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

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

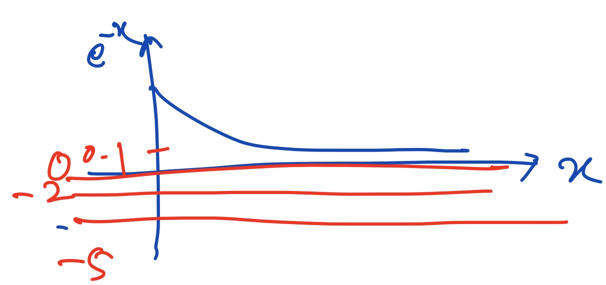
where

- 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.



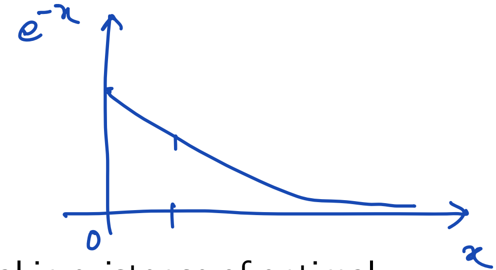
Examples

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

- Let $f(x) = e^{-x}$ and $X = [0, \infty)$. Find f^* and x^*. $= 0$ 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 $X = [0, 1)$? $\{x \in \mathbb{R} \mid x \leq 1, x \geq 0\}$



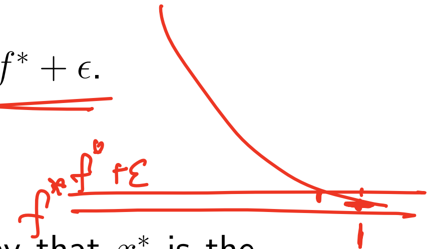
Moral of the story: Properties of feasibility set X is critical in existence of optimal solution.

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 $f(\bar{x}) < f^* + \epsilon$.

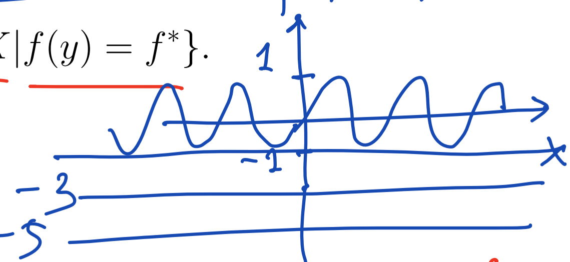


There are two possibilities:

- 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

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

$f(x) = \sin x$



$\subseteq X$

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

$f(x) = 0$

$X = [0, 10\pi]$

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

$-1 + \epsilon$ $= -1$

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

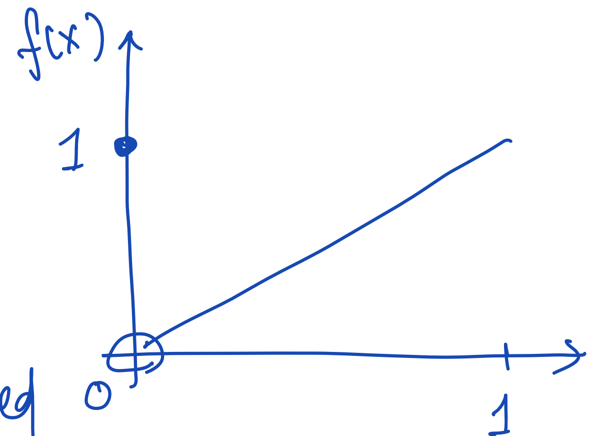
$g(x) = -f(x)$

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

Example

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

$X = [0, 1]$



Moral of the story:

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

optimal solution is not defined

optimal value is not attained

$$\underset{x \in X}{\operatorname{argmin}} f(x) = \phi$$

Infeasible optimization problem

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

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

$$X_1 = \{x_1 \geq 0\}$$

$$X_2 = \{x_2 \geq 1\}$$

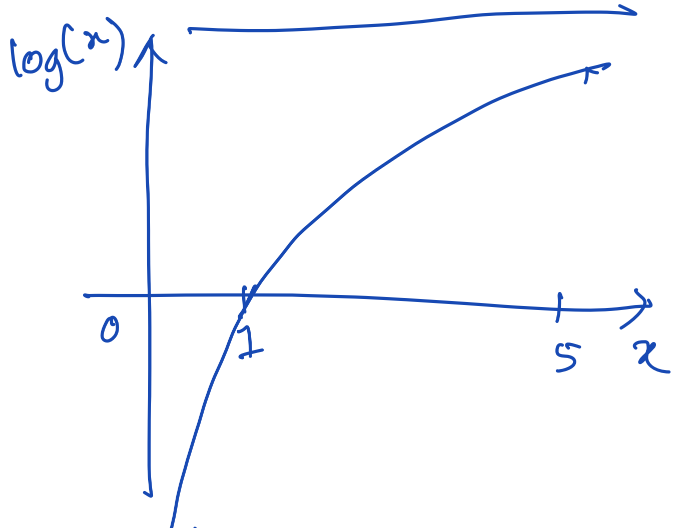
$$X_3 = \{x_1 + x_2 = -1\}$$

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

Unbounded optimization problem

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

- Example:

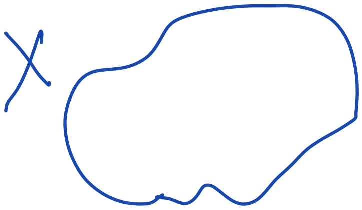


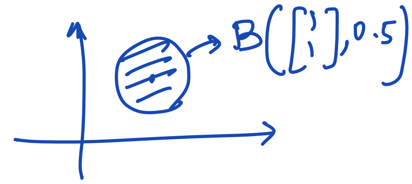
unbounded $\begin{cases} f(x) = \log x \\ X = [0, 5] \end{cases}$

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

$f^* = 0$, 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 $\texttt{int}(X)$ $= \{x_0 \in X \mid \exists$ a radius r for which $B(x_0, r) \subset X\}$
- Set X is called an **open** set if $X = \texttt{int}(X)$.
- 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:

$X = (0, 1]$.

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

X is not open

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

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

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

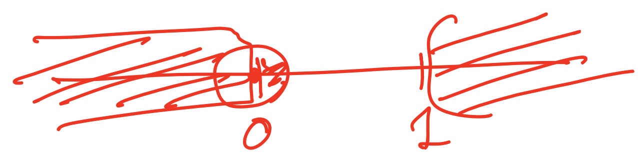
$\Rightarrow X^c$ is not open $\Rightarrow X$ is not closed.

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

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

$X^c = (-\infty, 1) \cup (1, \infty)$ 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 $x_1, x_2 \in X$, $||x_1 - x_2||_2 \leq B$.

  Is $[0, 10]$ bounded? yes

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

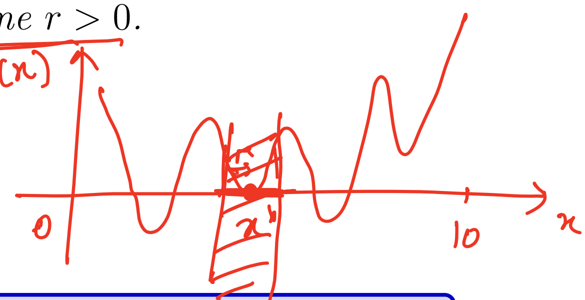
- A set X is compact if it is closed and bounded.

Global and Local Optimum

Definition 1 (Global Optimum). *A feasible solution $x^* \in X$ is a global optimum if $f(x^*) \leq f(x)$ 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 $f(x^*) \leq f(x)$ for all $x \in B(x^*, r)$ for some $r > 0$.*



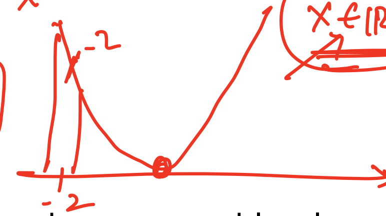
Existence of Optimal Solution:

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.

$f(x) = x^2$ $\min_{x \in \mathbb{R}} x^2$

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

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

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

is non-empty and bounded.

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

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

$= [-6, 6]$

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

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

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

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

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

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

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

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

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

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

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

Hessian ($H(x)$)

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

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

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

Compute Hessian of

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

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

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

$= A + A^T$

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

$H(x) = 2A^TA$

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

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

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

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

Compute $\phi'(0)$:

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

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

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

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

$g(t) = x + td$

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

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

$= d$

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: $\nabla f(x^*)^T d$

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

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

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

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

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

<u>Recall</u>: 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

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

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

Least - squares problem / Linear Regression

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

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

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

$$H(x) = 2A^TA.$$

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

$$\Rightarrow \boxed{A^TAx^* = A^Tb}$$

If $H(x)$ is positive definite, $x^* = (A^TA)^{-1}A^Tb$ is the unique solution which is a local optimum.

Note: A^TA is always positive semidefinite, but may not always be positive definite.

Convex Sets

Definition 3. *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 $\sum_{i=1}^{n} \lambda_i = 1$. 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** $x, y \in X$, $\lambda x + (1-\lambda)y \in X$ for any $\lambda \in [0, 1]$.
- it contains **all** convex combinations of any two of its elements.

Basic Examples of Convex Sets

Sets Defined by Linear Inequalities:

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

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

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.*

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.*

Operations that preserve convexity of sets

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

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

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.*

Ellipsoid:

Proposition 5. *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.*

Operations that preserve convexity of sets

Proposition 6 (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.

Example:

Proposition 7 (Weighted Sum)**.** *If $X_1, X_2, \ldots, X_m$ are convex sets, then $\sum_{i \in [m]} \alpha_i X_i := \{y \mid y = \sum_{i \in [m]} \alpha_i x_i, \quad x_i \in X_i\}$ is a convex set.*

Example:

Operations that preserve convexity of sets

Proposition 8 (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.*

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.,*

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

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:

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 $\lambda_i \in \mathbb{R}$.

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

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

Projection

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

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

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

Separating Hyperplane

Definition 8 (Separating Hyperplane). *Let X_1 and X_2 be two nonempty convex 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*

- $X_1 \subset H^- := \{x \in \mathbb{R}^n \mid a^\top x \leq b\}$,
- $X_2 \subset H^+ := \{x \in \mathbb{R}^n \mid a^\top x \geq b\}$,
- $X_1 \cap X_2 \not\subset H$.

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 $b' < b''$.*

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.

Separating Hyperplane Theorem

Theorem 5: Separating Hyperplane Theorem

Let X be a closed convex set and $x_0 \notin X$. Then, there exists a hyperplane that strictly separates x_0 and X.

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 $b = a^\top x_0 - \frac{||a||_2^2}{2}$.
2. Use properties of projection and convexity of X to verify that H is indeed the separating hyperplane.

Theorem of the Alternative (Farkas' Lemma)

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

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

Application: Linear Programming Duality

Consider the following pair of linear optimization problems.

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0. \end{aligned} \tag{P}$$

$$\begin{aligned} \max_{y\in\mathbb{R}^m} \quad & b^\top y \\ \text{s.t.} \quad & A^\top y \leq c, \end{aligned} \tag{D}$$

Theorem 6: LP Duality

If (P) has a finite optimal value, then (D) also has a finite optimal value and both optimal values are equal to each other.

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 $\texttt{dom}(f)$, is the set $\{x \in \mathbb{R}^n \mid |f(x)| < +\infty\}$.

- Example: $f(x) = \frac{1}{x}$. What is $\texttt{dom}(f)$?

- $f(x) = \sum_{i=1}^{n} x_i \log(x_i)$. What is $\texttt{dom}(f)$?

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

Convex Functions

Definition 9 (Convex Function)**.** *A function* $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *is convex if*

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

- $f(x) = x^2$
- $f(x) = e^x$
- $f(x) = a^\top x + b$ for $x \in \mathbb{R}^n$

Example: Norms

Definition 10 (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* $\alpha \in \mathbb{R}$,
- $\pi(x + y) \leq \pi(x) + \pi(y)$.

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.*

Example: Indicator Function

Definition 11. *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 if the set C is a convex set.*

Example: Support Function

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.*

Special Types of Convex Functions

Definition 12. *A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is*

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

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 $\sum_{i=1}^k \lambda_i = 1$.*

Proof is straightforward via induction.

Epigraph Characterization

Definition 13. *A epigraph of a function* $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ *is defined as the set*

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

Example: Norm cone: $\{(x,t)|||x|| \leq t\}$ is a convex set.

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* $\mathbb{R}^{n+1}$*.*

Level-set Characterization

Definition 14. *For any $\alpha \in \mathbb{R}$, the level set of function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ at level α is defined as*

$$\mathtt{lev}_\alpha(f) := \{x \in \mathtt{dom}(f) | f(x) \leq \alpha\}.$$

Proposition 15. *If a function f is a convex function, then* **every** *level set of f is a convex set.*

Converse is not true. A function is called quasi-convex if its domain and all level sets are convex sets.

HW: Give an example of a function which is quasi-convex but not convex.

Restriction of a Convex Function on a Line

Proposition 16. *If a function f is convex if and only if for any $x, h \in \mathbb{R}^n$, the function $\phi(t) = f(x + th)$ is a convex function on $\mathbb{R}$.*

If we know how to check convexity of functions defined on $\mathbb{R}$, then we can check convexity of functions defined on $\mathbb{R}^n$.

First Order Condition

Proposition 17. *If a function f is differentiable, then it is convex if and only if* $\mathtt{dom}(f)$ *is a convex set and for any* $x, y \in \mathtt{dom}(f)$*, we have*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

A global lower bound on the function can be obtained at any point based on local information $(f(x), \nabla f(x))$.

Second Order Condition

Proposition 18. *If a function f is twice differentiable, then it is convex if and only if $\mathtt{dom}(f)$ is a convex set and $\nabla^2 f(y) \succeq 0$ for every $y \in \mathtt{dom}(f)$.*

Convexity Preserving Operations

Proposition 19 (Conic Combination)**.** *Let* $\{f_i(x)\}_{i\in I}$ *be a collection of convex functions and let* $\alpha_i \geq 0$ *for all* $i \in I$*. Then,* $g(x) := \sum_{i\in I} \alpha_i f_i(x)$ *is a convex function.*

Proposition 20 (Affine Composition)**.** *If* $f : \mathbb{R}^m \to \mathbb{R}$ *is a convex function, then* $g(x) := f(Ax + b)$ *is also a convex function where* $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$*.*

Convexity Preserving Operations

Proposition 21 (Pointwise Maximum). *Let $\{f_i(x)\}_{i\in I}$ is a collection of convex functions, then $g(x) := \max_{i\in I} f_i(x)$ is a convex function.*

The set I need not be a finite set.

Proposition 22 (Pointwise Supremum). *Let $f(x, \omega)$ is convex in x for any $\omega \in \Omega$, then $g(x) := \sup_{\omega\in\Omega} f(x, \omega)$ is convex in x.*

Convexity Preserving Operations

Proposition 23 (Scalar Composition). *If a function f is convex in $\mathbb{R}^n$, and F is a convex non-decreasing function on $\mathbb{R}$, then $g(x) := F(f(x))$ is convex.*

Proposition 24 (Vector Composition). *Let $\{f_i\}_{i\in\{1,2,\ldots m\}}$ are convex functions on $\mathbb{R}^n$, and $F : \mathbb{R}^m \to \mathbb{R}$ is a convex function and non-decreasing in each argument, then the function $g(x) = F(f(x))$ is convex.*

Convexity Preserving Operations - 6

Proposition 25 (Partial Minimization)**.** *If $f(x, y)$ is convex in (x, y), and Y is a convex set, then $g(x) := \inf_{y \in Y} f(x, y)$ is a convex function.*

Examples of Convex Functions

Recall: Optimization Problem

An optimization problem can be stated as

$$\min_{x \in X} f(x), \tag{2}$$

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 $X := \big\{x \in \mathbb{R}^n | g_i(x) \leq 0, h_j(x) = 0, i \in \{1, 2, \ldots, m\}, j \in \{1, 2, \ldots, p\}\big\}$.
- $f : \mathbb{R}^n \to \mathbb{R}$ 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 $f(x^*) = f^*$.

Recall

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

- The problem is unbounded when $f^* = -\infty$.

Definition 15. *Recall that*

- *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:* $\operatorname{argmin}_{x\in X} f(x) := \{z \in X | f(z) = f^*\}$*,*
- *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* $r > 0$*.*

Theorem: 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.

Abstract vs. Standard Form

An optimization problem can be stated in abstract form as

$$\min_{x \in X} f(x), \tag{3}$$

where $X := \big\{x \in \mathbb{R}^n | g_i(x) \leq 0, h_j(x) = 0, i \in \{1, 2, \ldots, m\}, j \in \{1, 2, \ldots, p\}\big\}$, or in "standard form" as

$$\begin{array}{rll} \displaystyle\min_{x \in \mathbb{R}^n} & f(x) & \\ \text{subject to} & g_i(x) \leq 0, & i \in \{1, 2, \ldots, m\} \\ & h_j(x) = 0, & j \in \{1, 2, \ldots, p\}. \end{array}$$

Feasibility Problem

Goal: Find $x \in \mathbb{R}^n$ which satisfies a collection of inequality and equality constraints.

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & 0 \\ \text{subject to} \quad & g_i(x) \leq 0, \qquad i \in \{1, 2, \ldots, m\} \\ & h_j(x) = 0, \qquad j \in \{1, 2, \ldots, p\}. \end{aligned}$$

$f^* = 0$ if a feasible solution exists. Otherwise, $f^* = +\infty$.

Equivalent Optimization Problems

Consider the following two optimization problems:

$$\min_{x \in X} f(x). \tag{4}$$

$$\min_{y \in Y} g(y). \tag{5}$$

The above problems are equivalent if

- Given an optimal solution x^* of (4), we can find an optimal solution y^* of (5), and
- given an optimal solution y^* of (5), we can find an optimal solution x^* of (4).

Equivalence: Maximization

Equivalence: Epigraph Form

Equivalence: Slack Variables

Equivalence: From Equality to Inequality Constraints

Equivalence: From Constrained to Unconstrained

Equivalence: Scalar Multipliers and Constant Terms

Equivalence: Monotone Transformations

Inner Approximation

Relaxation and Soft Constraints

Convex Optimization Problems

An optimization problem in abstract form

$$\min_{x \in X} f(x), \tag{6}$$

is convex when the feasibility set X is a convex set and the cost function $f(x)$ is a convex function.

An optimization problem in standard form

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \qquad i \in \{1, 2, \ldots, m\} \\ & h_j(x) = 0, \qquad j \in \{1, 2, \ldots, p\}, \end{aligned}$$

is convex when

- f and g_i are convex functions.
- h_j **are affine functions.**

1. Local Optimum is Global

2. Necessary and Sufficient Optimality Condition

3. Set of Minimizers is a Convex Set

Linear Programming

Quadratic Programming

QCQP

SOCP

LMIs

Linear Programming (LP)

LP is a class of optimization problems where the cost function is linear in the decision variable and the feasibility set is a polyhedron.

LP in standard equality form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & c^\top x \\ \text{s.t.} & Ax = b, \\ & x \geq 0. \end{aligned} \tag{P}$$

Obtaining a lower bound on the cost function

Finding best possible lower bound

This happens to be another linear program:

$$\max_{y \in \mathbb{R}^m} b^\top y$$
$$\text{s.t.} A^\top y \leq c. \tag{D}$$

The above problem is referred to as the dual of problem (P).
A LP stated as above is called standard inequality form.
We can show that the dual of (D) is (P).

Properties

Theorem 7

For the primal-dual pair of optimization problems stated above, the following are true.

1. If (P) is infeasible, and (D) is feasible, then (D) is unbounded.
2. If (P) is unbounded, then (D) is infeasible.
3. **Weak Duality**: For any feasible solution $\bar{x}$ and $\bar{y}$ of the respective problems, we always have $c^\top \bar{x} \geq b^\top \bar{y}$.
4. **Strong Duality**: Show that for the respective optimal solutions x^* and y^*, we must have $c^\top x^* = b^\top y^*$.

HW: Give an example of (P) and (D) where both are infeasible.

Proof

Farkas' Lemma

To prove the strong duality theorem, we will make use of an alternative form of Farka's lemma.

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

Lemma 3 (Alternative form of 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. $\{x \in \mathbb{R}^n | Ax \leq b\}$
2. $\{y \in \mathbb{R}^m | y \geq 0, y^\top A = 0, y^\top b < 0\}$.

Lagrangian Function

Consider the following optimization problem in standard form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i \in [m] := \{1, 2, \ldots, m\}, \\ & h_j(x) = 0, j \in [p]. \end{aligned}$$

The Lagrangian function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p : \mathbb{R}$ is defined as

$$L(x, \lambda, \mu) := f(x) + \sum_{i \in [m]} \lambda_i g_i(x) + \sum_{j \in [p]} \mu_j h_j(x),$$

where

- λ_i is the Lagrange multiplier associated with $g_i(x) \leq 0$
- μ_j is the Lagrange multiplier associated with $h_j(x) = 0$.

Lower Bound Property:

Lemma 4. *If $\bar{x}$ is feasible and $\bar{\lambda} \geq 0$, then $f(\bar{x}) \geq L(\bar{x}, \bar{\lambda}, \mu)$.*

Lagrangian Dual

From the previous lemma, we know that if $\bar{x}$ is feasible and $\bar{\lambda} \geq 0$, then

$$f(\bar{x}) \geq L(\bar{x}, \bar{\lambda}, \mu) \geq \inf_x L(x, \bar{\lambda}, \mu) =: d(\bar{\lambda}, \mu).$$

where

$$d(\lambda, \mu) := \inf_x \Big[f(x) + \sum_{i \in [m]} \lambda_i g_i(x) + \sum_{j \in [p]} \mu_i h_j(x) \Big].$$

- $d(\lambda, \mu)$ requires solving an unconstrained optimization problem.
- Given any $\lambda \geq 0, \mu$, $d(\lambda, \mu) \leq f^*$ where f^* is the optimal value.
- $d(\lambda, \mu)$ may take value $-\infty$ for some choice of λ and μ.
- $d(\lambda, \mu)$ is concave in λ and μ.

Lagrangian Dual Optimization Problem

Let us compute the best lower bound on f^*:

$$\begin{aligned} \max_{\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p} \quad & d(\lambda, \mu) \\ \text{s.t.} \quad & \lambda \geq 0, \\ & (\lambda, \mu) \in \texttt{dom}(d). \end{aligned}$$

- The above is a convex optimization problem since $d(\lambda, \mu)$ is concave in λ and μ.
- Let the optimal value be denoted d^*.

Example 1: Lagrangian Dual of LP

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.} \quad & Ax = b, x \geq 0. \end{aligned}$$

Find L, d and $\texttt{dom}(d)$.

Example 2: Least Norm Solution

Least norm solution:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^\top x \\ \text{s.t.} \quad & Ax = b. \end{aligned}$$

Find L and d.

Weak and Strong Duality

Weak Duality: $d^* \leq f^*$ always holds (even for non-convex problems).

Strong Duality: $d^* = f^*$ is guaranteed to hold for convex problems satisfying certain conditions, referred to as *constraint qualification* conditions.

Example 3

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & -x^2 \\ \text{s.t.} \quad & x - 1 \le 0, \quad -x \le 0. \end{aligned}$$

Find the optimal value of the above problem, derive the dual and determine whether strong duality holds.

Example 4

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & -x_1^2 - x_2^2 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 1 \leq 0. \end{aligned}$$

Find the optimal value of the above problem, derive the dual and determine whether strong duality holds.

KKT Optimality Conditions

For the above primal and dual optimization problems, $\bar{x}, \bar{\lambda}$ and $\bar{\mu}$ are said to satisfy KKT optimality conditions if the following holds:

- **Primal Feasibility**: $g_i(\bar{x}) \leq 0, i \in [m], h_j(\bar{x}) = 0, j \in [p]$.
- **Dual Feasibility**: $\bar{\lambda} \geq 0$.
- **Complementary Slackness**: $\bar{\lambda}_i g_i(\bar{x}) = 0$ for all $i \in [m]$.
- **Stationarity**: $\nabla_x f(\bar{x}) + \sum_{i \in [m]} \bar{\lambda}_i \nabla_x g_i(\bar{x}) + \sum_{j \in [p]} \bar{\mu}_i \nabla_x h_j(\bar{x}) = 0$.

Sufficient Condition for Optimality

Let $\bar{x}, \bar{\lambda}$ and $\bar{\mu}$ satisfy KKT conditions stated above. From primal and dual feasibility we have

$$
\begin{aligned}
d(\bar{\lambda}, \bar{\mu}) &= \inf_x \left[f(x) + \sum_{i \in [m]} \bar{\lambda}_i g_i(x) + \sum_{j \in [p]} \bar{\mu}_i h_j(x) \right] \\
&\leq f(\bar{x}) + \sum_{i \in [m]} \bar{\lambda}_i g_i(\bar{x}) + \sum_{j \in [p]} \bar{\mu}_i h_j(\bar{x}) \leq f(\bar{x}).
\end{aligned}
$$

Further, both inequalities hold with equality.

Thus, when the primal problem is convex, we have:

- $d(\bar{\lambda}, \bar{\mu}) = f(\bar{x})$ (strong duality)
- $\bar{x}$ is optimal solution of primal problem.
- $(\bar{\lambda}, \bar{\mu})$ are optimal solution of dual problem.

Necessary and Sufficient Condition for Optimality

Theorem 8

Suppose the primal optimization problem is convex which satisfies Slater's constraint qualification condition: there exists $\bar{x} \in \texttt{int}(\mathcal{D})$ in the domain of the optimization problem for which $g_i(\bar{x}) < 0$ for all $i \in [m]$ and $h_i(\bar{x}) = 0$ for all $i \in [p]$.

Then, strong duality holds. Equivalently, a feasible solution x^* is optimal if and only if there exist λ^*, μ^* such that (x^*, λ^*, μ^*) satisfy KKT optimality conditions.

Constraint qualification is required for the necessity part of the proof.

Convex Theorem of the Alternative

Consider the following general form of optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \le 0, i \in [m] := \{1, 2, \ldots, m\}, \end{aligned}$$

where f and g_i are convex functions.

Theorem 9

Let the constraint functions g_i satisfy **slater's** condition: there exists $\bar{x}$ such that $g_i(\bar{x}) < 0$ for all $i \in [m]$. Then, exactly one of the following two systems must be empty.

- $\{x \in \mathbb{R}^n | f(x) < 0, g_i(x) \le 0, i \in [m]\}$
- $\{\lambda \in \mathbb{R}^m | \inf_{x \in \mathbb{R}^n} [f(x) + \sum_{i \in [m]} \lambda_i g_i(x)] \ge 0\}$.

Proof: Blackboard.

Strong Duality Theorem

Consider the following general form of optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, i \in [m] := \{1, 2, \ldots, m\}, \end{aligned}$$

where f and g_i are convex functions satisfying Slater's condition.

Theorem 10

x^* is an optimal solution to the above problem if and only if there exists $\lambda^* \geq 0$ such that $\inf_{x \in \mathbb{R}^n}[f(x) + \sum_{i \in [m]} \lambda_i g_i(x)] \geq f(x^*)$.

Proof: Blackboard.

Other notions of constraint qualification

- If all the constraint functions $g_i(x)$ and $h_j(x)$ are affine, then constraint qualification holds.

- **Relaxed Slater Condition:** If some of the inequality constraints are affine, then they need not hold with strict inequality. It is sufficient to find $\bar{x} \in \texttt{relint}(\mathcal{D})$ such that $g_i(\bar{x}) < 0$ for all g_i that are not affine.

- **Linear Independence Constraint Qualification** holds at a feasible solution x^* if the vectors

$$\begin{aligned} &\nabla h_j(x^*), \quad j \in [p], \\ &\nabla g_i(x^*), \quad i \in \{k \in [m] | g_k(x^*) = 0\} \end{aligned}$$

  are linearly independent.