \varepsilon d) > f(x^*)$.

Least-squares problem / Linear Regression

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

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

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

$$\underline{H(x) = 2A^T A.}$$

If x^* is an optimal solution, then $\nabla f(x^*) = 0$

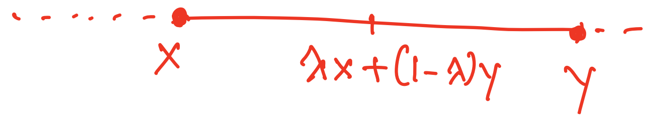
$$\Rightarrow \boxed{A^T A x^* = A^T b}$$

If $H(x)$ is <u>positive definite</u>, $x^* = (A^T A)^{-1} A^T b$ is the unique solution which is a local optimum.

<u>Note</u>: $A^T A$ is always positive semidefinite, but may not always be positive definite.

Convex Sets

Definition 3. *Given a collection of points $x_1, x_2, \dots, x_k$, the combination $\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k$ is called* **Convex** *if $\lambda_i \geq 0$ and $\sum_{i=1}^{k} \lambda_i = 1$. A set X is convex if all convex combinations of its elements are in the set.*

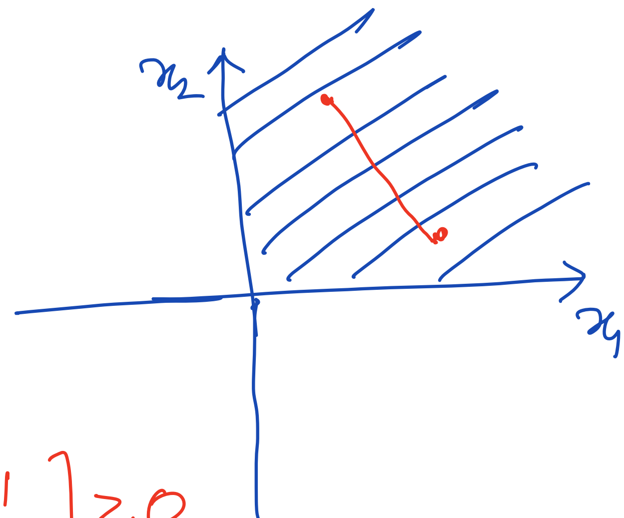


Equivalently, X is a convex set if

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

Examples :

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

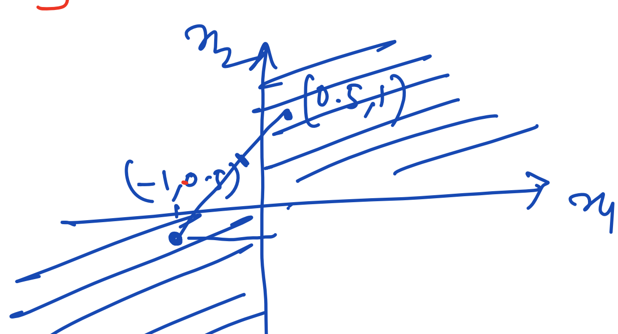


$y, z \in X \Rightarrow y_1 \geq 0, \ y_2 \geq 0$
$z_1 \geq 0, \ z_2 \geq 0$

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

$p = \lambda y + (1-\lambda) z = \begin{bmatrix} \lambda y_1 + (1-\lambda) z_1 \\ \lambda y_2 + (1-\lambda) z_2 \end{bmatrix} \geq 0 \Rightarrow p \in X$

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



$y_1 = \begin{bmatrix} -1 \\ -0.5 \end{bmatrix}, \ z_1 = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$

$\lambda = 1/2$, $p = \lambda y_1 + (1-\lambda) z_1$

$= \dfrac{y_1 + z_1}{2} = \begin{bmatrix} -0.25 \\ +0.25 \end{bmatrix} \notin X_2$

$\Rightarrow X_2$ is not 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 $x_1, x_2 \in H \Rightarrow a^T x_1 = b$, $a^T x_2 = b$.

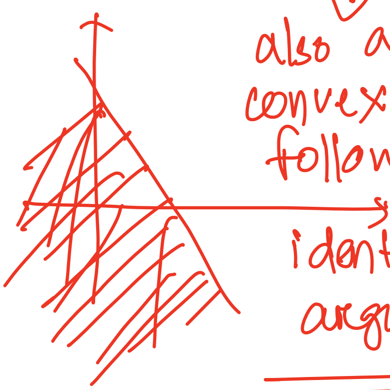
let $\lambda \in [0,1)$

$p = \lambda x_1 + (1-\lambda) x_2$

we need to check whether $a^T p = b$.

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

$\Downarrow$ also a convex set following identical arguments.



$$a^T p = a^T(\lambda x_1 + (1-\lambda) x_2)$$

$$= \lambda a^T x_1 + (1-\lambda) a^T x_2 = \lambda b + (1-\lambda) b$$

$$= b$$

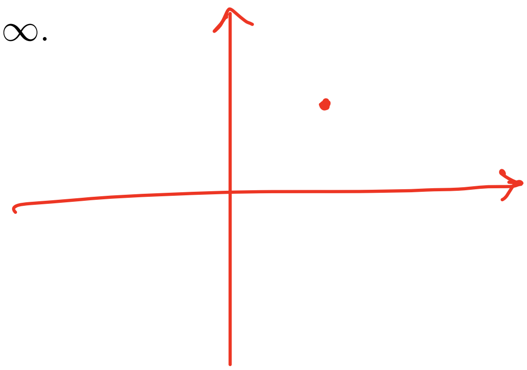
$\Rightarrow p \in H \Rightarrow$ H is a convex set.

Sets Defined by Norms

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

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

We define $[n] := \{1, 2, \ldots, n\}$.



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

Let $x_1, x_2 \in B_p(c,R)$

$\Rightarrow ||x_1 - c||_2 \leq R, \quad ||x_2 - c||_2 \leq R$

Let $\lambda \in [0,1]$, $z := \lambda x_1 + (1-\lambda)x_2$.

Let us check whether z belongs to $B_p(c,R)$.

$$\begin{aligned} ||z - c||_p &= ||\lambda x_1 + (1-\lambda)x_2 - c||_p \\ &= ||\lambda x_1 + (1-\lambda)x_2 - \lambda c - (1-\lambda)c||_p \\ &= ||\lambda(x_1 - c) + (1-\lambda)(x_2 - c)||_p \\ &\leq ||\lambda(x_1 - c)||_p + ||(1-\lambda)(x_2 - c)||_p \quad \text{(triangle inequality)} \\ &= \lambda ||x_1 - c||_p + (1-\lambda)||x_2 - c||_p \quad \text{(the homogeneity)} \\ &\leq \lambda R + (1-\lambda)R = R. \quad (\because x_1, x_2 \in B_p(c,R)) \end{aligned}$$

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

Positive Semidefinite Matrices

$A \succeq B \Rightarrow (A-B) \succeq 0$

$\Downarrow$

$A-B$ is +ve semidefinite

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.*

Let $X_1, X_2 \in S_n^+ \Rightarrow X_1 \succeq 0$

$X_2 \succeq 0$

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

$Z = \lambda X_1 + (1-\lambda) X_2$

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

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

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

$\Rightarrow Z$ is positive semidefinite.

$\Rightarrow S_n^+$ is a convex set.