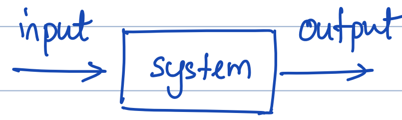
5th Jan 2026: Introduction

Most engineered systems are systems with memory, with input and output governed by an ordinary differential equation (ODE).

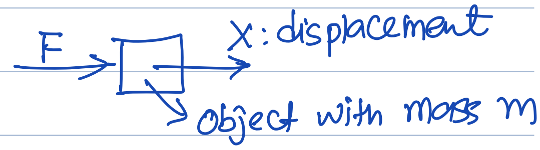


$\underline{y = f(x)}$: memoryless system.

x_1 : position
x_2 : velocity

$$\ddot{x}(t) = \frac{d^2 x(t)}{dt^2} = \frac{F(t)}{m}$$

$$\frac{dv(t)}{dt} = \frac{F(t)}{m}$$



State-space representation: Let $\underline{x \in \mathbb{R}^n}$ denote the states of the system.

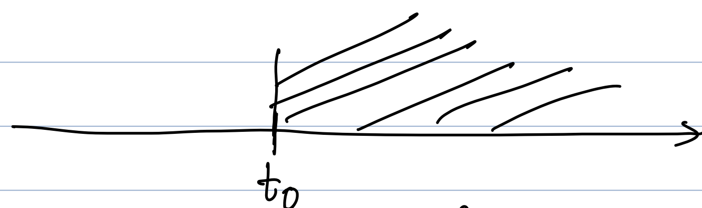
$$\dot{x}(t) = f(x(t), u(t), t) \quad \| \quad \underline{y(t) = g(x(t), u(t), t)}$$

$= x(t)$

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = \frac{F(t)}{m},$$
$$u(t) = F(t)$$

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} f_1(x(t), u(t), t) \\ f_2(x(t), u(t), t) \\ \vdots \\ f_n(x(t), u(t), t) \end{bmatrix}$$

$$f_1(\cdot) = x_2$$
$$f_2(\cdot) = \frac{u(t)}{m}$$



State variables summarize all necessary information so that we can predict future output.

$$\dot{x} = Ax + Bu, \quad \underline{B = -Kx}$$

$$\underline{(A - BK)}$$

The goal is to find a control signal that optimizes certain criteria.

Goal: drive the states from an initial state x_0 to a desired state x^* while minimizing the energy of the control signal.

$u(t)$ from $t \in [0,T]$

$$\min_{u(t)} \int_0^T \|u(t)\|_2^2 \, dt$$

$$\text{s.t.} \quad x(T) = x^*, \quad x(0) = x_0$$
$$\dot{x} = f(x,u,t)$$

DT:

$$\min_{u_0, u_1, u_2, \dots, u_{T-1}} \sum_{k=0}^{T-1} \|u_k\|_2^2$$

$$\text{s.t.} \quad x(0) = x_0, \quad x(T) = x^*$$
$$x_{k+1} = f(x_k, u_k, k)$$

$$\boxed{\min_{u} \; g(u)}$$

Homework: Watch week 1 & week 2 of NPTEL Lectures "Convex Optimization in Control"

LECTURE-2 : 12th Jan

An optimization problem can be written as $\min_{x \in X} f(x)$,

where x : decision variable

X : set of feasible solutions

f : real-valued function, $f : X \to \mathbb{R}$

Given an optimization problem, we want to determine

(i) optimal value : $f^* \in \mathbb{R}$ s.t. $\boxed{f^* \leq f(x)}$ for all $x \in X$,

and f^* is the largest real-number satisfying the above property.

Notation

$\inf_{x \in X} f(x) \longrightarrow$ greatest lower bound on $f(x)$ when $x \in X$.

(ii) optimal solution: $\operatorname{argmin}_{x \in X} f(x)$

$x^* \in X$ is called an optimal solution if $f(x^*) = f^*$.

If we cannot find x^* s.t. $f(x^*) = f^*$, then, optimal solution does not exist for that problem.

Ex: $\min_{x \in [0,\infty)} e^{-x}$

optimal value $= 0$.

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

optimal solution does not exist

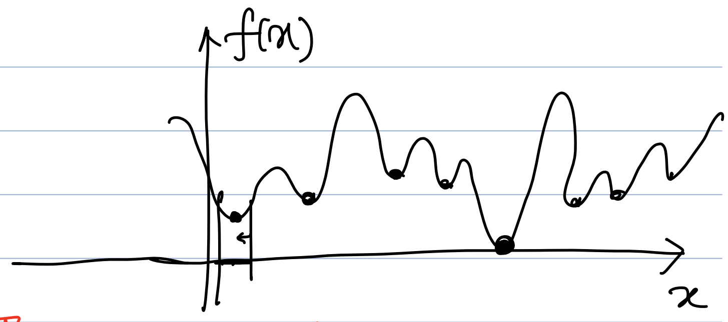
Global optimum: The quantity $x^* \in X$ is a global optimum if $f(x^*) \leq f(x)$ at every $x \in X$.

Local optimum: The quantity $x^* \in X$ is a local optimum if there exists $r > 0$ such that $f(x^*) \leq f(x)$ for all $x \in X$ and $\|x - x^*\|_2 \leq r$.

Examples:

$f(x) = \sin x$

$X = \mathbb{R}$



optimal value $= -1$

optimal solution: $\frac{3\pi}{2} + 2n\pi$, n: integer

each of which are global optima.

In this course, the default convention is that all optimization problems are minimization problems.

(A) $\max_{x \in X} f(x) \equiv \min_{x \in X} g(x)$, (B) $g(x) := -f(x)$

Optimal value of A = $-$ optimal value of B

The set of optimal solutions of A = set of optimal solutions of B.

"Optimal solution" be default means globally optimal solution.

The set of optimal solutions is denoted by

$$\boxed{\underset{x \in X}{\operatorname{argmin}}\ f(x)} = \{x^* \mid f(x^*) \le f(x) \text{ for all } x \in X\}$$

(1) If the set of feasible solutions X is an empty-set, then the problem is called infeasible. For such problems, we sometimes say optimal value $f^* = +\infty$.

(2) If the optimal value $f^* = -\infty$, then we say the optimization problem is unbounded.

Examples: ①
$$\min_{x \in \mathbb{R}^2} \ x_1 + x_2 \quad \text{s.t.} \quad x \geq 0, \ x_1 + x_2 = -1$$
: infeasible.

②
$$\min_{x \in \mathbb{R}} \ \tan x \quad \text{s.t.} \quad x \in [0, 2\pi]$$
: unbounded

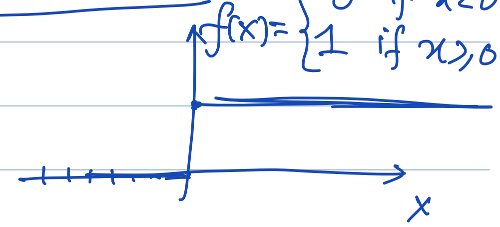
Existence of an optimal solution

Theorem: If the cost function f is continuous and the feasibility set X is closed and bounded, then there exists at least one global optimum.

Def^n: Let $(x_1, x_2 \dots)$ be a sequence s.t. $\lim_{n \to \infty} x_n = \bar{x}$ ← limit point

The function f is continuous if and only if $\lim_{n \to \infty} f(x_n) = f(\bar{x})$, for every sequence that converges to $\bar{x}$, $\forall \bar{x}$.

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$



$$x_n = \frac{-1}{n}$$

$$f(x_n) = 0 \neq f(0)$$

Def^n: A set X is closed if it contains all of its limit points.
equivalently, X is closed if its complement X^c is open.

$x \in [0, 1)$: neither open, nor closed.

open sets ⇍ $x \in (0, 1)$: $x > 0$, $x < 1$

$X = \mathbb{R}^n$: both open & closed.

closed set ⇐ $x \in [0, 1]$: $x \geq 0$, $x \leq 1$

Def^n: A set X is bounded if there exists $B>0$ s.t.

$$\|x-y\|_2 \leq B \quad \forall\ x,y \in X.$$

LECTURE 3: 12th Jan.

Consider an unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$, with $X = \mathbb{R}^n$, f: twice continuously differentiable. $\quad$ -(1)

Q What are the necessary and sufficient conditions for $x^* \in \mathbb{R}^n$ to be an optimal solution of (1)?

Derivative of $f: \mathbb{R}^n \to \mathbb{R}$ is denoted by $Df \in \mathbb{R}^{1\times n}$

$$Df(\bar{x}) = \left[\frac{\partial f}{\partial x_1}(\bar{x}),\ \frac{\partial f}{\partial x_2}(\bar{x}),\ \dots\ \frac{\partial f}{\partial x_n}(\bar{x})\right].$$

Recall that if $f: \mathbb{R} \to \mathbb{R}$, $\displaystyle\lim_{h\to 0} \frac{f(\bar{x}+h) - f(\bar{x})}{h} = \frac{df}{dx}(\bar{x})$

or $\underline{f(\bar{x}+h)} \simeq \underline{f(\bar{x})} + \boxed{\frac{df}{dx}(\bar{x})}\, h$

when h is sufficiently small.

When $f: \mathbb{R}^n \to \mathbb{R}$, both $\bar{x}$ and h are vectors.

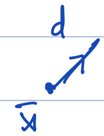
To ensure that $Df(\bar{x}) \cdot h$ remains a scalar, the convention of $Df(\bar{x})$ being a row vector is adopted.

Since most of the analysis is carried out with column vectors, we define gradient of the function $\nabla f(\bar{x}) = Df(\bar{x})^T \in \mathbb{R}^{n\times 1}$

$$= \begin{bmatrix} \partial f/\partial x_1(\bar{x}) \\ \vdots \\ \partial f/\partial x_n(\bar{x}) \end{bmatrix}.$$

$$f(\bar{x}+h) \simeq f(\bar{x}) + \underbrace{\nabla f(\bar{x})^T h}_{\text{directional derivative.}}$$

Directional derivative of f along direction $d \in \mathbb{R}^n$ at $\bar{x} \in \mathbb{R}^n$ is $\nabla f(\bar{x})^T d$.



$D f_1(x) = a^T$

Determine gradient of the following functions:

(1) $f_1(x) = a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$

(2) $f_2(x) = x^T Q x$, $\nabla f_2(x) = (Q + Q^T) x$

(3) $f_3(x) = \|Ax - b\|_2^2 = (Ax-b)^T(Ax-b) = x^T A^T A x - 2b^T A x + b^T b$,

$$\frac{\partial f_1}{\partial x_1} = a_1, \quad \frac{\partial f_1}{\partial x_2} = a_2 \; \cdots, \quad \nabla f(x) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

$$\nabla f_3(x) = 2A^T A x - 2A^T b$$

(2) $$f_2(x) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} Q_{11} & \cdots & Q_{1n} \\ \vdots & & \\ Q_{n1} & \cdots & Q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \sum_{j=1}^n Q_{1j} x_j \\ \sum_{j=1}^n Q_{2j} x_j \\ \vdots \\ \sum_{j=1}^n Q_{nj} x_j \end{bmatrix} = \sum_{i=1}^n x_i \left(\sum_{j=1}^n Q_{ij} x_j \right)$$

$$= x_1 \left(\sum_{j=1}^n Q_{1j} x_j \right) + \sum_{\substack{k \neq 1 \\ k=1}}^n x_k \left(\sum_{j=1}^n Q_{kj} x_j \right)$$

$x^T Q x = Q_{11} x_1^2 + Q_{12} x_1 x_2 + Q_{21} x_2 x_1 + Q_{22} x_2^2$

$2Q_{11} x_1 + Q_{12} x_2 + Q_{21} x_2$

$Q_{11} x_1 + Q_{12} x_2$

$+ Q_{11} x_1 + Q_{21} x_2$

$$= Q_{11} x_1^2 + \left(\sum_{j=2}^n Q_{1j} x_j \right) x_1 + \sum_{k=2}^n x_k \left(\sum_{j=1}^n Q_{kj} x_j \right)$$

$[Qx]_1 = Q_{11} x_1 + Q_{12} x_2$

$[Q^T x]_1 = Q_{11} x_1 + Q_{21} x_2$

$$\frac{\partial f_2}{\partial x_1} = \sum_{\substack{j=1 \\ j\neq 1}}^{n} Q_{1j} x_j + 2Q_{11}x_1 + \sum_{\substack{k\neq 1 \\ k=1}}^{n} x_k Q_{k1}$$

$$= \sum_{j=1}^{n} Q_{1j} x_j + \sum_{k=1}^{n} Q_{k1} x_k = [Qx]_1 + [Q^T x]_1$$

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

$[v]_1$: first element of the vector v

Definition: The Hessian of a function $f: \mathbb{R}^n \to \mathbb{R}$ at point $\bar{x} \in \mathbb{R}^n$ denoted by $\nabla^2 f(\bar{x}) \in \mathbb{R}^{n\times n}$ with (i,j)-th entry given by $\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x})$.

In other words,
$$\nabla^2 f(\bar{x}) = D\nabla f(x) = \begin{bmatrix} D\frac{\partial f}{\partial x_1} \\ D\frac{\partial f}{\partial x_2} \\ \vdots \\ D\frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \text{———} \\ \text{———} \\ \text{———} \end{bmatrix}$$

For the functions stated earlier,

i) $\nabla f_1(x) = a$, $\nabla^2 f_1(x) = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}$

ii) $\nabla f_2(x) = (Q+Q^T)x$, $\nabla^2 f(x) = Q+Q^T$

iii) $\nabla f_3(x) = 2A^T A$

For a function $f: \mathbb{R}^n \to \mathbb{R}$ at point $\bar{x} \in \mathbb{R}^n$, a direction d is called a direction of descent if the directional derivative $\nabla f(x)^T d < 0$. $\Rightarrow$ if $d = \nabla f(x)$,

$\nabla f(x)^T \nabla f(x) \geq 0$,

$$f(\bar{x}+\varepsilon d) = f(\bar{x}) + \varepsilon \nabla f(\bar{x})^T d + (h.o.t)$$

if $d = -\nabla f(x)$,

$-\nabla f(x)^T \nabla f(x) \leq 0$

$$\Rightarrow f(\bar{x}+\varepsilon d) - f(\bar{x}) = \varepsilon \nabla f(\bar{x})^T d + (h.o.t)$$

If ε is sufficiently close to zero, the R.H.S. is strictly negative.

$$\Rightarrow f(\bar{x}+\varepsilon d) < f(\bar{x}),$$

Note: $d = -\nabla f(x)$ is always a direction of descent whenever $\nabla f(x) \neq 0$.

Thus, if at any $\bar{x}$, $\nabla f(\bar{x}) \neq 0$, then $\bar{x}$ cannot be an optimal solution

Theorem: For an unconstrained optimization problem, if $\bar{x}$ is a (local) optimum, then $\nabla f(\bar{x}) = 0$.

Defn: For a function $f: \mathbb{R}^n \to \mathbb{R}$, if $\nabla f(\bar{x}) = 0$, then $\bar{x}$ is called a stationary point.

→ the above theorem specifies a necessary condition for optimality.

→ Not all stationary points correspond to optimal solutions.

LECTURES 4 & 5 : 13^{th} Jan

Let $\bar{x}$ be a stationary point.

Second order Taylor-series expansion at $\bar{x}+\varepsilon d$ for small ε is

$$f(\bar{x}+\varepsilon d) \simeq f(\bar{x}) + \varepsilon \nabla f(\bar{x})^T d + \frac{1}{2}\varepsilon^2 d^T \nabla^2 f(\bar{x}) d + (h.o.t.)$$

$$= f(\bar{x}) + \frac{1}{2}\varepsilon^2 \boxed{d^T \nabla^2 f(\bar{x}) d} + (h.o.t)$$

$\Rightarrow f(\bar{x}+\varepsilon d) - f(\bar{x})$ is negative if $d^T \nabla^2 f(\bar{x}) d$ is negative.

If $\exists\, \bar{d}$ s.t. $(\bar{d})^T \nabla^2 f(\bar{x}) \bar{d} < 0$, then $\bar{x}$ is not a minimizer.

Therefore, if $\bar{x}$ is a (local) minimizer, then

$$(\bar{d})^T \nabla^2 f(\bar{x}) \bar{d} \geq 0 \quad \forall\, \bar{d} \in \mathbb{R}^n$$

or equivalently, $\nabla^2 f(\bar{x})$ is a positive semidefinite matrix.

necessary condition.

$$\boxed{\nabla^2 f(\bar{x}) \succeq 0}$$

sufficient condition

Theorem: Let $\bar{x} \in \mathbb{R}^n$ be such that $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is a positive definite matrix. Then, $\bar{x}$ is a local minimizer of $\min_{x \in \mathbb{R}^n} f(x)$.

proof idea: From Taylor series expansion, we obtain

$$f(\bar{x}+\varepsilon d) = f(\bar{x}) + \frac{1}{2}\varepsilon^2 \, d^T \nabla^2 f(\bar{x}) d + \phi(\bar{x}, \varepsilon, d),$$

where there exists some ε^* s.t. whenever $\varepsilon < \varepsilon^*$,

$$\frac{1}{2}\varepsilon^2 d^T \nabla^2 f(\bar{x}) d > |\phi(\bar{x}, \varepsilon, d)|.$$

$$\Rightarrow f(\bar{x}+\varepsilon d) > f(\bar{x}) \quad \forall\, \varepsilon \in (0, \varepsilon^*), \quad \forall\, d \in \mathbb{R}^n.$$

Note: For the maximization problem $\max_{x \in \mathbb{R}^n} f(x)$,

(i) If $\bar{x}$ is a (local) maximizer, then $\nabla f(\bar{x}) = 0$

(necessity) and $\nabla^2 f(\bar{x}) \preceq 0$.

(negative semidefinite)

(ii) (sufficiency) If at some point $\bar{x}$, $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is negative definite, then $\bar{x}$ is a local maximizer.

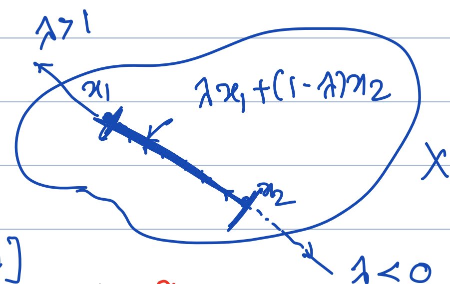
Q) When is a local optimum global?

Q) For a constrained optimization problem, what are the conditions for optimality?

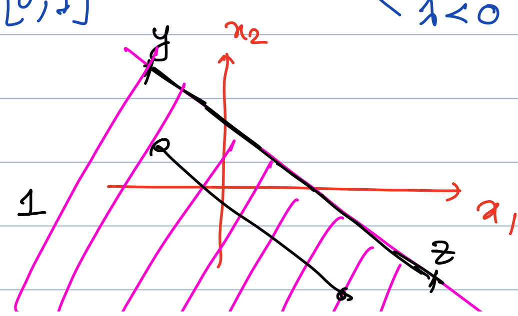
→ to answer this, we need to introduce convex sets and convex functions.

CONVEX SETS

Def^n: A set X is convex if for any $x_1, x_2 \in X$, we have $\lambda x_1 + (1-\lambda) x_2 \in X$, whenever $\lambda \in [0,1]$.



To show that a set is not a convex set, we only need to find one instance that falsifies the definition, i.e., find $\bar{x}_1, \bar{x}_2$ and $\bar{\lambda} \in [0,1]$ s.t. $\bar{\lambda}\bar{x}_1 + (1-\bar{\lambda})\bar{x}_2 \notin X$.



Examples: $X_1 = \{ x \in \mathbb{R}^2 \mid x_1 + x_2 = 1 \}$

Let $y, z \in X_1 \Rightarrow y_1 + y_2 = 1,\ z_1 + z_2 = 1$

$$w = \lambda y + (1-\lambda) z = \begin{bmatrix} \lambda y_1 + (1-\lambda) z_1 \\ \lambda y_2 + (1-\lambda) z_2 \end{bmatrix}$$

$X_2 = \{ x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 1 \}$

to check $w \in X_1$, we will calculate $w_1 + w_2$

$$= \lambda y_1 + (1-\lambda) z_1 + \lambda y_2 + (1-\lambda) z_2$$
$$= \lambda (y_1 + y_2) + (1-\lambda)(z_1 + z_2)$$
$$= \lambda + (1-\lambda) = 1.$$

This proves that X_1 is a convex set.

In general, $X = \{ x \in \mathbb{R}^n \mid \underbrace{a^T x \leq b}_{\text{half-space}} \}$ is a convex set.

Properties of convex sets

(1) Intersection of two convex sets is a convex set.
(by induction, it follows that intersection of arbitrarily many convex sets is a convex set)

proof: Let X_1 and X_2 be two convex sets.
Let $Z = X_1 \cap X_2$.
Let $z_1, z_2 \in Z$. $\Rightarrow$ $z_1 \in X_1$, $z_2 \in X_1$, and $z_1 \in X_2$ and $z_2 \in X_2$.

To establish convexity of the set Z, we need to check whether $\lambda z_1 + (1-\lambda) z_2$ belongs to Z when $\lambda \in [0,1]$.

Observe that because X_1 is a convex set and $z_1 \in X_1$, $z_2 \in X_1$, we have $\lambda z_1 + (1-\lambda) z_2 \in X_1$ whenever $\lambda \in [0,1]$.
Similarly, $\lambda z_1 + (1-\lambda) z_2 \in X_2$ whenever $\lambda \in [0,1]$.
$\Rightarrow$ $\lambda z_1 + (1-\lambda) z_2 \in X_1 \cap X_2$ for any $\lambda \in [0,1]$. ∎

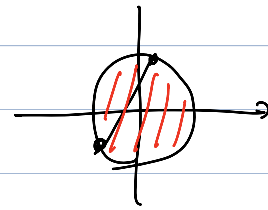
$X_1 = \{x \in \mathbb{R}^n \mid a^T x \leq b\}$

$X_2 = \{x \in \mathbb{R}^n \mid a^T x \geq b\}$ → convex sets

$\Rightarrow X_1 \cap X_2 = \{x \in \mathbb{R}^n \mid \underbrace{a^T x = b}_{\text{hyperplanes}}\}$ → convex set

Few more examples:

$X_1 = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$: not a convex set



$X_2 = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$: convex set

$X_3 = \{x \in \mathbb{R}^n \mid \|x\|_2^2 \leq c\}$

$X_4 = \{x \in \mathbb{R}^n \mid \|x\|_2^2 = c\}$ → not a convex set.

Convex functions

Defn: A function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be a convex function if

a) dom f is a convex set.

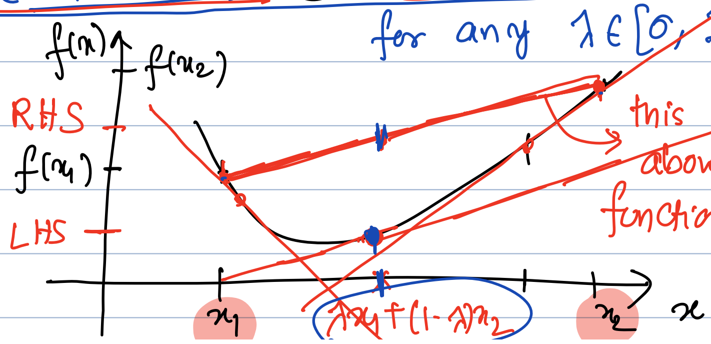
$$\text{dom} f = \{x \in \mathbb{R}^n \mid f(x) < \infty\}$$

b) If $x_1, x_2 \in \text{dom} f$,

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2),$$

for any $\lambda \in [0,1]$.

A function is strictly convex if the inequality is strict for $\lambda \in (0,1)$

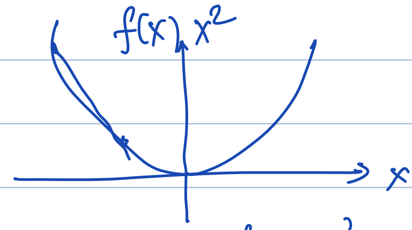


this line lies above the function for $x \in [x_1, x_2]$.

Examples: $f_1(x) = a^T x$, $x \in \mathbb{R}^n \rightarrow$ domain: $\mathbb{R}^n$ convex

$$f_1(\lambda x_1 + (1-\lambda) x_2)$$
$$= a^T \lambda x_1 + (1-\lambda) a^T x_2$$
$$= \lambda f(x_1) + (1-\lambda) f(x_2)$$

$f_2(x) = x^2$, $x \in \mathbb{R}$

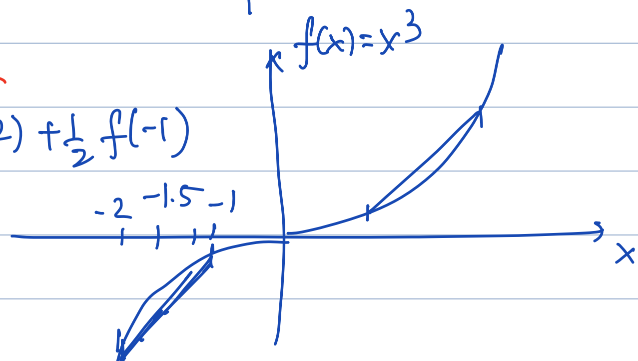


$f_3(x) = x^3$, $x \in \mathbb{R}$

$$f(-1.5) > \frac{1}{2} f(-2) + \frac{1}{2} f(-1)$$

$\downarrow$

not a convex function.



To show that a function is not convex, only one counter-example is sufficient.

Alternative Equivalent definitions

Defn: Let f be a differentiable function. f is convex if & only if dom f is a convex set, and for any $x, y \in \text{dom} f$,

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

$\ell_x(y)$: linear function of y if x is assumed to be fixed, and it is a tangent to f at x.

$\ell_x(x) = f(x)$

$\nabla_x \ell(y) = \nabla f(x)$

$\ell_x(y)$ is constructed only using local information at point x, but acts as a lower bound on the function $f(y)$ everywhere in the domain.

Def^n: Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. Then, the function is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom} f$, with $\text{dom} f$ being a convex set.

Example: $f_1(x) = q^T x \rightarrow \nabla^2 f_1(x) \in \mathbb{R}^{n \times n}$ with all entries 0.

$\Rightarrow \nabla^2 f_1(x) \succeq 0$

$\Rightarrow f_1$ is a convex function.

$f_2(x) = x^T Q x$

$f_3(x) = \|Ax - b\|_2^2$

$\nabla^2 f_2(x) = Q + Q^T \Rightarrow f_2$ is convex if and only if $Q + Q^T$ is positive semidefinite.

$\nabla^2 f_3(x) = 2A^T A$

$w^T \nabla^2 f_3(x) w = 2 w^T A^T A w = 2 \|Aw\|_2^2 \geq 0$

$\Rightarrow f_3$ is always a convex function.

Note: A function is concave if $\text{dom} f$ is a convex set, and $-f(x)$ is a convex function. Equivalently,

i) $\forall x_1, y_1 \in \text{dom} f$, $f(\lambda x_1 + (1-\lambda) x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$ $\forall \lambda \in [0,1]$.

(ii) If f is differentiable,

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

(iii) If f is twice-differentiable, $\nabla^2 f(x)$ is negative semidefinite. $\forall x \in \text{dom} f$.

Examples: Determine if the following functions are convex, concave or neither.

$f_1(x) = \log x$

$f_2(x) = \sqrt{x}$

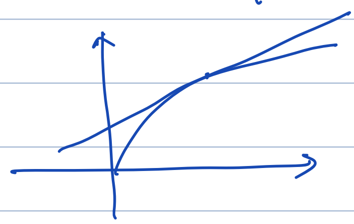
$f_3(x) = -x^2$, $x \in \mathbb{R}$

$f_4(x) = x_1 x_2$, $x \in \mathbb{R}^2$

Solution (1) dom f_1 : $(0,\infty)$

$$\frac{d^2 f_1}{dx^2} = -\frac{1}{x^2} < 0 \Rightarrow f_1 \text{ is concave}$$

(2) dom f_2 : $[0,\infty)$, $\frac{df_2}{dx} = \frac{1}{2}x^{-1/2}$, $\frac{d^2 f_2}{dx_2^2} = -\frac{1}{4}x^{-3/2} < 0$



$\Rightarrow f_2$ is concave

(3) concave since x^2 is convex.

(4) $\nabla f_4(x) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$, $\nabla^2 f_4(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$: eigenvalues: $\{1, -1\}$

$\rightarrow f_4$ is neither convex, nor concave.

LECTURE 6 & 7 : 19th Jan

Recap of previous week

- defined optimization problem
  - $\rightarrow$ local & global optima
  - $\rightarrow$ optimal value & optimal sol^n
- conditions that guarantee existence of at least one global optimum
- necessary and sufficient conditions for local optimum for unconstrained problems
- convex sets
- convex functions.

Grading policy.

Midsem: 30%

endsem: 30%

attendance: 10%

tests : 30%

best 3 out of 4 $\begin{cases} \rightarrow 2 \text{ surprise} \\ \rightarrow 2 \text{ announced} \end{cases}$

Properties of convex functions

(A) If $f_1, f_2, \ldots, f_K$ constitute a collection of convex functions, then

$$g(x) := \sum_{j=1}^{K} \alpha_j f_j(x), \quad \text{with} \quad \alpha_j \geq 0$$

is also a convex function.

to prove this, show that (1) $\text{dom}\, g = \bigcap_{j=1}^{K} \underbrace{\text{dom}\, f_j}_{\text{each convex set}}$ → intersection is also a convex set

(2) $g(\lambda x + (1-\lambda)y) = \sum_{j=1}^{K} \alpha_j f_j(\lambda x + (1-\lambda)y)$

$$\leq \sum_{j=1}^{K} \alpha_j \left[\lambda f_j(x) + (1-\lambda) f_j(y)\right]$$

$$= \lambda g(x) + (1-\lambda) g(y) \quad \forall x, y \in \text{dom}\, g, \ \lambda \in [0,1].$$

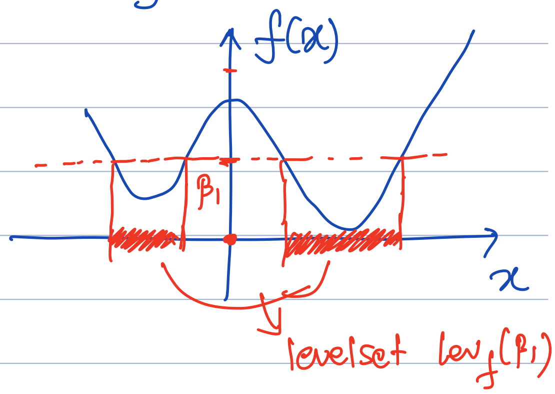
pointwise maximum

(B) $g(x) := \max_{1 \leq j \leq K} f_j(x)$ is also a convex function. → HW.

(C) Defn: (Sub-level set) For a function $f: \mathbb{R}^n \to \mathbb{R}$, its sublevel set at level $\beta \in \mathbb{R}$ is

$$\text{lev}_f(\beta) := \{x \in \mathbb{R}^n \mid f(x) \leq \beta\}.$$

If f is a convex function, then $\text{lev}_f(\beta)$ is a convex set for any $\beta \in \mathbb{R}$.



The converse is not true. think of an example.

convex optimization problem

An optimization problem $\min_{x \in X} f(x)$ is a convex optimization problem if (i) X is a convex set
(ii) $f(x)$ is a convex function.

Examples:

(1) $\min_{x \in \mathbb{R}} (5-x)^2 \rightarrow 25 + x^2 - 10x$ → convex
s.t. $x \geq 0$

(2) $\max_{x \in \mathbb{R}} \log x$ s.t. $x > 5$ $\equiv$ $\min_{x \in \mathbb{R}} -\log x$ → convex function
s.t. $x > 5$ → convex set.

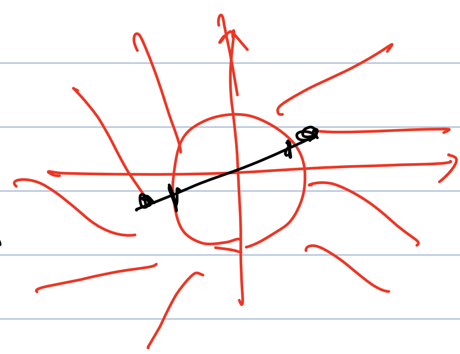
$x \geq 0 \iff -x \leq 0$
$x \leq e^2$

(3) $\min_{x \in \mathbb{R}} x^2$ → convex function
s.t. $\log_e(x) \leq 2$ → $x \in [0, e^2]$: convex set

$\log_e(x) \leq \log_e(e^2) = 2$

(4) $\min_{x \in \mathbb{R}} -e^{-x}$ → not a convex function
s.t. $x \in (1, 5)$

(5) $\min_{x \in \mathbb{R}^n} c^T x$ → ✓
s.t. $\|x\|_2^2 \geq 5$
↓
not a convex set



Note: (1) There is no such thing as a concave set.
(2) There is nothing called concave optimization problem.

Properties of convex optimization problems

$$\min_{x \in \mathbb{R}^n} \; -\|x\|_2^2 \quad \text{s.t.} \quad \|x\|_2^2 \le 1$$

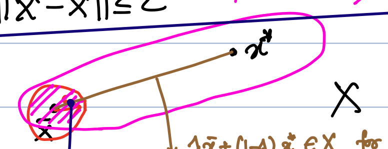
$\rightarrow$ optimal solution is not unique.

① For a convex optimization problem, a locally optimal solution is globally optimal.

proof: Let $\bar{x}$ be a locally optimal solution.

$\Rightarrow \exists\, \varepsilon > 0$ s.t. $f(\bar{x}) \le f(x) \;\; \forall\, \|x - \bar{x}\| \le \varepsilon$ and $x \in X$

Let x^* be any other feasible solution, i.e., $x^* \in X$.



$\lambda\bar{x} + (1-\lambda)x^* \in X$ for any $\lambda \in [0,1]$. Since X is a convex set.

let this point be $z = \lambda_1 \bar{x} + (1-\lambda_1)x^*$, with $\lambda_1 \in [0,1]$ such that $\|z - \bar{x}\| \le \varepsilon$.

$z \in X. \Rightarrow f(\bar{x}) \le f(z)$

$$f(\bar{x}) \le f(z) = f(\lambda_1 \bar{x} + (1-\lambda_1)x^*)$$
$$\le \lambda_1 f(\bar{x}) + (1-\lambda_1) f(x^*)$$
$$\Rightarrow (1-\lambda_1) f(\bar{x}) \le (1-\lambda_1) f(x^*)$$
$$\Rightarrow f(\bar{x}) \le f(x^*)$$

$\longrightarrow$ this is true for any $x^* \in X$.

$\Rightarrow \bar{x}$ is a globally optimal solution.

② If the function f is strictly convex, then there exist at most one optimal solution.

proof: Suppose $\exists$ two optimal solutions, $x_1, x_2 \in X$, $x_1 \ne x_2$.

$\min_{x \in X} f(x)$ $\rightarrow$ convex set

$$\Rightarrow \underbrace{f(x_1)}_{f^* =} = f(x_2) \le f(x) \quad \forall\, x \in X.$$

Let $z = \frac{x_1 + x_2}{2}$. Is $z \in X$? Yes since X is a convex set.

the inequality is strict as f is strictly convex.

$$\underline{\underline{f(z)}} = f\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) < \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = f^*.$$

$\Rightarrow$ x_1 and x_2 are not optimal since $f(z) < f^*$.

This is a contradiction to the initial hypothesis.

$\Rightarrow$ When f is strictly convex, we have at most one optimal solution.

Classes of convex optimization problems

In most applications, the set of feasible solutions X is defined in terms of inequality and equality constraints.

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} g_1(x) \le 0, \; \ldots \; g_m(x) \le 0, \\ h_1(x) = 0, \; \ldots \; h_p(x) = 0 \end{array} \right\}$$

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

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

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

subject to (s.t.)

$$\left.\begin{array}{l} g_1(x) \le 0 \\ g_2(x) \le 0 \\ \vdots \\ g_m(x) \le 0 \end{array}\right\} \text{inequality constraints.}$$

inequality constraints with '<' relation may render the feasibility set to be an open set, for which optimal solution may not exist.

$$\left.\begin{array}{l} h_1(x) = 0 \\ h_2(x) = 0 \\ \vdots \\ h_p(x) = 0. \end{array}\right\} \Rightarrow \text{equality constraints}$$

$a^T x \ge b$

$\Updownarrow$

$-a^T x \le -b$

$\Updownarrow$

$-a^T x + b \le 0$

$a^T x = b$

$\Updownarrow$

$a^T x - b = 0$

<u>Claim</u>: The feasibility set X is guaranteed to be a convex set if $g_i(x)$ are convex functions and $h_j(x)$ are affine functions ($h_j(x) = \underline{a_j^T x + b_j}$).

$$X = \bigcap_{i=1}^{m} \underbrace{\{x \in \mathbb{R}^n \mid g_i(x) \le 0\}}_{\text{sub-levelset of } g_i \text{ at level } 0.} \quad \bigcap_{j=1}^{p} \{x \in \mathbb{R}^n \mid \underline{h_j(x) = 0}\}$$

$\{x \in \mathbb{R}^n \mid h_j(x) \le 0\}$

$\cap \{x \in \mathbb{R}^n \mid -h_j(x) \le 0\}$

If all g_i's are convex functions, the intersection is a convex set.

If g_i is a convex function, $\{x \in \mathbb{R}^n \mid g_i(x) \le 0\}$ is a convex set.

For the same argument to hold, we need h_j to be convex and $-h_j$ to be convex

$\Leftarrow$ h_j needs to be convex and concave simultaneously.

$\Downarrow$

only affine functions have this property.

<u>(A) Linear programming (LP)</u>: The functions $f, (g_i)_{i=1,2..m}, (h_j)_{j=1,2..p}$ are all affine functions.

Since affine functions are convex, each LP is a <u>convex optimization</u> problem.

$$\min_{x \in \mathbb{R}^n} \quad c^T x + d \rightarrow \text{usually dropped.}$$

s.t.

$a_i^T x \le b_i, \quad i = 1, 2 \dots m \iff \boxed{Ax \le b}$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

$g_j^T x = h_j, \quad j = 1, 2 \dots p \iff Gx = h$

(B) Quadratic programming (QP) : cost function is a quadratic f^n; constraints are defined in terms of affine functions.

Q: symmetric

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2} x^T Q x + c^T x + d$$

$$\text{s.t.} \quad Ax \leq b$$
$$Gx = h$$

$\nabla^2 f(x) = Q$

The problem is convex if Q is positive semidefinite.

(C) Quadratically constrained Quadratic programs (QCQPs)

$$\min_{x \in \mathbb{R}^n} \ \frac{1}{2} x^T Q x + c^T x + d$$

$$\text{s.t.} \quad \frac{1}{2} x^T A_i x + b_i^T x + c_i \leq 0 \ , \quad i = 1, 2 \ldots m$$

$$\frac{1}{2} x^T G_i x + h_i^T x + f_i = 0 \ , \quad i = 1, 2 \ldots p$$

(Q, A_i, G_i) are symmetric matrices.

When is a QCQP convex: $Q \succeq 0$, $A_i \succeq 0$ $\forall i$, $G_i = 0$, $\forall i$

LECTURE 8 & 9 : 20th Jan.

Duality and optimality conditions

Consider an optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } g_i(x) \le 0, \quad i = 1, 2, \dots m$$
$$h_j(x) = 0, \quad j = 1, 2 \dots p$$

λ_i : multiplier associated with $g_i(x) \le 0$

μ_j : multiplier associated with $h_j(x) = 0$

(P) – primal optimization problem.

$\mu \in \mathbb{R}^p$

Lagrangian of (P) is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$\lambda \in \mathbb{R}^m$

defined as :

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Suppose $\bar{x}$ is feasible to (P), and $\bar{\lambda} \ge 0$ (each $\lambda_i \ge 0 \ \forall i = 1, 2 \dots m$)

Then,

$$f(\bar{x}) \ge L(\bar{x}, \bar{\lambda}, \bar{\mu}) \ge \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}, \bar{\mu}) =: d(\bar{\lambda}, \bar{\mu})$$

d : dual function

In other words, whenever $\bar{\lambda} \ge 0$,

$d(\bar{\lambda}, \bar{\mu}) \le f(x)$ for any $x \in \mathbb{R}^n$ that is feasible to (P).

Therefore, $d(\bar{\lambda}, \bar{\mu}) \le f^*$, where f^* is the optimal value of (P).

Whenever $\lambda \ge 0, \mu \in \mathbb{R}^p$, $d(\lambda, \mu)$ acts as a lower bound on f^*.

The best possible lower bound is obtained by maximizing the dual function

$$\max_{\lambda, \mu} d(\lambda, \mu) \quad \text{s.t. } \lambda \ge 0 \qquad \text{(D)}$$

not a convex problem, but satisfies strong duality.

Examples :

① $\begin{array}{ll} \min\limits_{x \in \mathbb{R}^2} & -x_1^2 - x_2^2 \\ \text{s.t.} & x_1^2 + x_2^2 - 1 \le 0. \quad \lambda \end{array}$: optimal value: -1

$$L(x,\lambda) = -x_1^2 - x_2^2 + \lambda(x_1^2 + x_2^2 - 1)$$
$$= (\lambda - 1)(x_1^2 + x_2^2) - \lambda$$

$$d(\lambda) = \inf_{x \in \mathbb{R}^2} \left[(\lambda - 1)(x_1^2 + x_2^2) - \lambda \right] = \begin{cases} -\infty & \text{when } \lambda < 1 \\ -\lambda & \text{when } \lambda \ge 1 \end{cases}$$

$$\max_{\lambda \ge 0} d(\lambda) \equiv \max_{\lambda \ge 1} -\lambda$$

→ optimal solution $\lambda^* = 1$
optimal value $= -1$

In this problem, optimal value of dual = optimal value of primal, which gives us certificate of optimality.

optimization problems that satisfy the above are called to satisfy strong duality property.

constraint qualification

If (P) is a convex optimization problem, and there exists a feasible solution $\bar{x}$ that satisfies

$\boxed{g_i(\bar{x}) < 0}$ $\forall\ i = 1, 2 \dots m$,

then we say that (P) satisfies Slater's constraint qualification condition.

Theorem: If (P) is a convex optimization problem that satisfies slater's constraint qualification condition, then strong duality holds.
(optimal value of (P) = optimal value of its dual)

Example:
$$\min_{x \in \mathbb{R}} \; -x^2 \quad \text{s.t.} \begin{cases} x - 1 \leq 0 & : \lambda_1 \\ -x \leq 0 & : \lambda_2 \end{cases}$$
$x \in [0,1]$

: check whether strong duality holds.

$$\mathcal{L}(x,\lambda) = -x^2 + \lambda_1(x-1) - \lambda_2 x$$

$$d(\lambda) = \underline{\inf_{x \in \mathbb{R}} \left[-x^2 + (\lambda_1 - \lambda_2)x \right]} - \lambda_1 = -\infty \quad \forall \lambda_1, \lambda_2$$

$\rightarrow$ strong duality does not hold

Example:
$$\min_{x \in \mathbb{R}^n} \; c^T x \qquad A \in \mathbb{R}^{m \times n}$$
$$\text{s.t.} \quad Ax = b \quad : \mu \in \mathbb{R}^m$$
$$x \geq 0 \iff -x \leq 0 \quad : \lambda \in \mathbb{R}^n$$

$$\mathcal{L}(x,\lambda,\mu) = c^T x - \lambda^T x + \mu^T [Ax - b]$$
$$= \underline{[c^T - \lambda^T + \mu^T A]} x - \mu^T b$$

$$\inf_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda,\mu) = \begin{cases} -\mu^T b & \text{when } \underline{c^T - \lambda^T + \mu^T A = 0} \\ -\infty & \text{otherwise.} \end{cases}$$

Dual problem:
$$\max_{\lambda,\mu} \; -\mu^T b$$
$$\text{s.t.} \quad \lambda \geq 0, \quad c^T - \lambda^T + \mu^T A = 0$$

$$\Updownarrow \qquad \underline{c^T + \mu^T A = \lambda^T \geq 0}$$

$$\begin{bmatrix} \max_{\mu} & -\mu^T b \\ \text{s.t.} & c^T + \mu^T A \geq 0 \end{bmatrix}$$

- Dual of a LP is also a LP.

Theorem: Strong duality always holds for a LP.

Optimality conditions

When $\lambda \geq 0$, and $\bar{x}$ is feasible to (P),

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

$$d(\lambda, \mu) = \inf_{x \in \mathbb{R}^n} \left[f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x) \right]$$

$$\text{(i)} \leq f(\bar{x}) + \sum_{i=1}^{m} \lambda_i g_i(\bar{x}) + \sum_{j=1}^{p} \mu_j h_j(\bar{x})$$

$$\text{(ii)} \leq f(\bar{x}).$$

Both inequalities hold with equality when

(ii) $\sum_{i=1}^{m} \lambda_i g_i(\bar{x}) = 0 \iff \lambda_i g_i(\bar{x}) = 0 \quad \forall i = 1, 2, \ldots m$

(i) $\bar{x}$ is the minimizer of $L(x, \lambda, \mu)$, which is an unconstrained optimization problem.

First order necessary condition: $\nabla L_x(x, \lambda, \mu) = 0$

Second order necessary condition: $\nabla^2 L_{xx}(x, \lambda, \mu) \succeq 0$

Second order sufficient condition: $\nabla^2 L_{xx}(x, \lambda, \mu) \succ 0$

Necessary condition under strong duality

If x^* is optimal solution of (P), then there exists $\lambda^* \geq 0$ and μ^* s.t.

(i) $g_i(x^*) \leq 0$, $h_j(x^*) = 0$ $\forall\, i = 1, 2 \ldots m$, $j = 1, 2 \ldots p$ (primal feasibility)

(ii) $\lambda^* \geq 0$

(iii) $\lambda_i^* g_i(x^*) = 0$ $\forall\, i = 1, 2 \ldots m$

(iv) $\nabla f(x^*) + \sum_{i=1}^{N} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0.$

Sufficient conditions: If (P) is a convex optimization problem, and we find (x^*, λ^*, μ^*) that satisfy (i) – (iv), then

i) x^* is optimal solution of (P)

ii) (λ^*, μ^*) are optimal solution of (D)

iii) strong duality holds.

Special case: when the problem does not have inequality constraints, the condition boils down to

$$\nabla f(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0.$$

$$h_j(x^*) = 0 \;\forall\, j = 1, 2 \ldots p$$

number of unknowns: $n + p$

number of equations: $n + p$