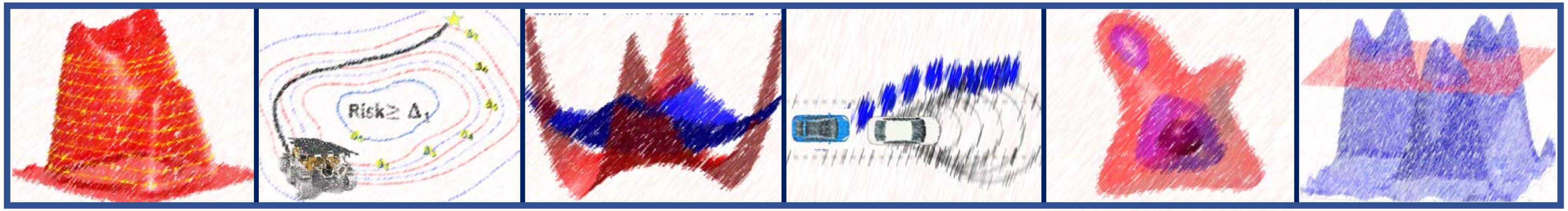
Lecture 2

Nonlinear Optimization

Overview

MIT 16.S498: Risk Aware and Robust Nonlinear Planning
Fall 2019

Ashkan Jasour



Overview of Nonlinear Optimization:

- Optimality Conditions
- Newton's Method
- Interior Point Method
- Convex Optimization
- Dual Optimization

Overview of Nonlinear Optimization:

- Optimality Conditions
- Newton's Method
- Interior Point Method
- Convex Optimization
- Dual Optimization

Why do wee need to look at these techniques:

- SDP solvers work based on the same principals (interior point method).
- We need these techniques to develop state-of-the art solvers for large SDP's in the lecture 9.
- We can formulate uncertain (robust, risk aware)optimization problems as a deterministic nonlinear optimization problem.

How to solve optimization problems?

- Optimality Conditions
- Newton's Method
- Interior Point Method

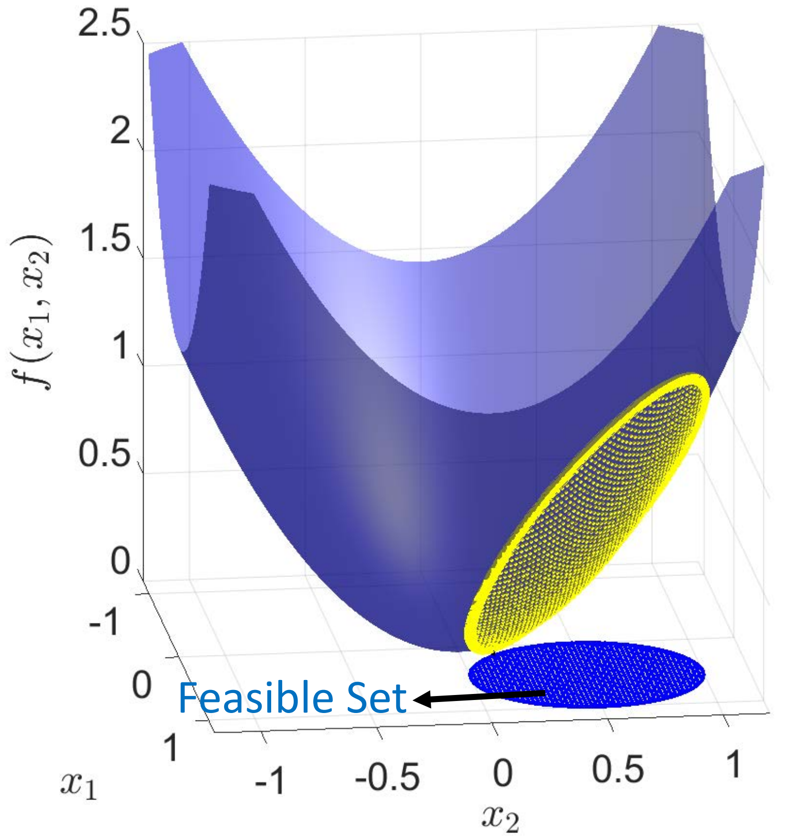
Optimization:

Minimize Objective-function(*design parameters*)

design parameters

Subject to Constraints(*design parameters*)

Design variables : $x = [x_1, ..., x_n]^T \in \mathbb{R}^n$



Real-Valued Scalar Continuous functions

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

$$g_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, ..., n_g$$

$$h_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, ..., n_h$$

Nonlinear Optimization:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) && \leftarrow \text{Objective function} \\ \text{subject to} \quad & \left.\begin{aligned} & g_i(x) \geq 0, \ i = 1, ..., n_g \\ & h_i(x) = 0, \ i = 1, .., n_h \end{aligned}\right\} \text{Feasible Set} && \begin{aligned} & \leftarrow \text{Inequality Constraints} \\ & \leftarrow \text{Equality Constraints} \end{aligned} \end{aligned}$$

Optimization:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \;\; i = 1, ..., n_g \\ & h_i(x) = 0, \;\; i = 1, .., n_h \end{aligned}$$

Step 1

Optimality Conditions

Given the objective function f and constraints g_i, h_i :

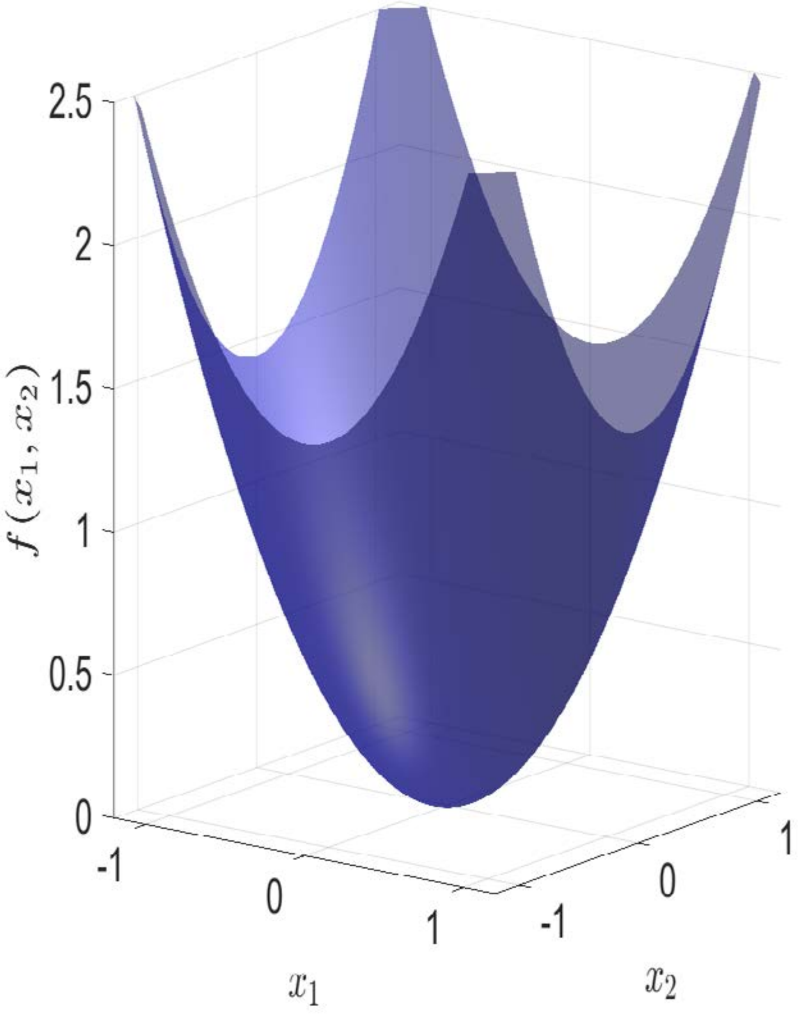
What are the conditions for x^* to be an optimal solution ?

Optimality Conditions: system of nonlinear equations or inequalities

Step 2

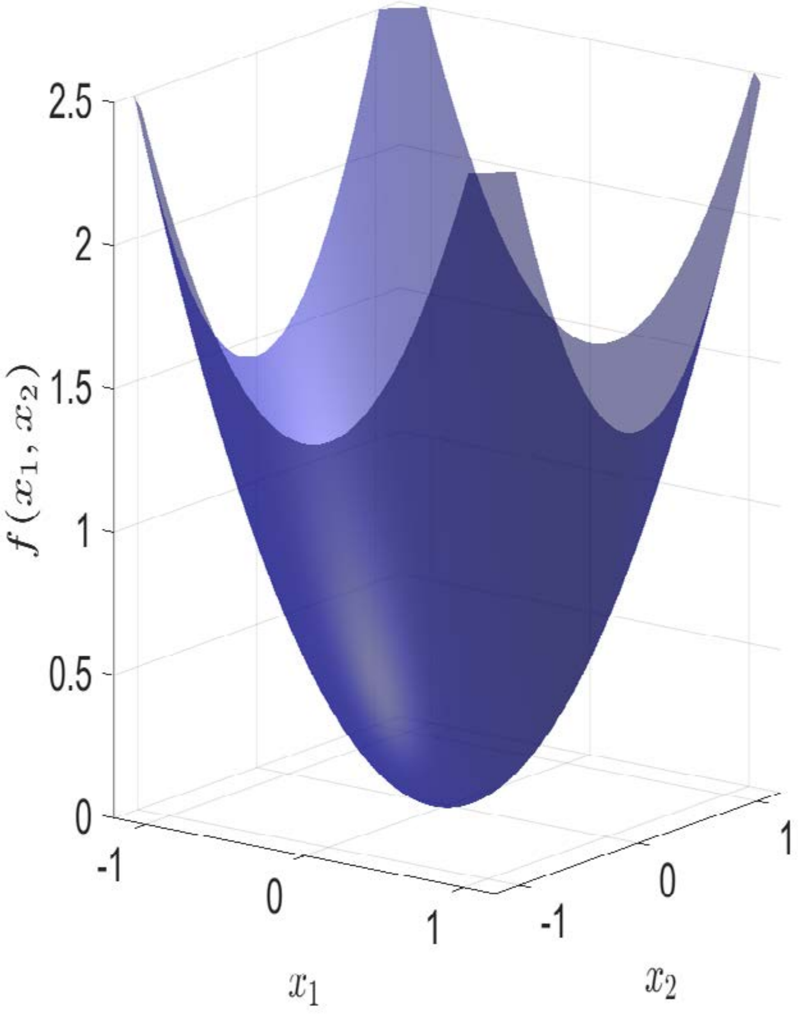
To find x^*, we solve the system of nonlinear equations or inequities (root finding problem).

Unconstrained Optimization



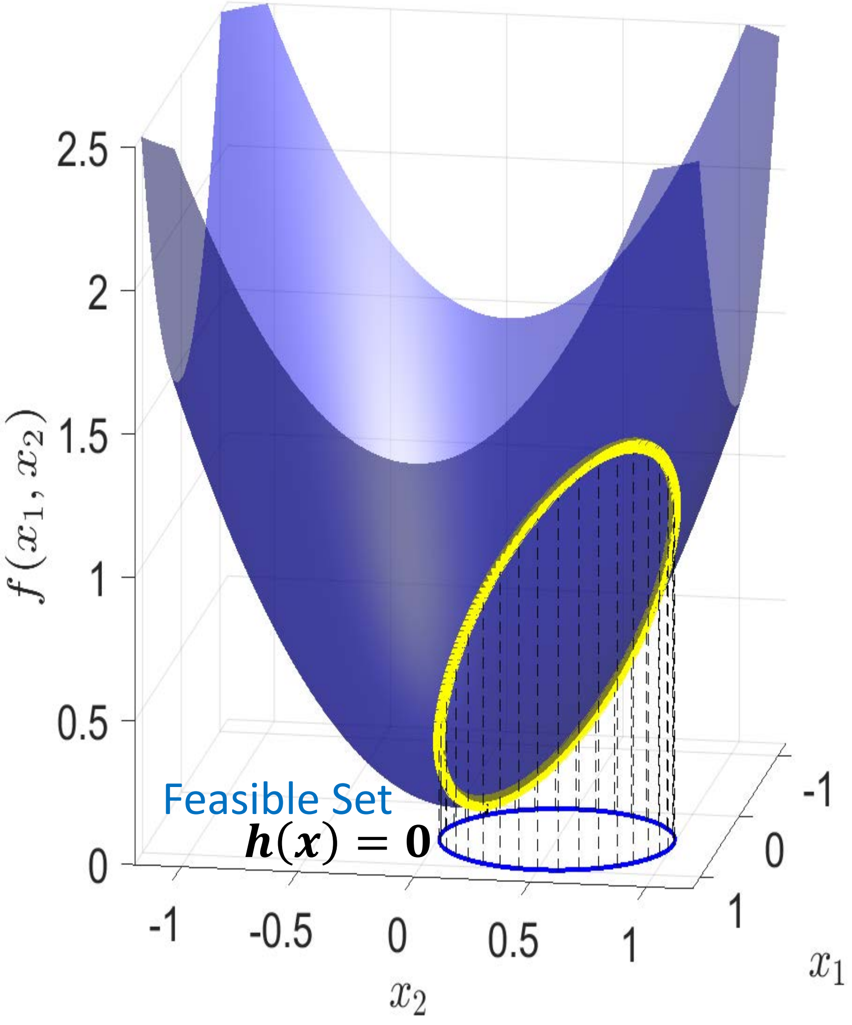
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Unconstrained Optimization



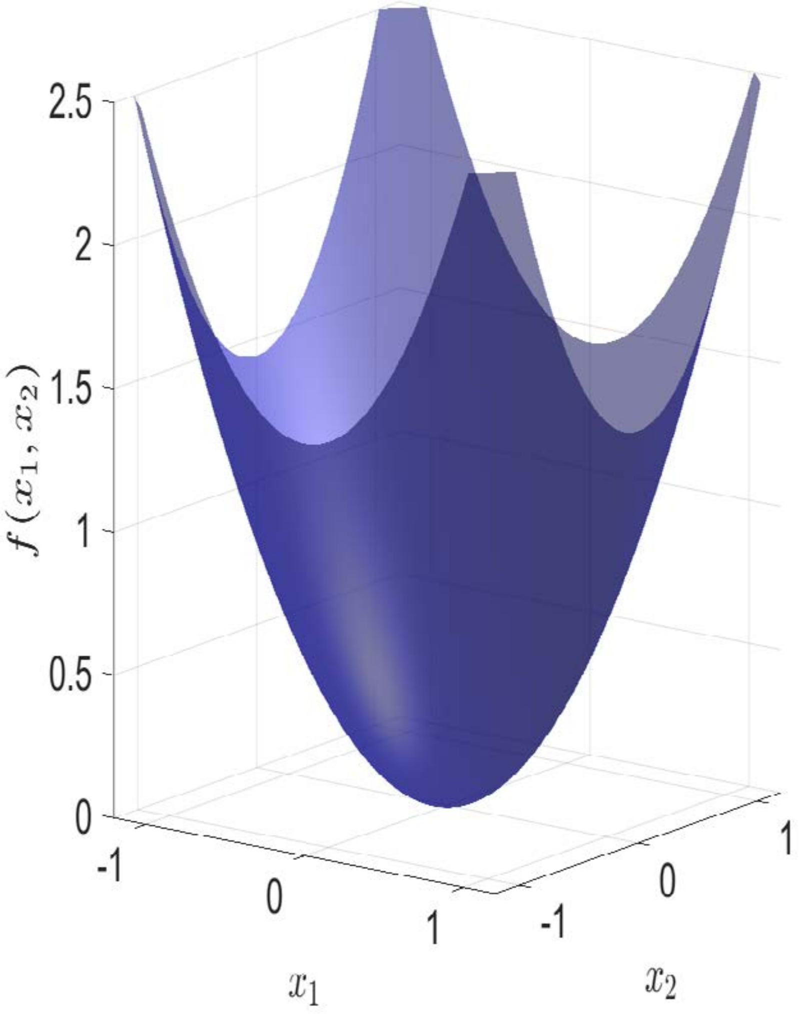
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Optimization with equality Constraints



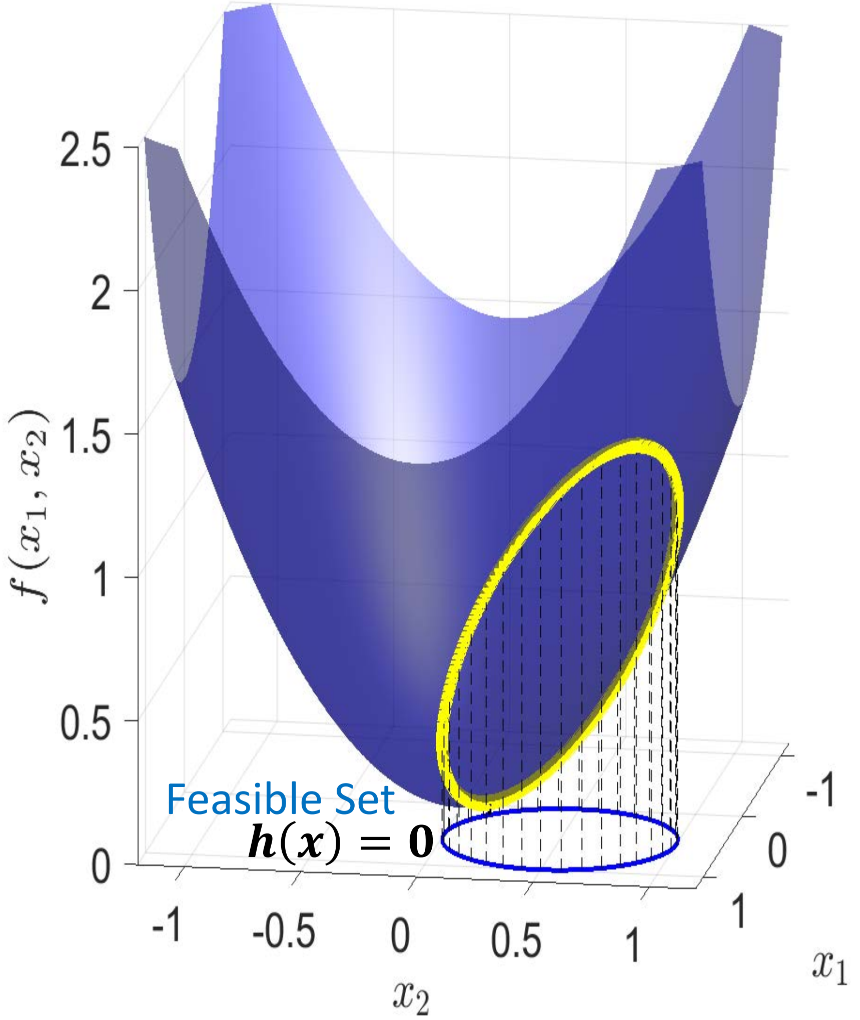
$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\ &\text{subject to} && h_i(x) = 0, \ i = 1, .., n_h \end{aligned}$$

Unconstrained Optimization



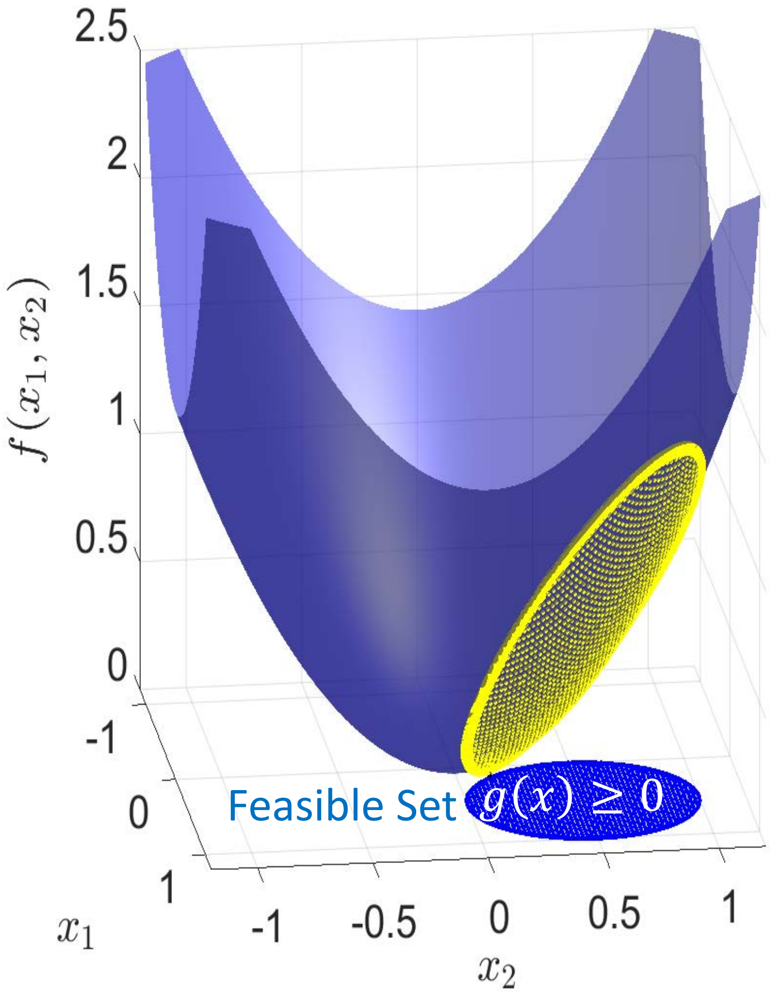
$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Optimization with equality Constraints



$$\begin{aligned}&\underset{x\in\mathbb{R}^n}{\text{minimize}} && f(x)\\ &\text{subject to} && h_i(x) = 0,\ \ i = 1, .., n_h\end{aligned}$$

Optimization with Inequality Constraints



$$\begin{aligned}&\underset{x\in\mathbb{R}^n}{\text{minimize}} && f(x)\\ &\text{subject to} && g_i(x) \geq 0,\ \ i = 1, ..., n_g\end{aligned}$$

Basic Definitions

- Gradient
- Hessian
- Global and Local Minima

Gradient

Gradient

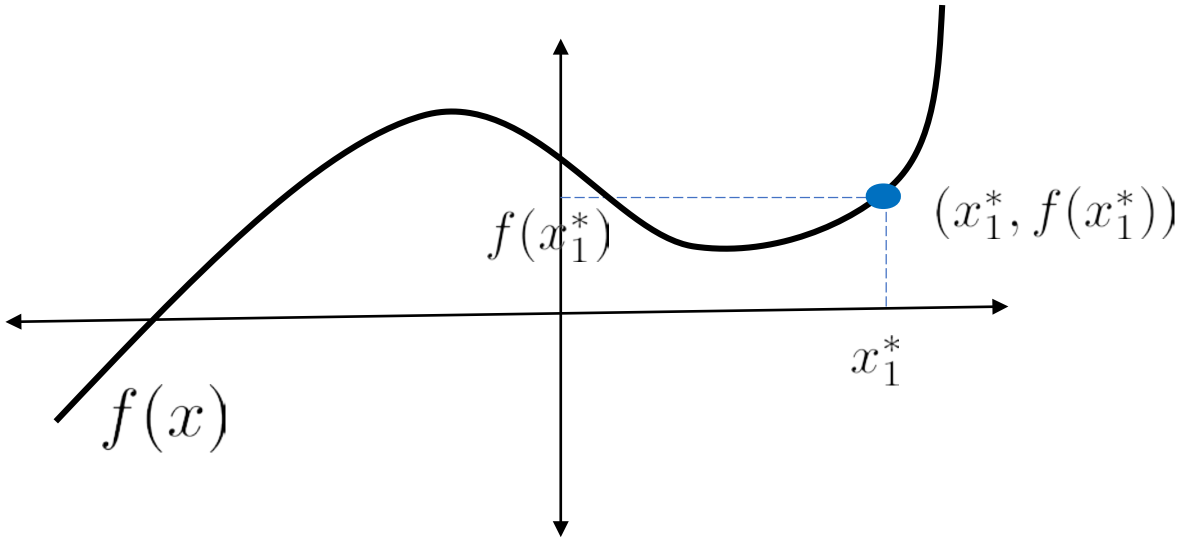
Real-valued scalar function $f(x): \mathbb{R}^n \rightarrow \mathbb{R}$

Gradient vector : $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$

$$f(x) \Rightarrow \boxed{\nabla} \Rightarrow \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Gradient and Tangent Line:

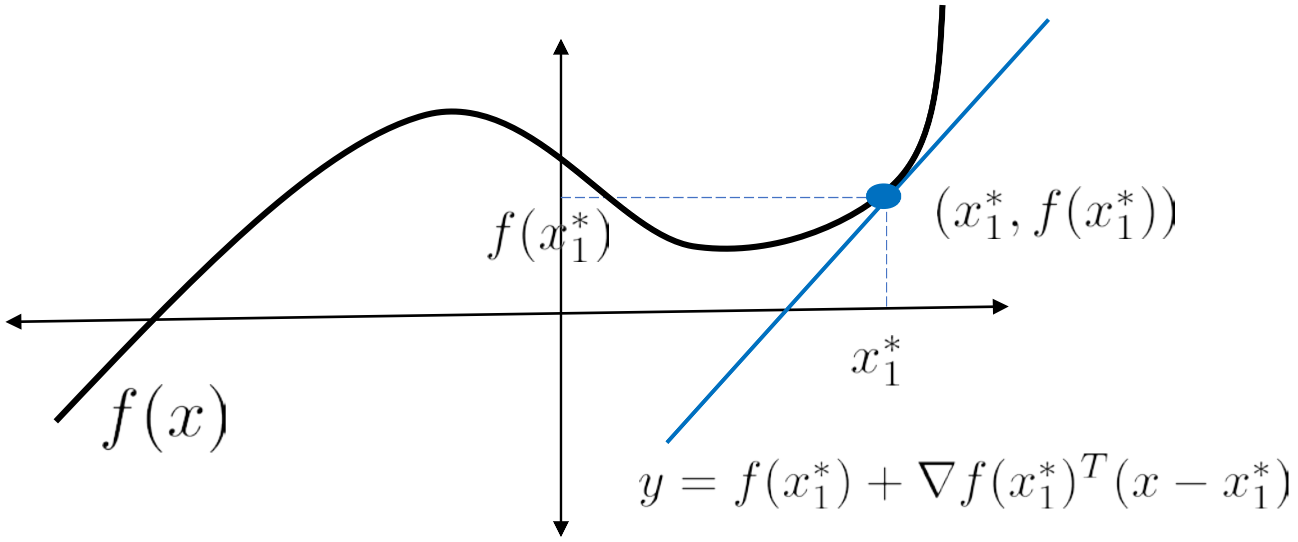
Tangent line (Hyperplane) at point $(x^*, f(x^*))$:



Gradient and Tangent Line:

Tangent line (Hyperplane) at point $(x^*, f(x^*))$:

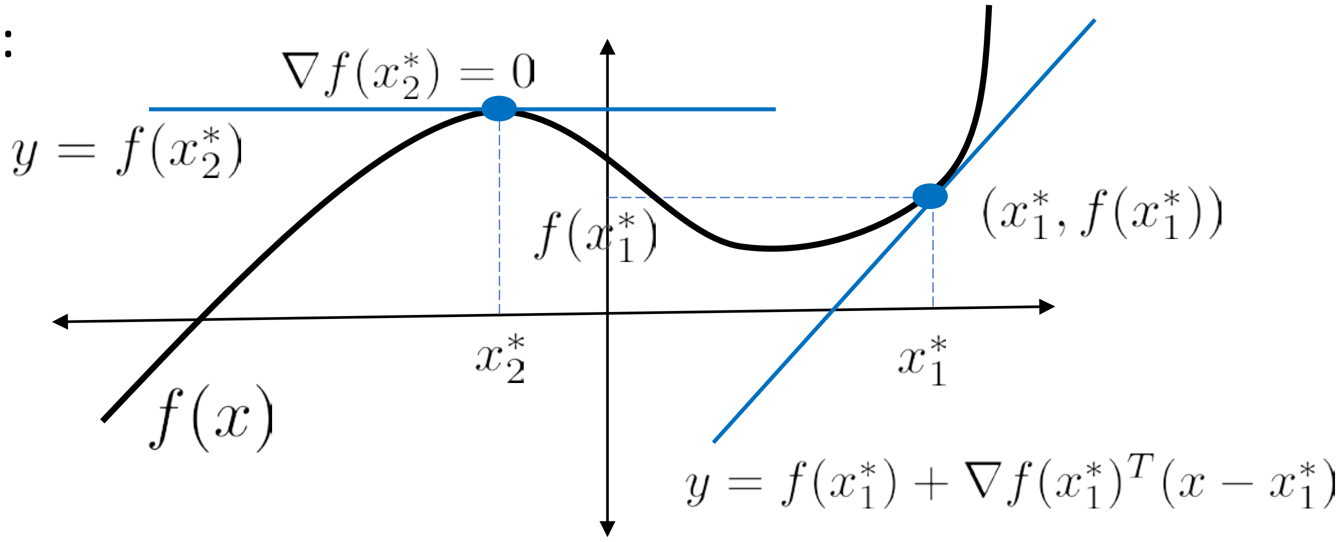
$$y = f(x^*) + \nabla f(x^*)^T (x - x^*)$$



Gradient and Tangent Line:

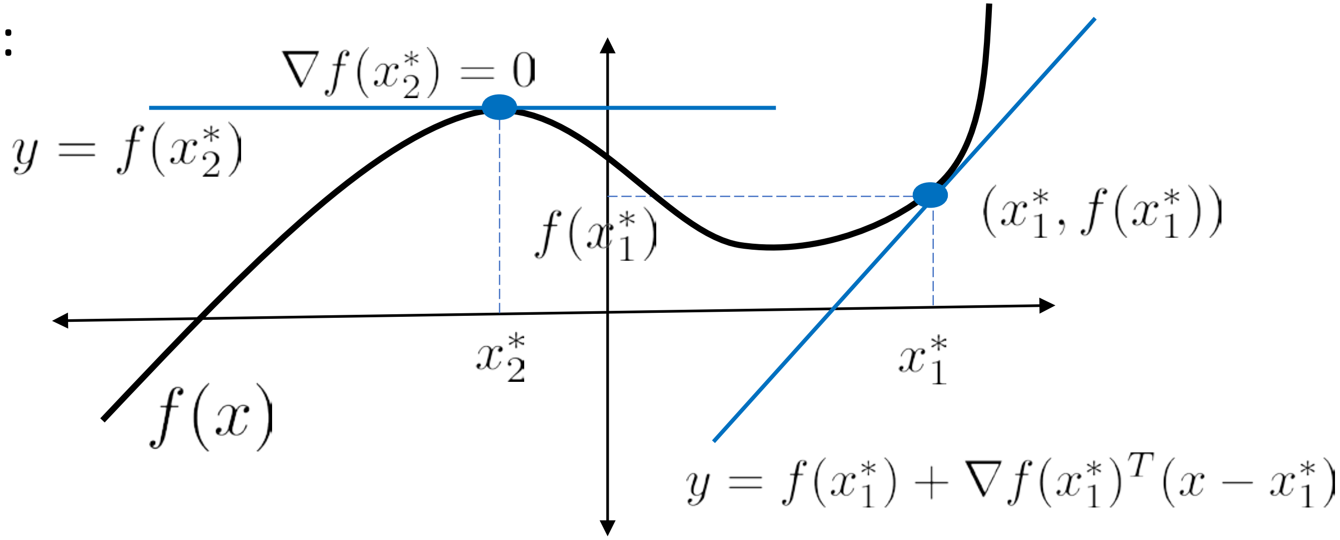
Tangent line (Hyperplane) at point $(x^*, f(x^*))$:

$$y = f(x^*) + \nabla f(x^*)^T (x - x^*)$$



Gradient and Tangent Line:

Tangent line (Hyperplane) at point $(x^*, f(x^*))$:



$$y = f(x^*) + \nabla f(x^*)^T (x - x^*)$$

$$\nabla f(x^*) = 0 \rightarrow x^* : \text{maximum/minimum}$$

Hessian

Hessian

Real-valued scalar function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

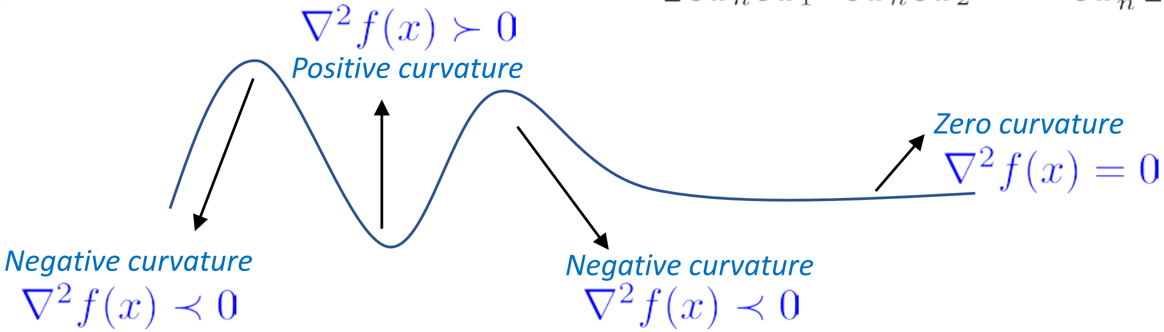
$$f(x) \xrightarrow{\textit{Gradient vector}} \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \xrightarrow{\textit{Hessian matrix}} \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian

Real-valued scalar function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(x) \xrightarrow{\textit{Gradient vector}} \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \xrightarrow{\textit{Hessian matrix}} \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian matrix : describes the local curvature of the function

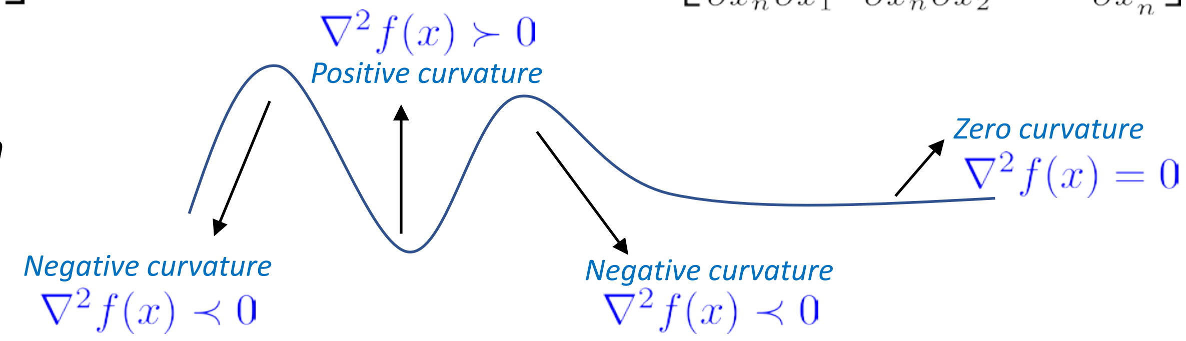


Hessian

Real-valued **scalar** function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(x) \xrightarrow{\textit{Gradient vector}} \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \xrightarrow{\textit{Hessian matrix}} \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian matrix : describes the local curvature of the function



Jacobian

Real-valued **vector** function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$

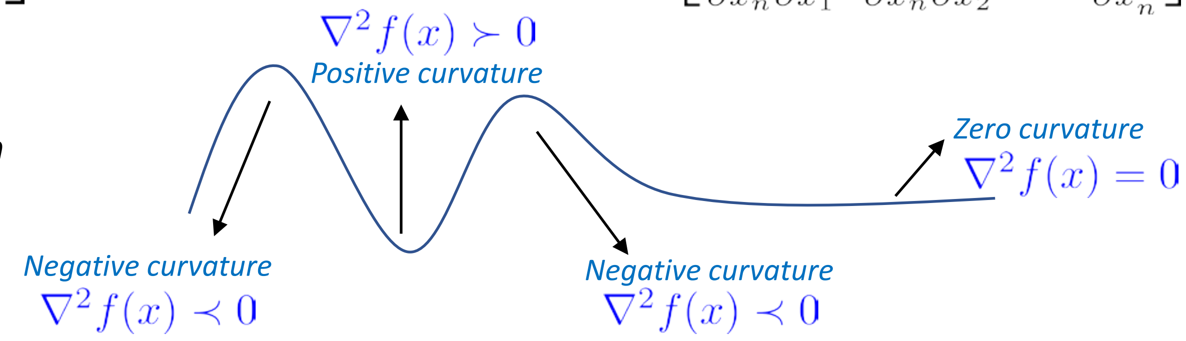
$$f = \begin{bmatrix} f_1(x) \\ \cdots \\ f_m(x) \end{bmatrix} \xrightarrow{\textit{Jacobian matrix}} \mathbf{J} = \nabla f(x) = \begin{bmatrix} [\nabla f_1(x)]^T \\ \vdots \\ [\nabla f_m(x)]^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Hessian

Real-valued **scalar** function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$

$$f(x) \xrightarrow{\text{Gradient vector}} \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \cdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \xrightarrow{\text{Hessian matrix}} \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Hessian matrix : describes the local curvature of the function



Jacobian

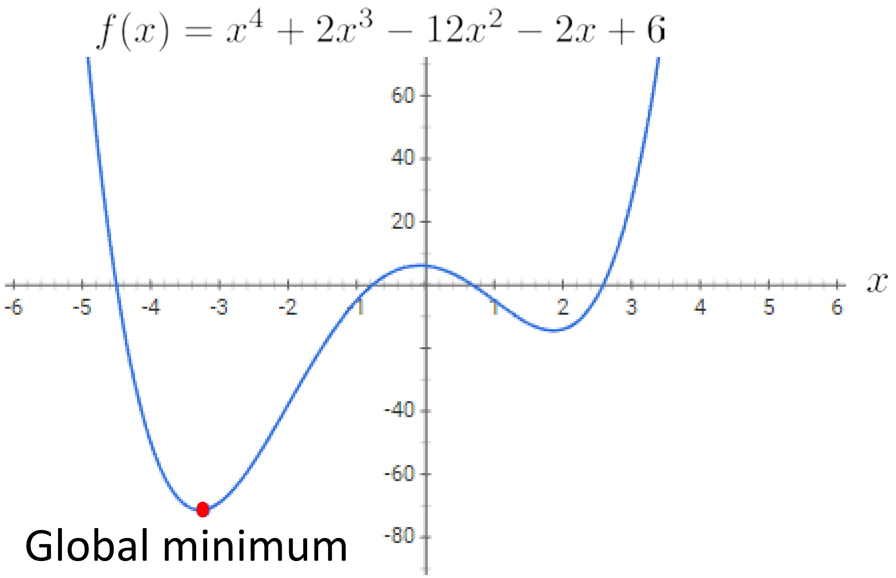
Real-valued **vector** function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$f = \begin{bmatrix} f_1(x) \\ \cdots \\ f_m(x) \end{bmatrix} \xrightarrow{\text{Jacobian matrix}} \mathbf{J} = \nabla f(x) = \begin{bmatrix} [\nabla f_1(x)]^T \\ \vdots \\ [\nabla f_m(x)]^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Hessian matrix : Jacobian matrix of gradient vector.

Global minimum x^*:

$$\underbrace{f(x^*) \leq f(x) \quad \forall x \in \mathbb{R}^n}_{\text{(Global)}}$$

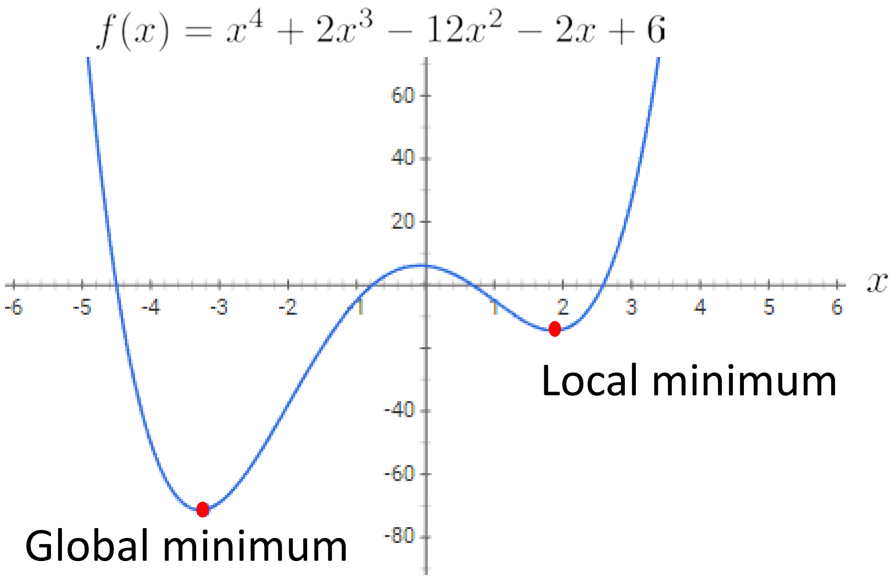


Global minimum x^*:

$$f(x^*) \leq f(x) \quad \underbrace{\forall x \in \mathbb{R}^n}_{\text{(Global)}}$$

Local minimum x^*:

$$f(x^*) \leq f(x) \quad \underbrace{\forall x \in B_{r>0}(x^*)}_{\substack{\text{Neighborhood (ball) around } x^* \\ \text{(Local)}}}$$



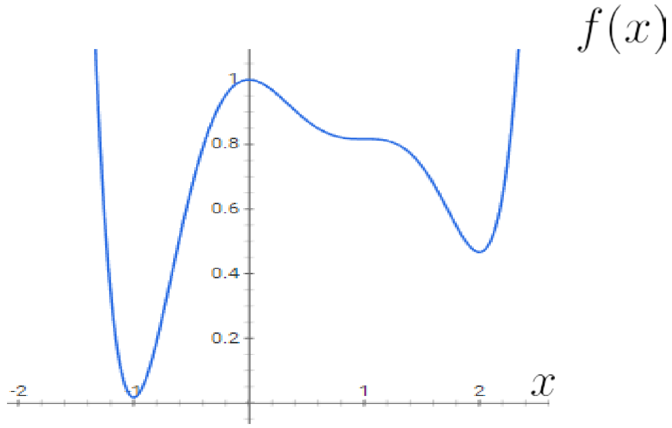
Optimality Conditions

- Unconstrained Optimization
- Constrained Optimization

1) Optimality Conditions: Unconstrained Optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

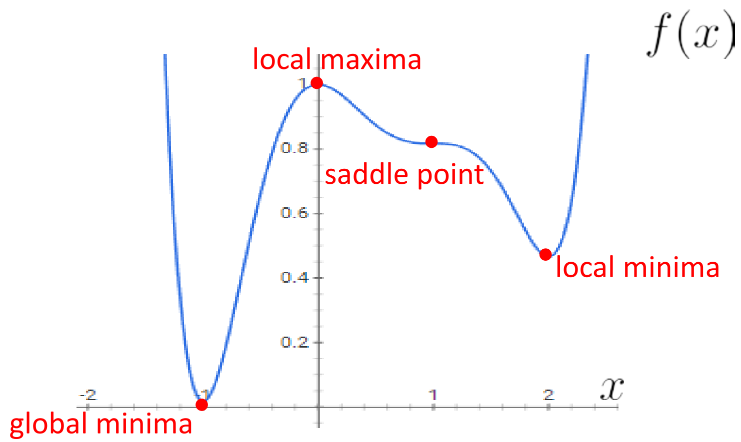
What are the conditions for x to be a minimum point ?



1) Optimality Conditions: Unconstrained Optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\nabla f(x^*) = 0$$

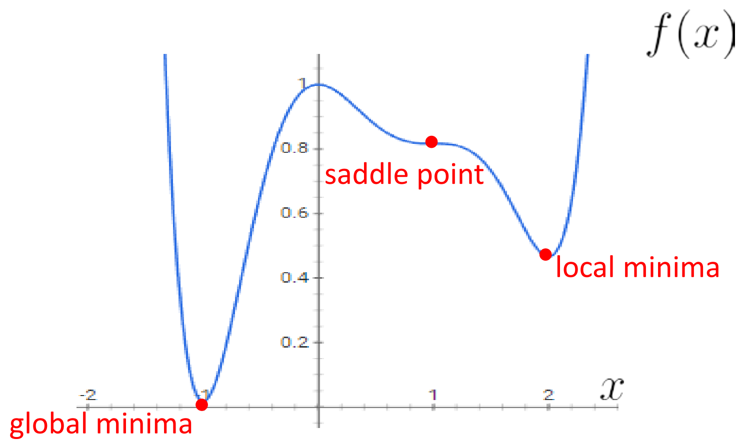


1) Optimality Conditions: Unconstrained Optimization

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\nabla f(x^*) = 0$$

$$\nabla^2 f(x^*) \succeq 0$$



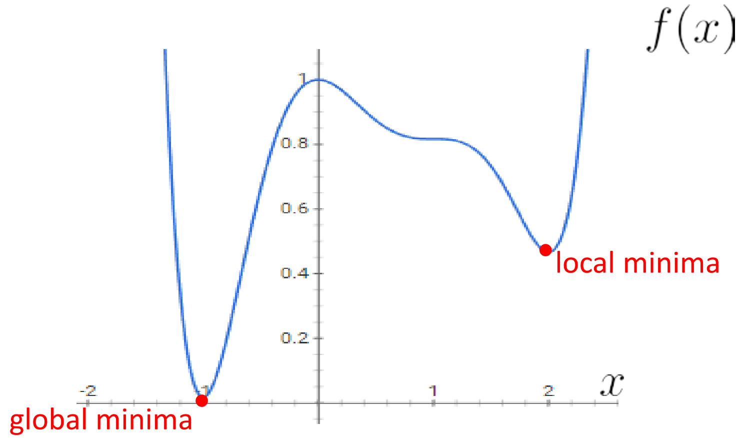
1) Optimality Conditions: Unconstrained Optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\nabla f(x^*) = 0$$

$$\nabla^2 \cancel{f(x^*)} \succeq 0$$

$$\nabla^2 f(x^*) \succ 0$$



2) Optimality Conditions: Unconstrained Optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Necessary Condition: $\nabla f(x) = 0, \quad \nabla^2 f(x) \succeq 0$

Sufficient Condition: $\nabla f(x) = 0, \quad \nabla^2 f(x) \succ 0$

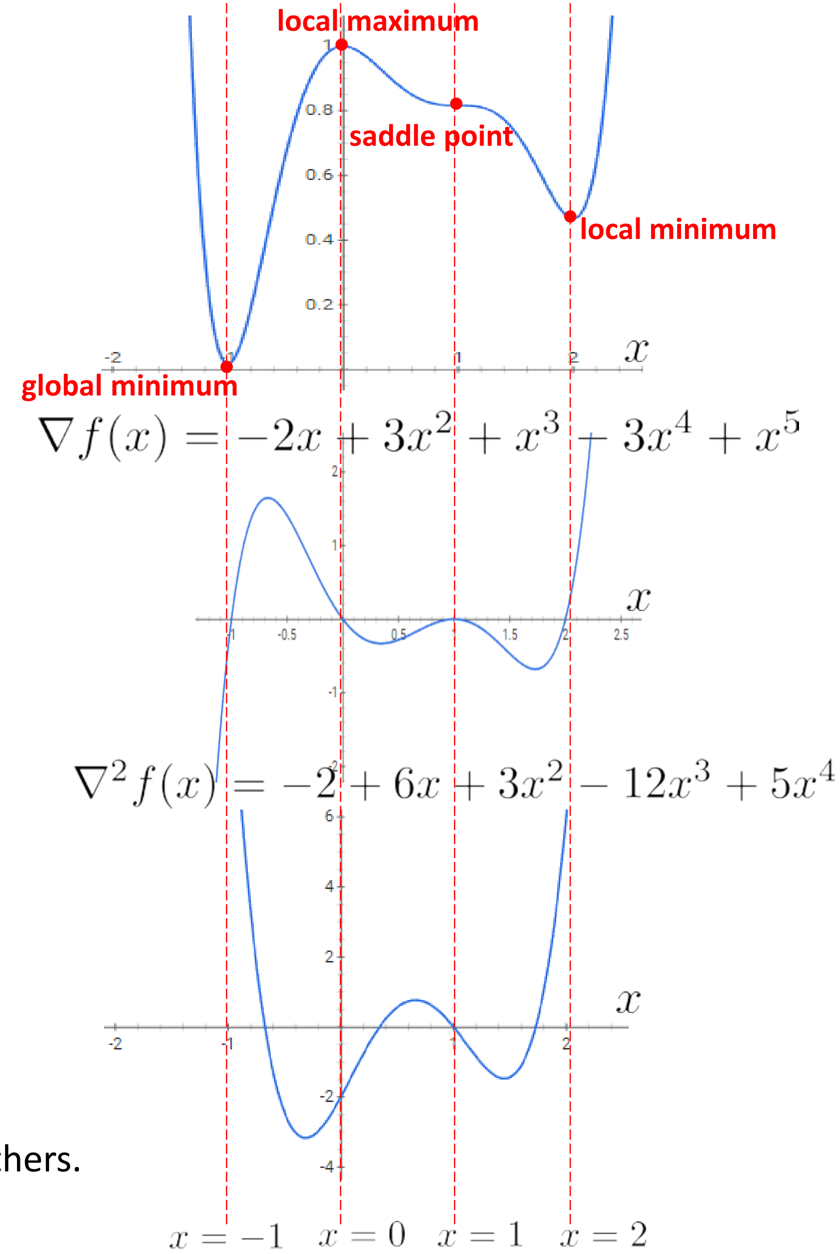
First order optimality condition

Second order optimality condition

$f(x) = 1 - x^2 + x^3 + \frac{1}{4}x^4 - \frac{3}{5}x^5 + \frac{1}{6}x^6$

$\nabla f(x) = -2x + 3x^2 + x^3 - 3x^4 + x^5$

$\nabla^2 f(x) = -2 + 6x + 3x^2 - 12x^3 + 5x^4$



Necessary Condition: $x = -1, 2, 1$

Saddle point:
maximum in some directions, but minimum in others.

Sufficient Condition: $x = -1, 2$

Local and global minima

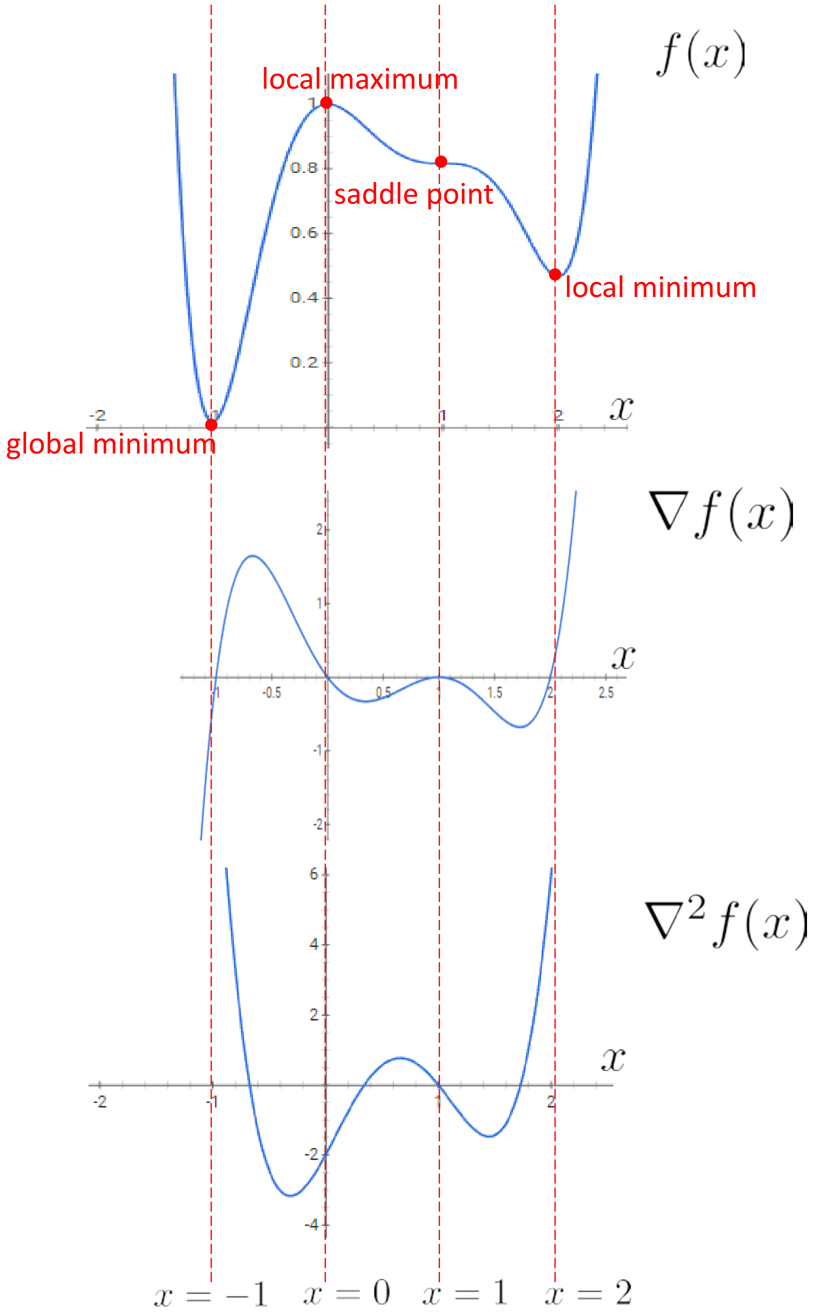
2) Optimality Conditions: Unconstrained Optimization

Unconstrained Optimization: $\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$

Necessary Condition: $\nabla f(x) = 0, \quad \nabla^2 f(x) \succeq 0$

Sufficient Condition: $\nabla f(x) = 0, \quad \nabla^2 f(x) \succ 0$

- To find x^* (argmin $f(x)$), Find the *roots* of $\nabla f(x^*)$.
- If obtained root satisfies $\nabla^2 f(x^*) \succ 0$, point x^* is (local/global) minimum.

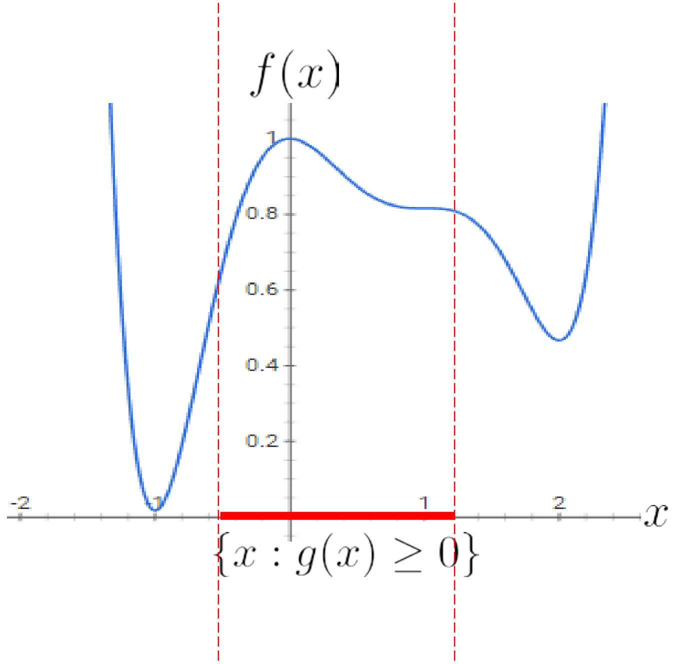


Optimality Conditions for Constrained Optimization

Optimality Conditions: Constrained Optimization

Constrained Optimization:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \geq 0 \end{array}$$

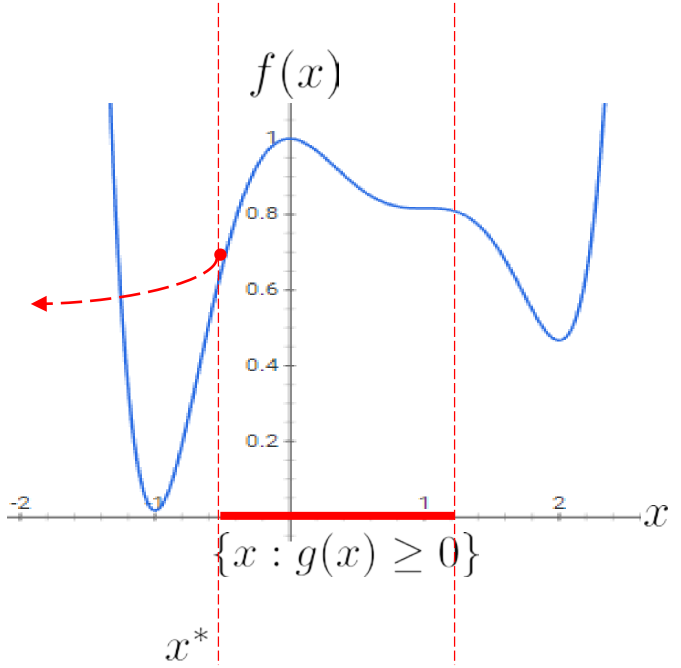


Optimality Conditions: Constrained Optimization

Constrained Optimization:

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \geq 0 \end{array}$$

minimum of constrained optimization



$$\nabla f(x^*) \neq 0$$

- Optimality conditions of *unconstrained optimization* are *not valid* for *constrained optimization*.

Optimization:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0,\ \ i = 1, ..., n_g \\ & h_i(x) = 0,\ \ i = 1, .., n_h \end{aligned}$$

KKT (Karush-Kuhn-Tucker) Necessary Optimality Condition:

Lagrange function $L(x, \mu, \lambda) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$

Lagrange multiplier

Optimality Cond.

$\nabla_x L(x, \mu, \lambda) = 0$ — *Stationarity*

$(\nabla_{\lambda_i} L(x, \lambda) = 0)$ $-h_i(x^*) = 0,\ \ i = 1, ..., n_h$

$g_i(x^*) \geq 0, \quad i = 1, ..., n_g$ — *Primal Feasibility*

$\mu_i \geq 0$ — *Dual Feasibility*

$\mu_i^* g_i(x^*) = 0, \quad i = 1, ..., n_g$ — *Dual Complementary Slackness*

KKT (Karush-Kuhn-Tucker) Optimality Condition

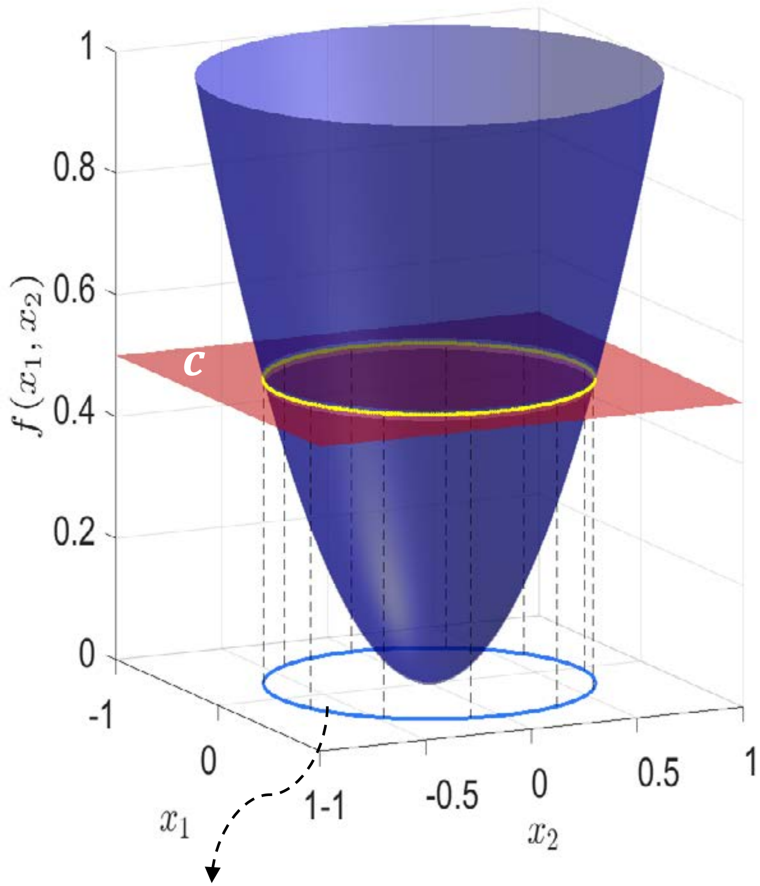
- Geometrical Interpretation

Basic Definitions

- Level Set
- Level Set and Gradient Vector
- Tangent Level Sets

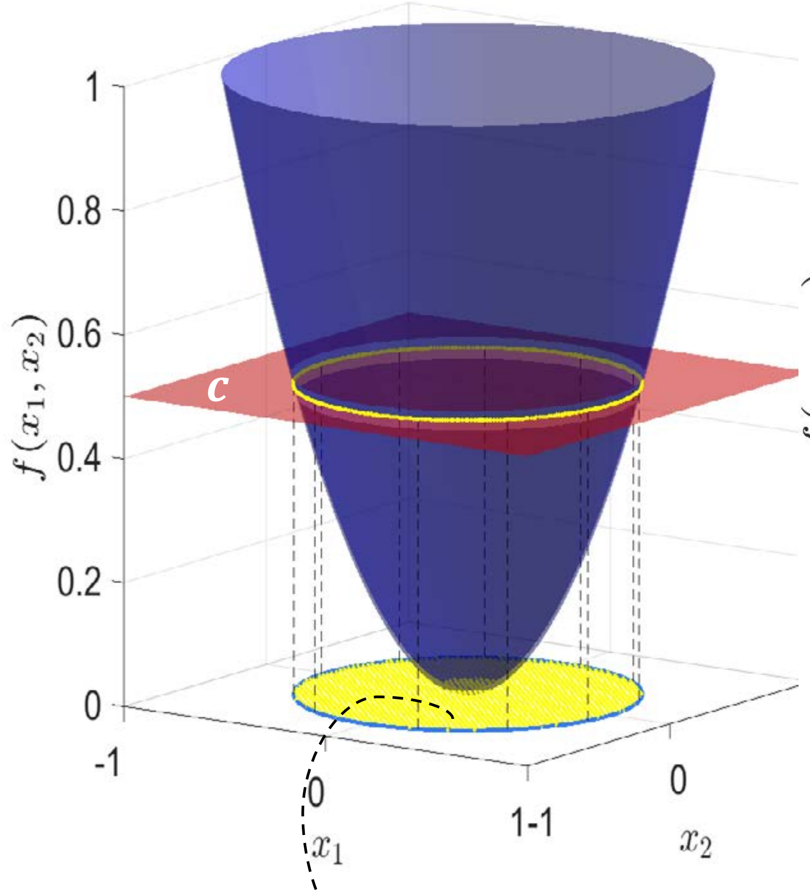
Level, Sub-level, and Super-level Sets

- For a given function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ and $c \in \mathbb{R}$, we define:



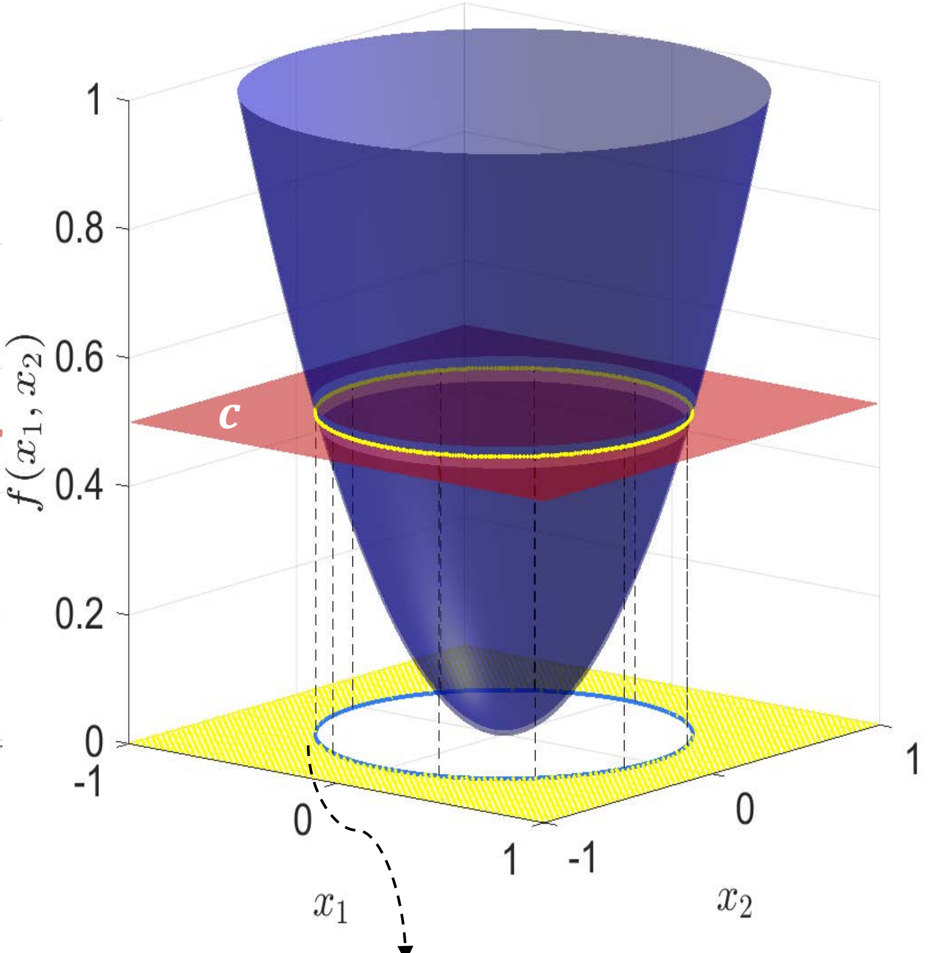
Level Set (Contour):

$$\{x \in \mathbb{R}^n : f(x) = c\}$$



Sub-Level Set :

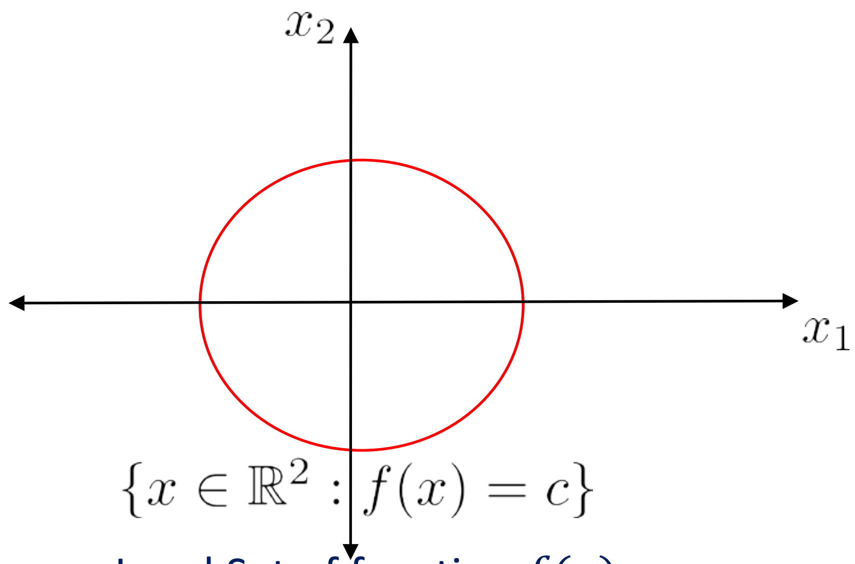
$$\{x \in \mathbb{R}^n : f(x) \leq c\}$$



Supper-Level Set :

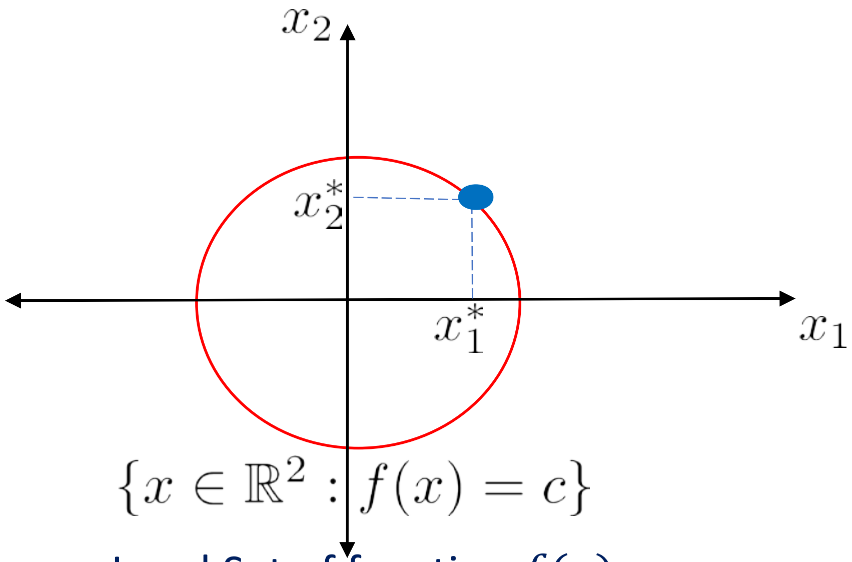
$$\{x \in \mathbb{R}^n : f(x) \geq c\}$$

Gradient and Level Sets



Level Set of function $f(x)$

Gradient and Level Sets

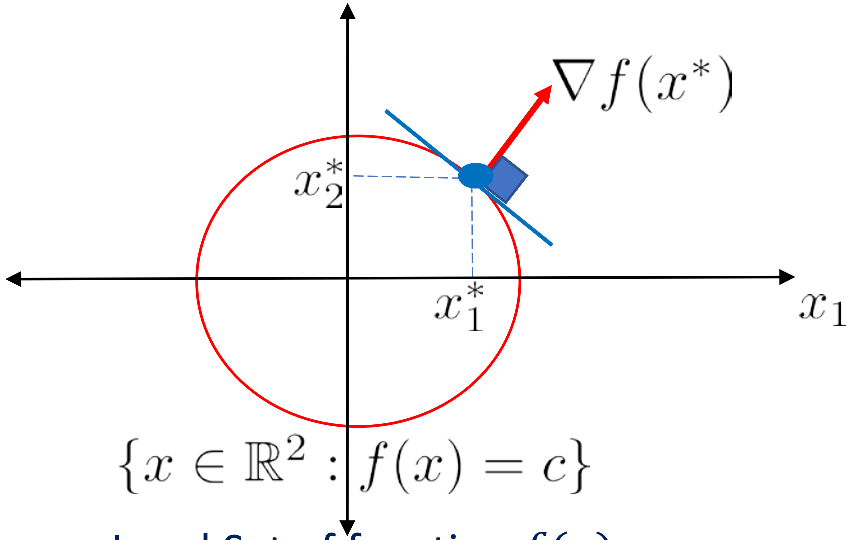


$\{x \in \mathbb{R}^2 : f(x) = c\}$

Level Set of function $f(x)$

Gradient and Level Sets

$\nabla f(x^*)$ is perpendicular vector to the level set.

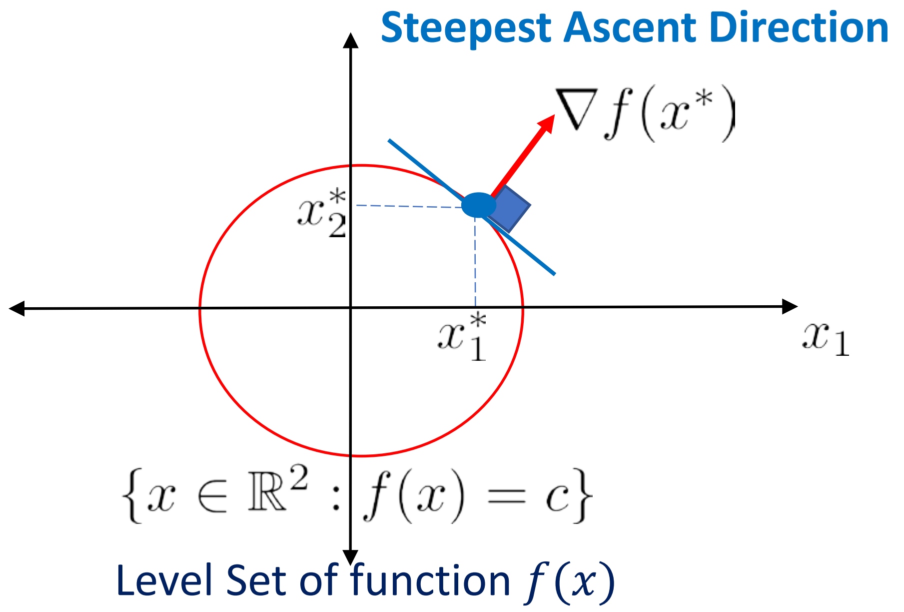


Level Set of function $f(x)$

Gradient and Level Sets

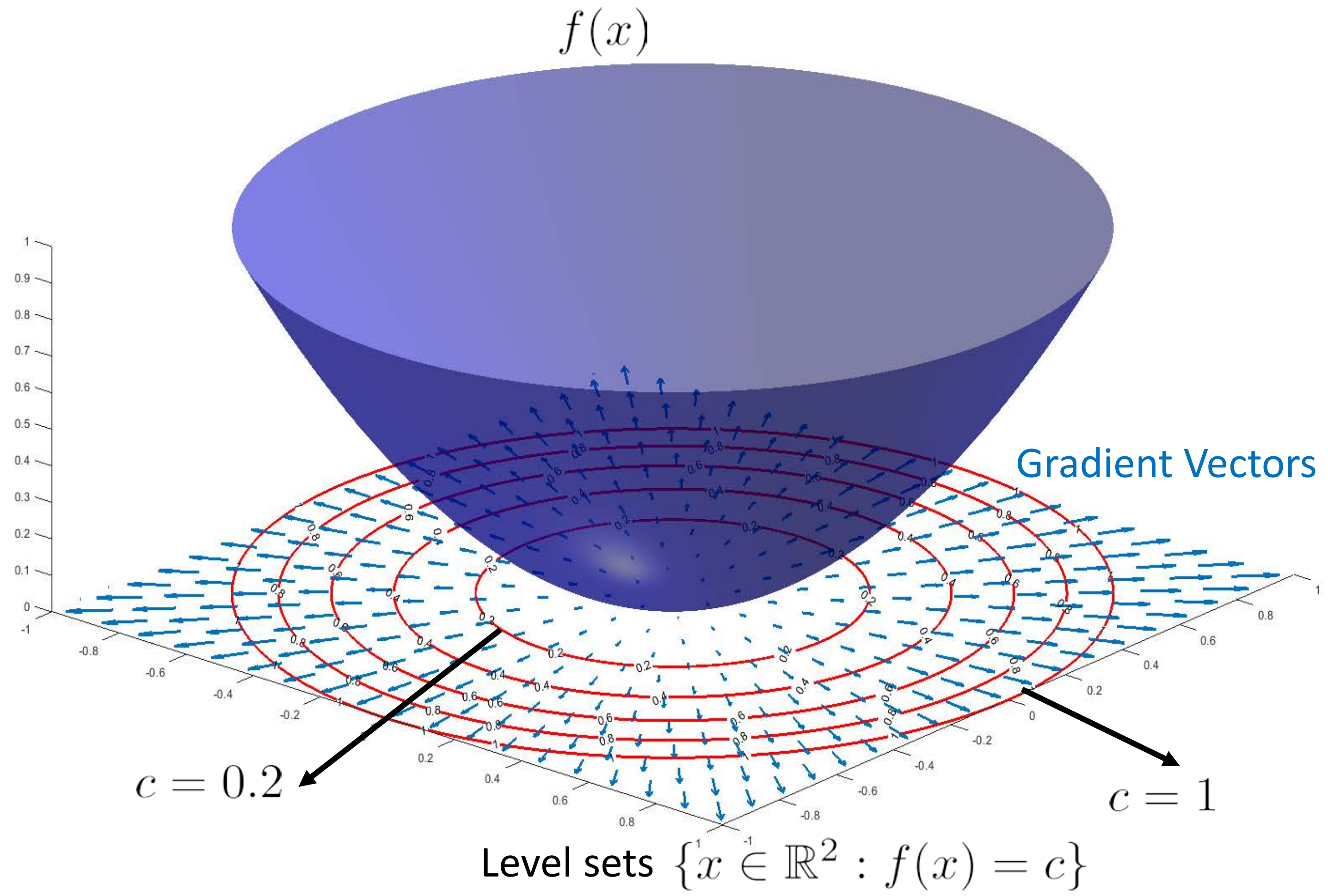
$\nabla f(x^*)$ is perpendicular vector to the level set.

$\nabla f(x^*)$ shows the direction of the steepest ascent.



Example:

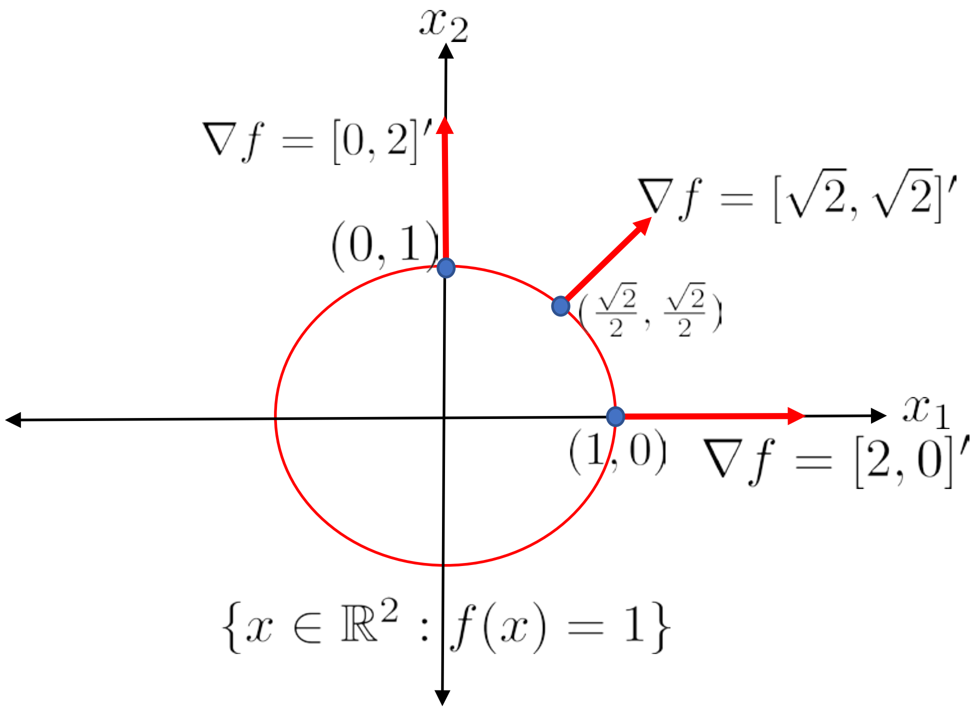
$$f(x) = x_1^2 + x_2^2 \quad \Rightarrow \quad \textbf{Gradient vector :} \ \nabla f = [2x_1, 2x_2]^T$$

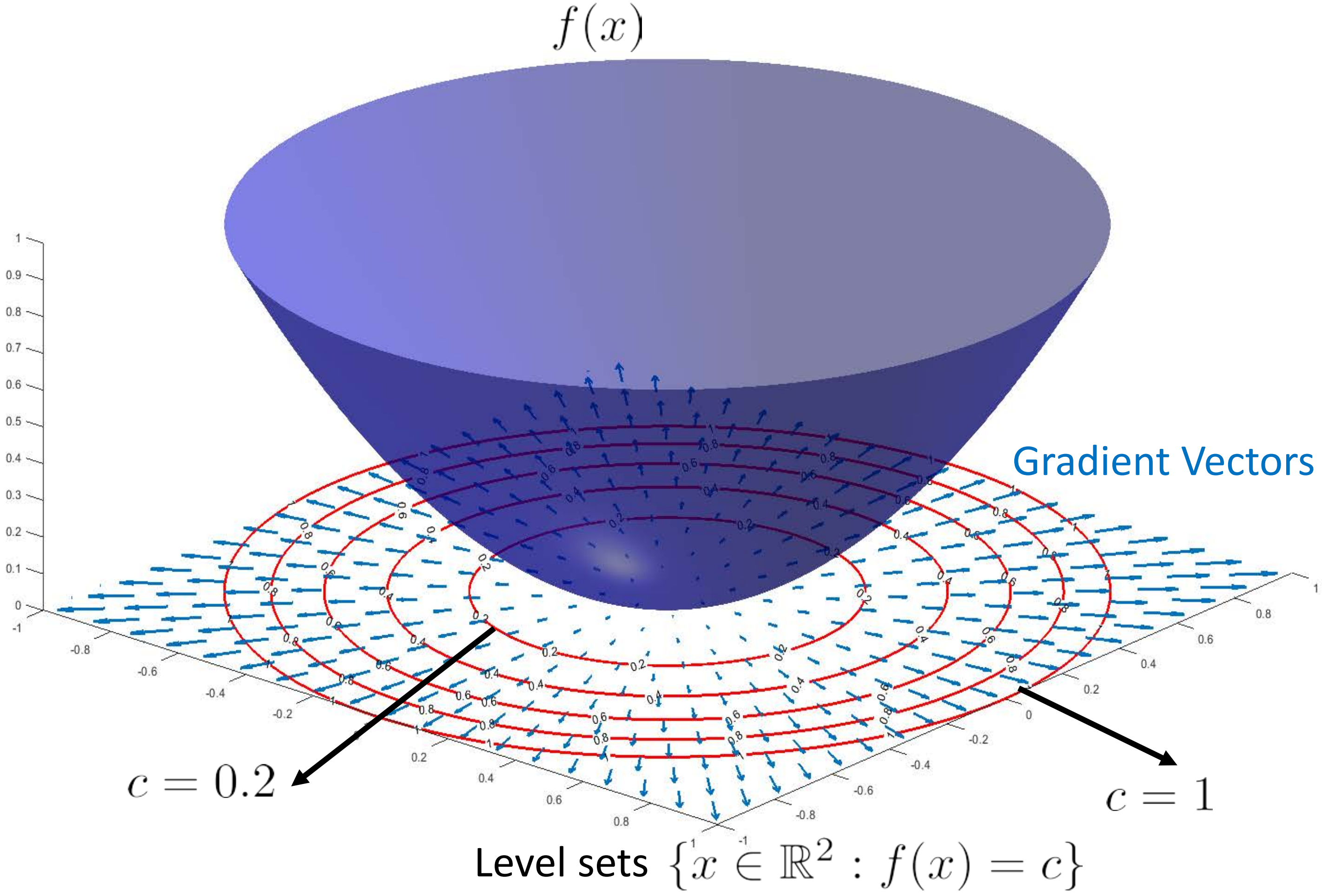


Example:

$$f(x) = x_1^2 + x_2^2$$

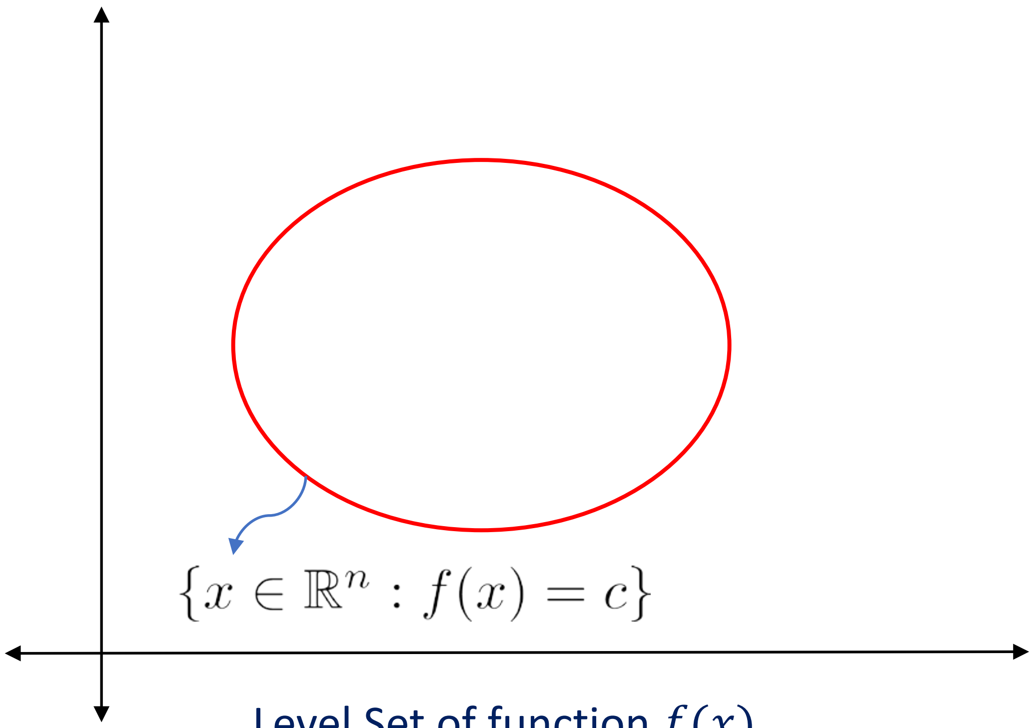
Gradient vector : $\nabla f = [2x_1, 2x_2]^T$





Example:

- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

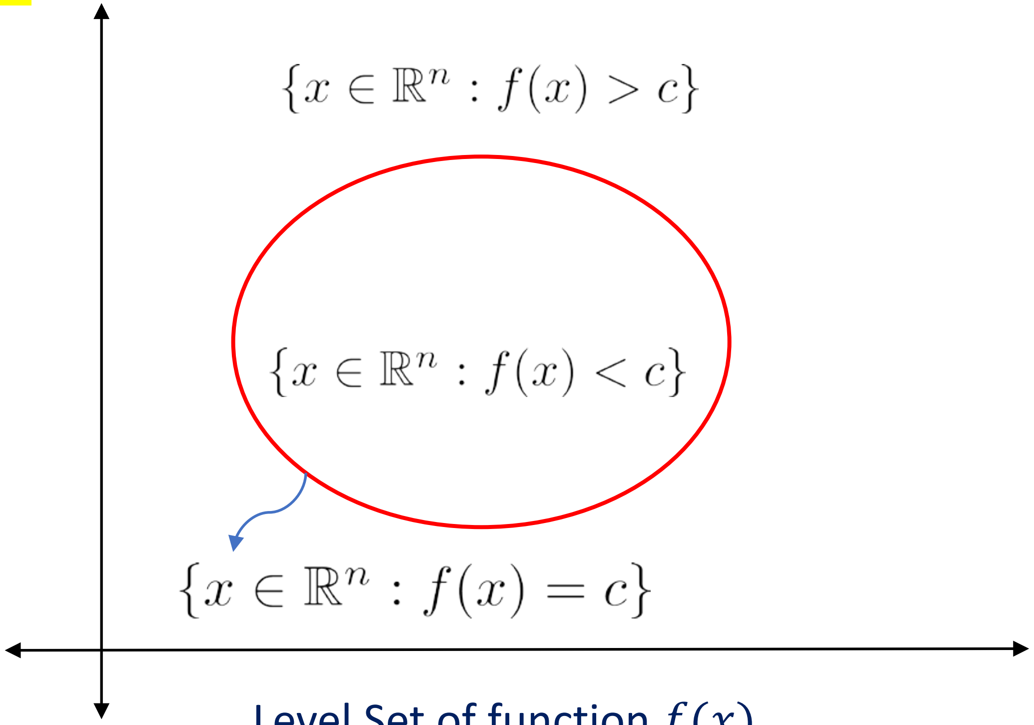


Level Set of function $f(x)$

Example:

- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1

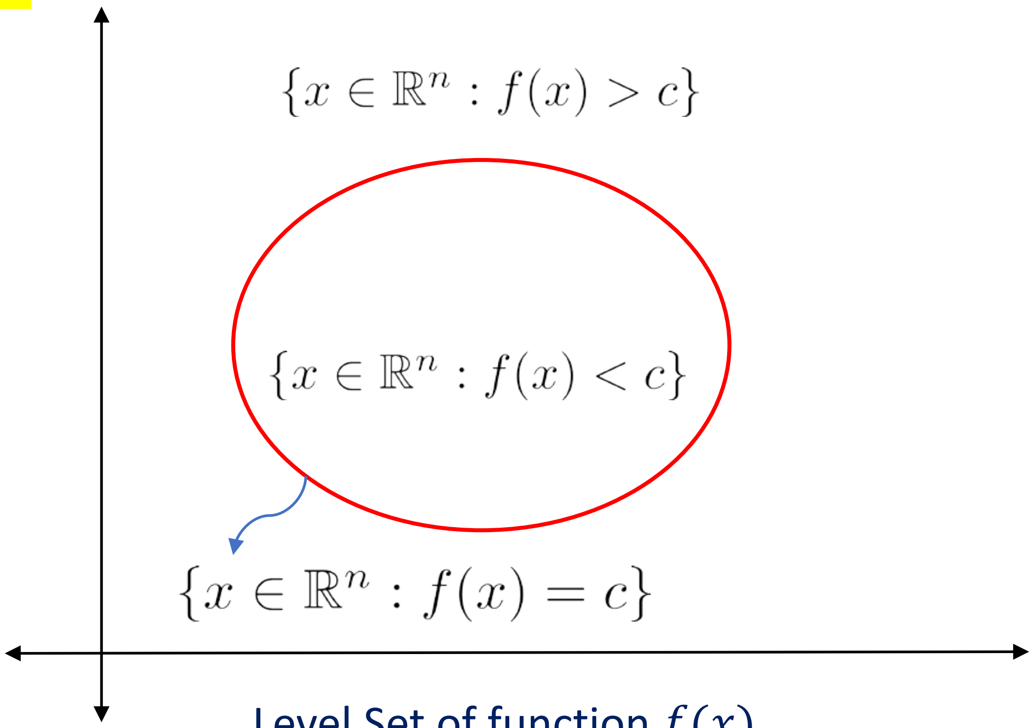


Level Set of function $f(x)$

Example:

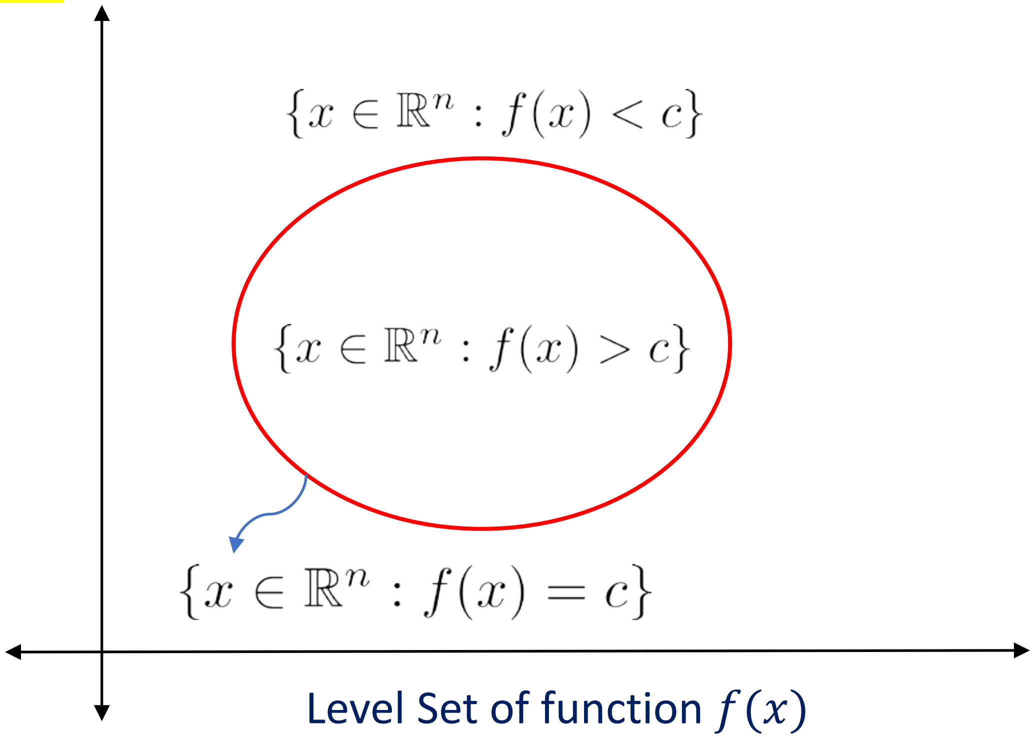
- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1



Level Set of function $f(x)$

Case 2

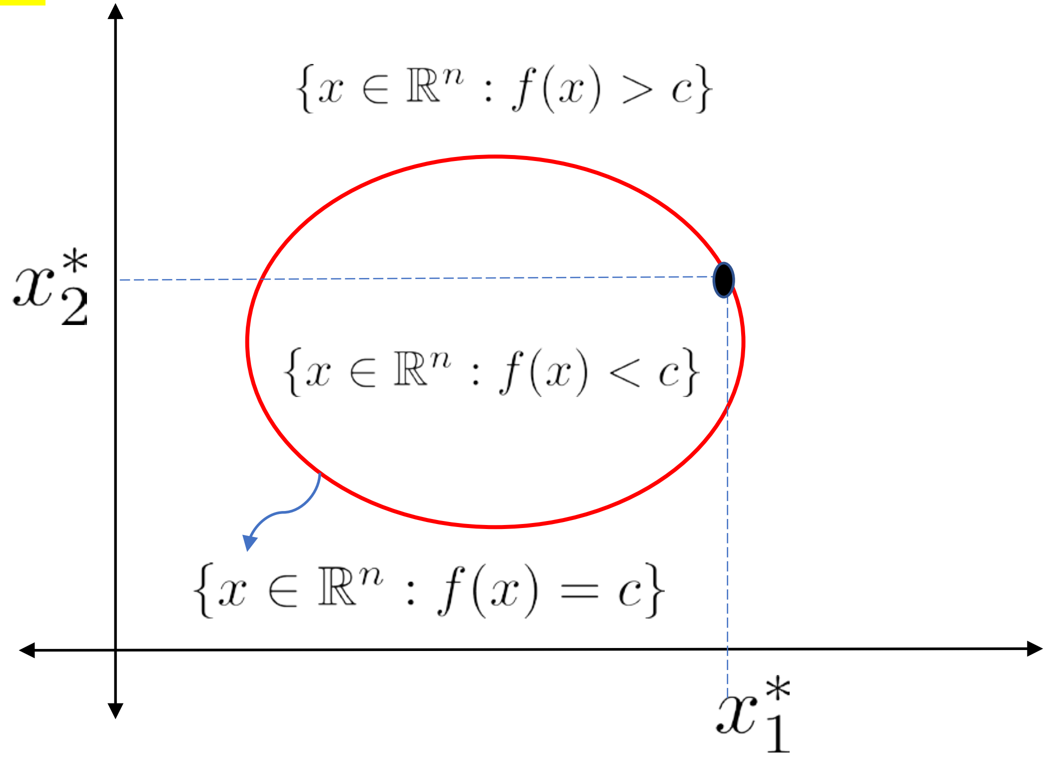


Level Set of function $f(x)$

Example:

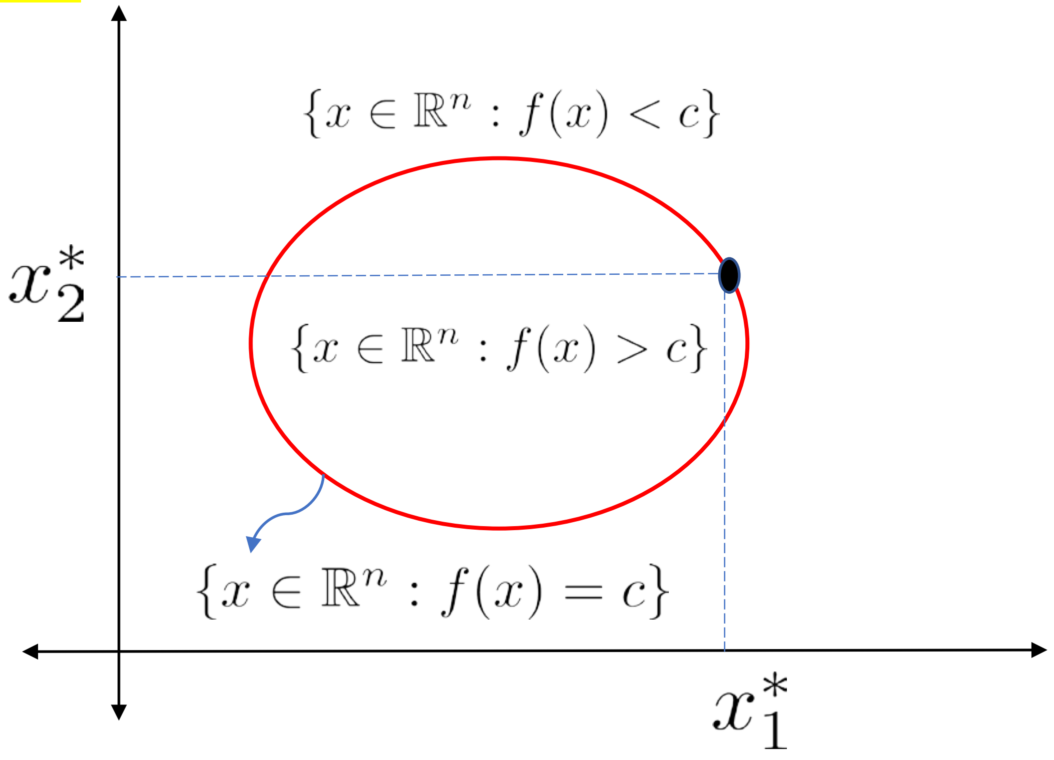
- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1



Level Set of function $f(x)$

Case 2

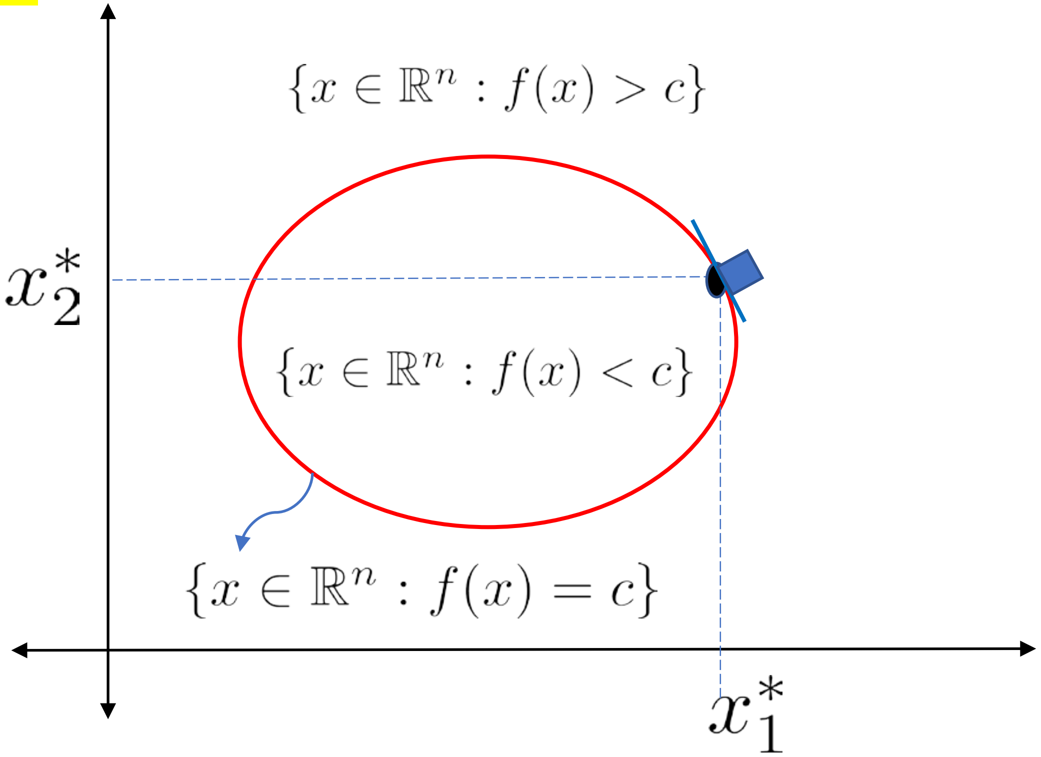


Level Set of function $f(x)$

Example:

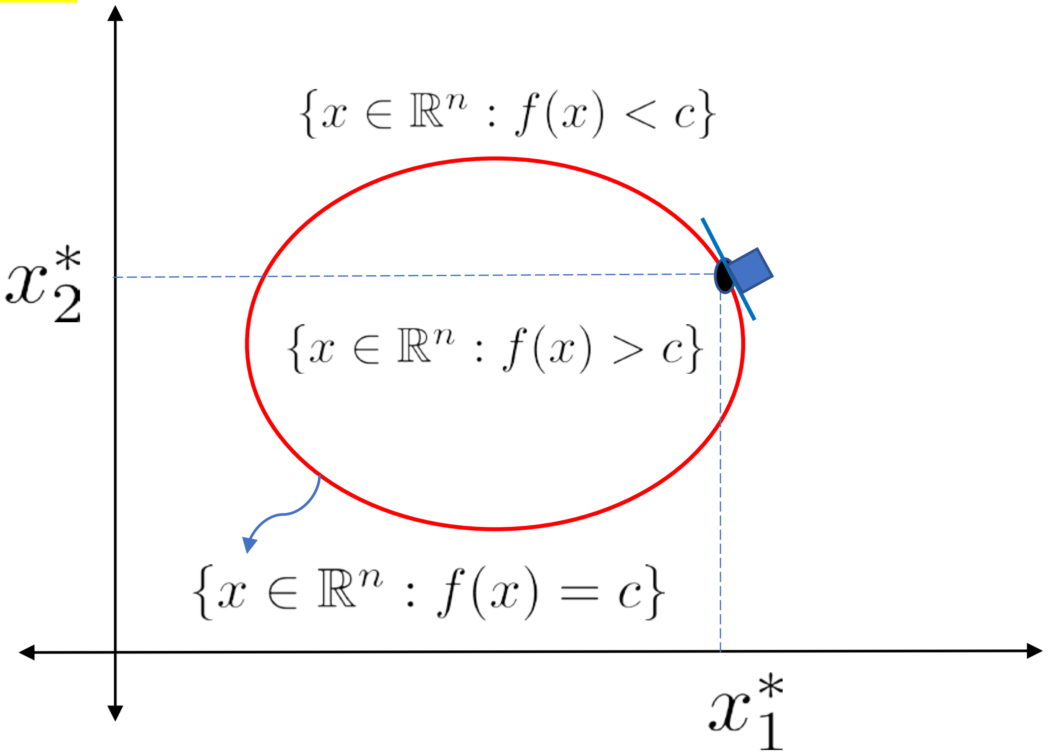
- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1



Level Set of function $f(x)$

Case 2

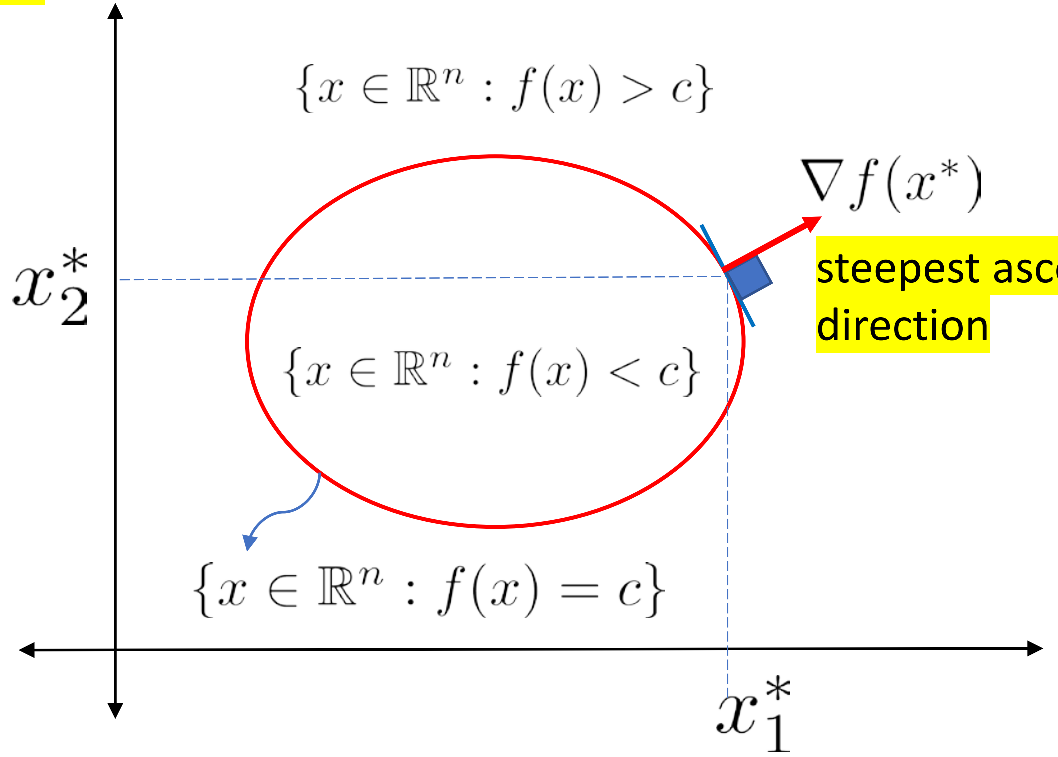


Level Set of function $f(x)$

Example:

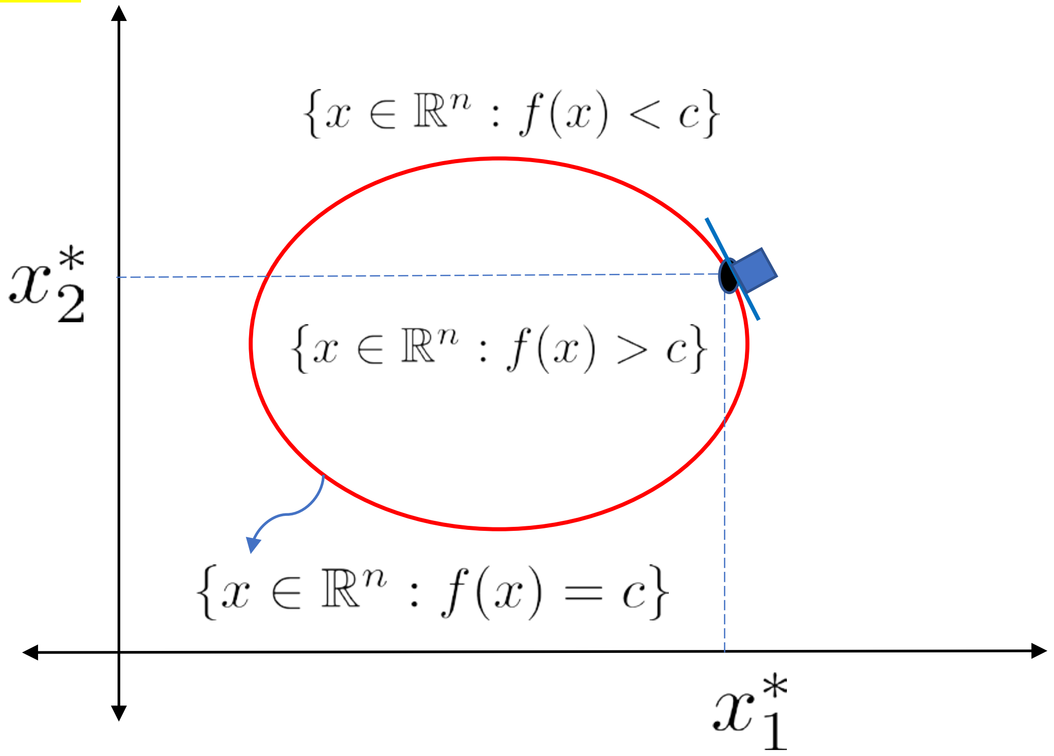
- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1



Level Set of function $f(x)$

Case 2

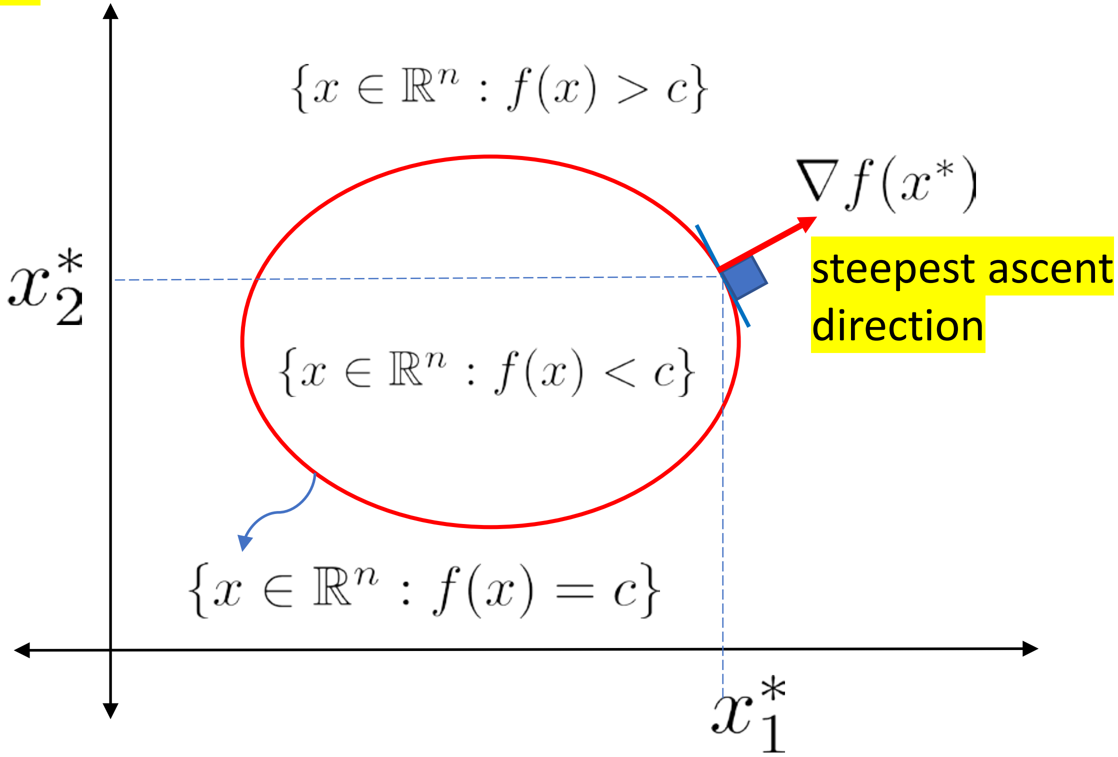


Level Set of function $f(x)$

Example:

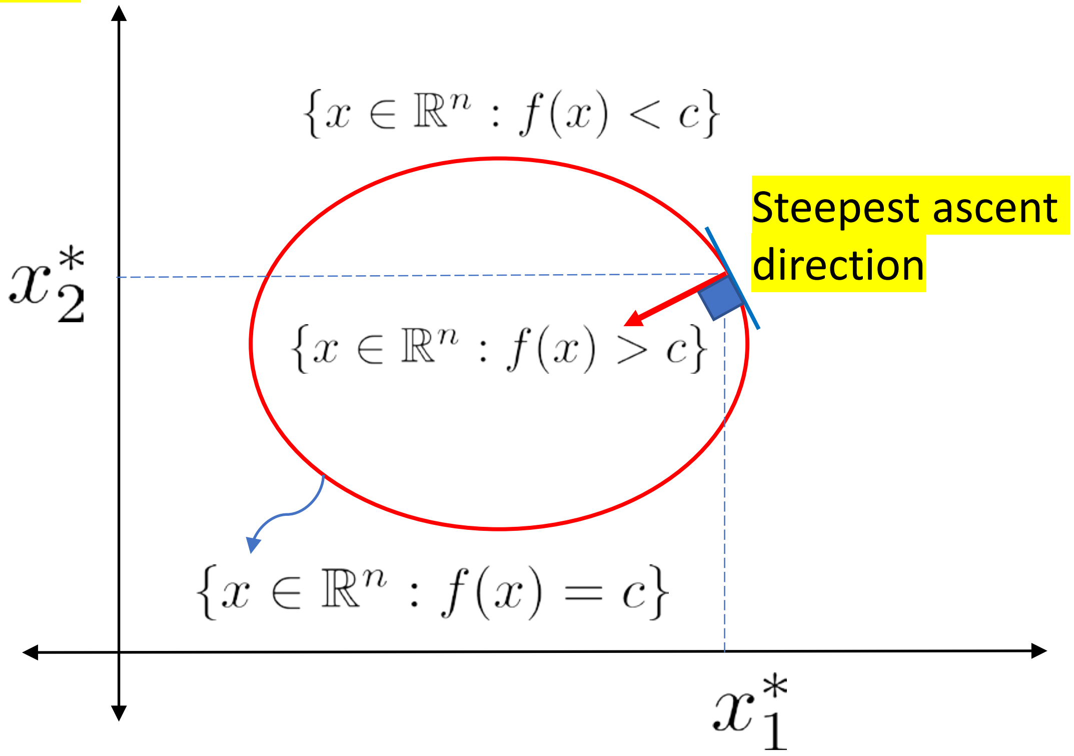
- $\nabla f(x^*)$ shows the direction of the steepest ascent.
- $\nabla f(x^*)$ is perpendicular vector to the level set.

Case 1



Level Set of function $f(x)$

Case 2



Level Set of function $f(x)$

Tangent Level sets

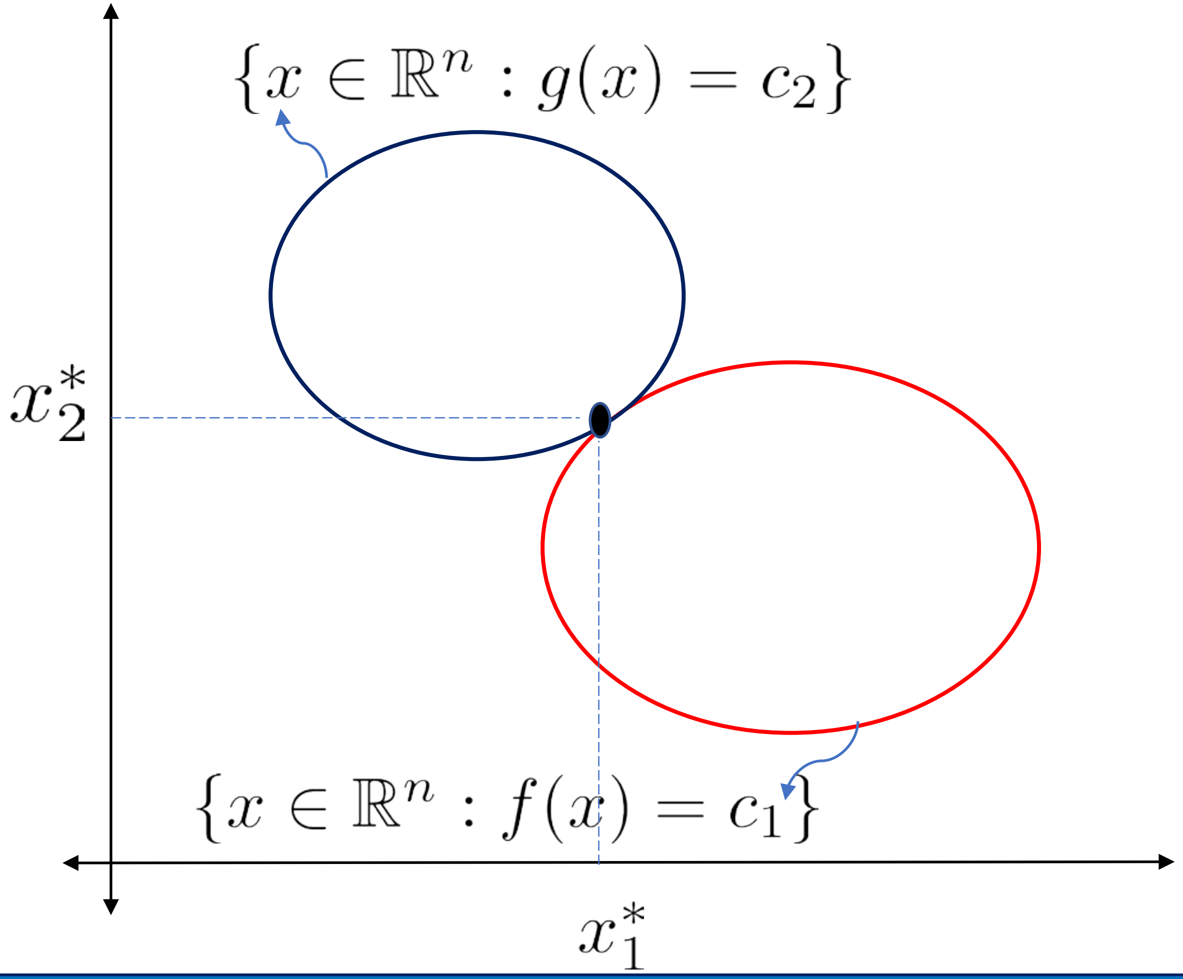
Tangent Level sets

- Function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g(x) : \mathbb{R}^n \to \mathbb{R}$

- Assume the level sets

$\{x \in \mathbb{R}^n : f(x) = c_1\}$ and $\{x \in \mathbb{R}^n : g(x) = c_2\}$

are tangent at the point x^*.



What is the relationship of the gradient vectors of $f(x)$ and $g(x)$ at point x^*?

Tangent Level sets

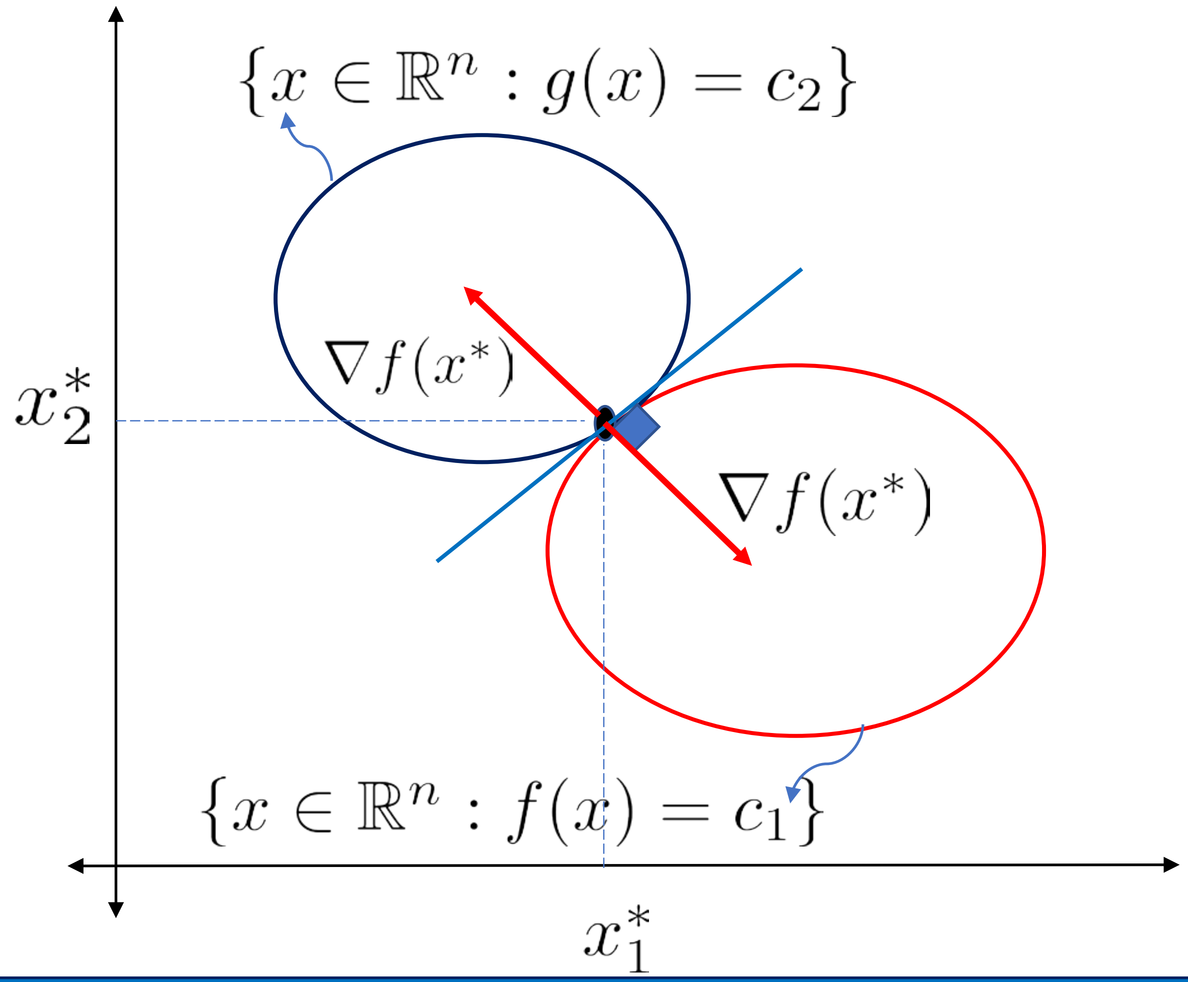
- Function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g(x) : \mathbb{R}^n \to \mathbb{R}$

- Assume the level sets

$$\{x \in \mathbb{R}^n : f(x) = c_1\} \text{ and } \{x \in \mathbb{R}^n : g(x) = c_2\}$$

are tangent at the point x^*.

What is the relationship of the gradient vectors of $f(x)$ and $g(x)$ at point x^*?



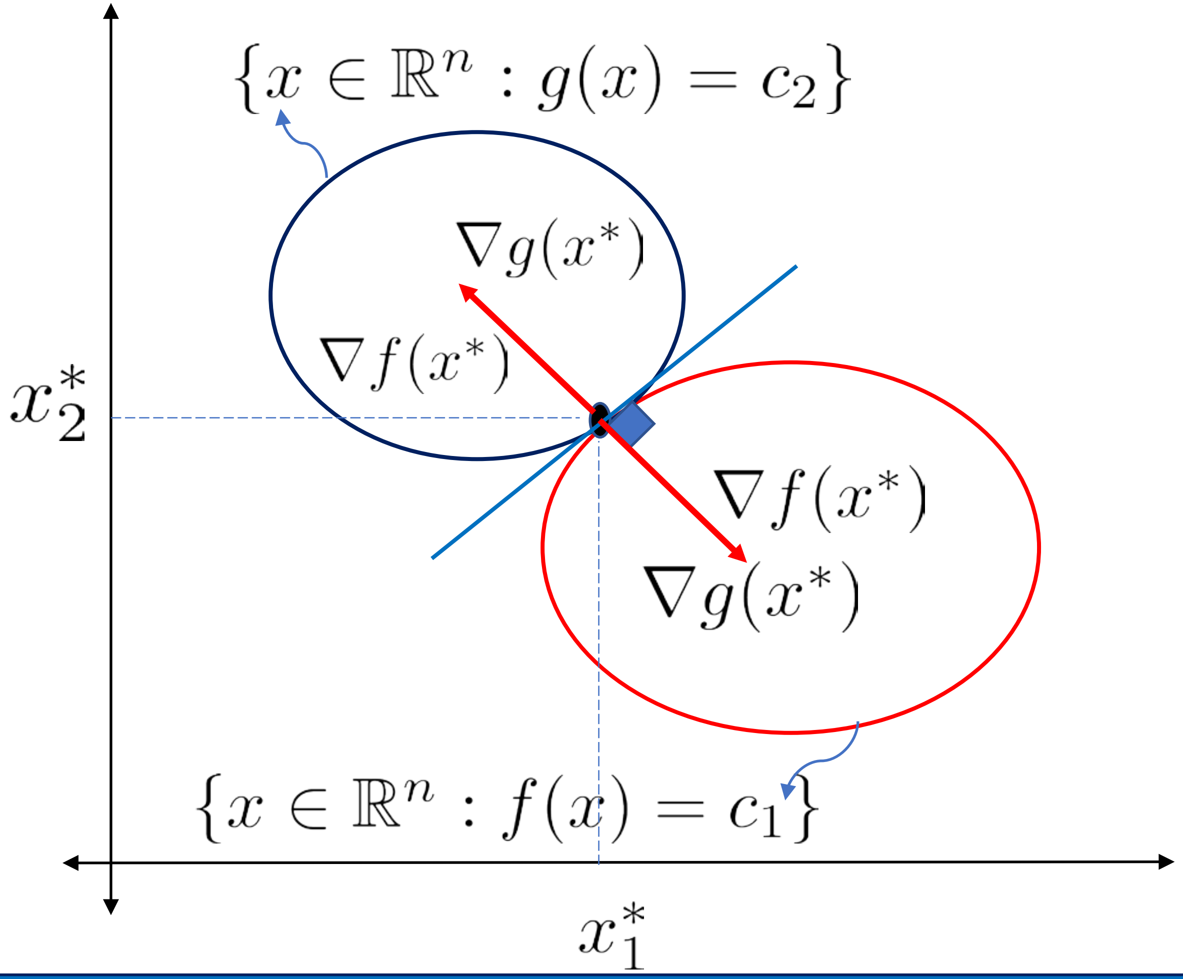
Tangent Level sets

- Function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g(x) : \mathbb{R}^n \to \mathbb{R}$

- Assume the level sets

$$\{x \in \mathbb{R}^n : f(x) = c_1\} \text{ and } \{x \in \mathbb{R}^n : g(x) = c_2\}$$

are tangent at the point x^*.



What is the relationship of the gradient vectors of $f(x)$ and $g(x)$ at point x^*?

Tangent Level sets

- Function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g(x) : \mathbb{R}^n \to \mathbb{R}$

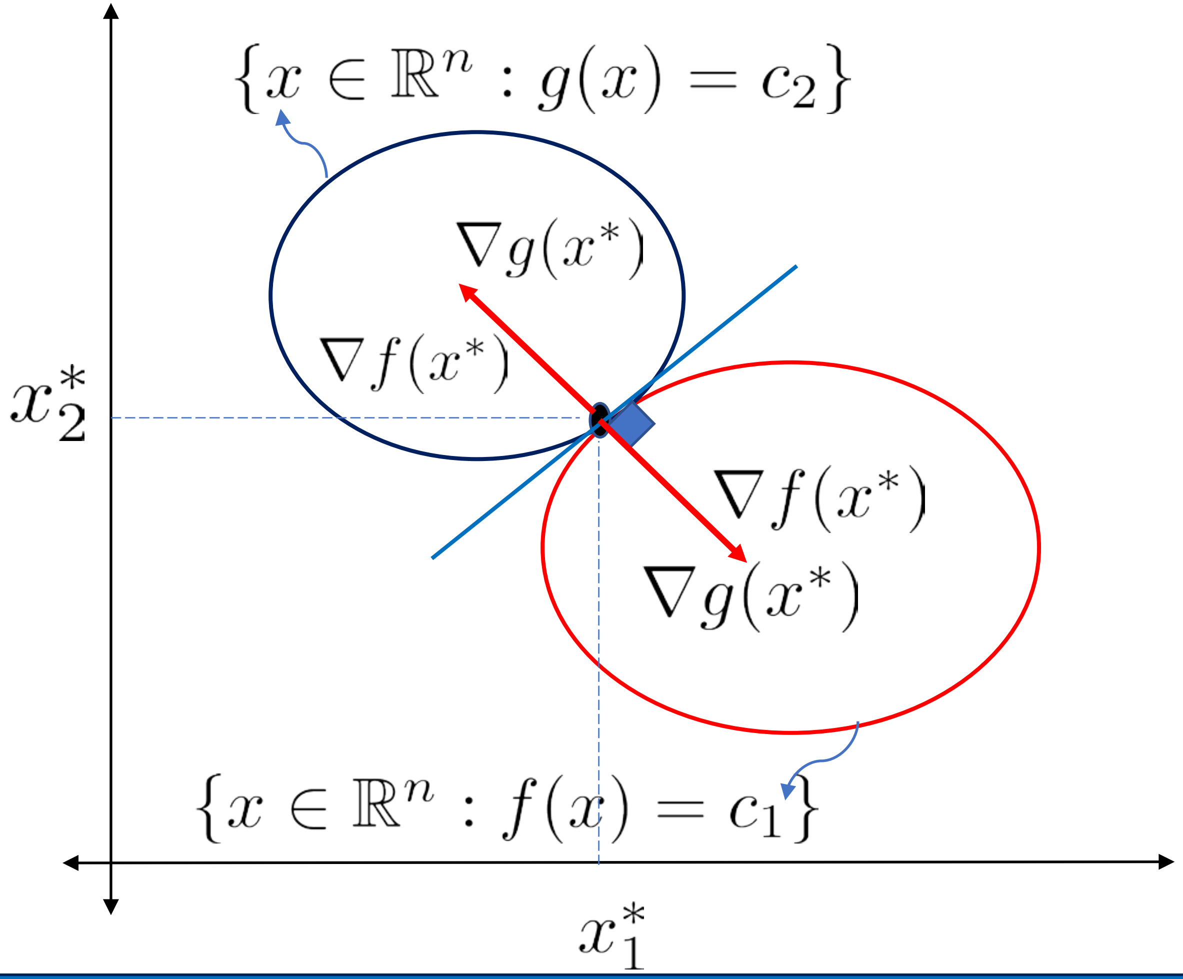
- Assume the level sets

$$\{x \in \mathbb{R}^n : f(x) = c_1\} \text{ and } \{x \in \mathbb{R}^n : g(x) = c_2\}$$

are tangent at the point x^*.

What is the relationship of the gradient vectors of $f(x)$ and $g(x)$ at point x^*?

➢ Gradient vectors $\nabla f(x^*)$ and $\nabla g(x^*)$ are parallel at the tangent point x^*.



Tangent Level sets

- Function $f(x):\mathbb{R}^n \rightarrow \mathbb{R}$ and $g(x):\mathbb{R}^n \rightarrow \mathbb{R}$

- Assume the level sets

$\{x \in \mathbb{R}^n : f(x) = c_1\}$ and $\{x \in \mathbb{R}^n : g(x) = c_2\}$

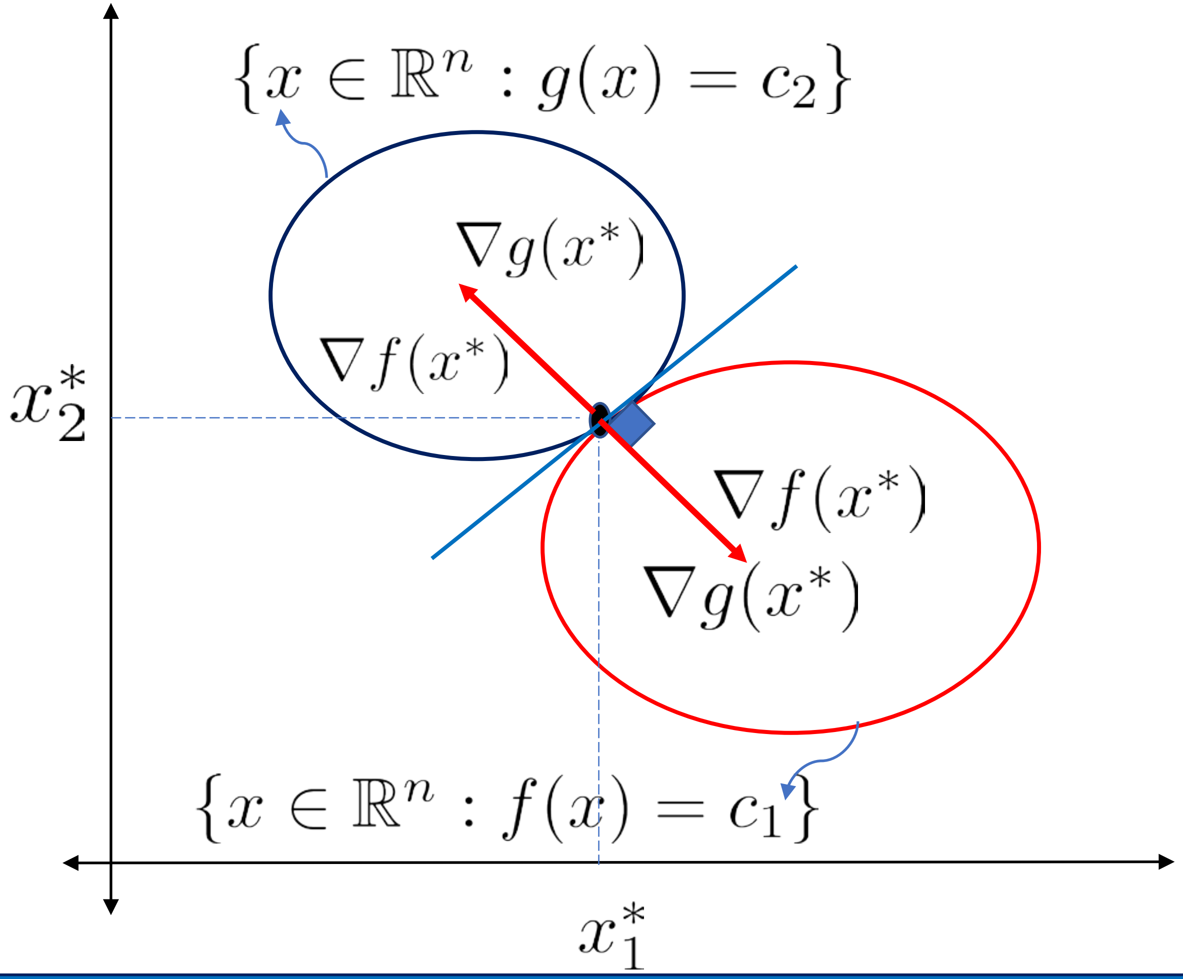
are tangent at the point x^*.

What is the relationship of the gradient vectors of $f(x)$ and $g(x)$ at point x^*?

➢ Gradient vectors $\nabla f(x^*)$ and $\nabla g(x^*)$ are parallel at the tangent point x^*.

$$\nabla f(x^*) = \lambda \nabla g(x^*)$$

Constant (+/-)

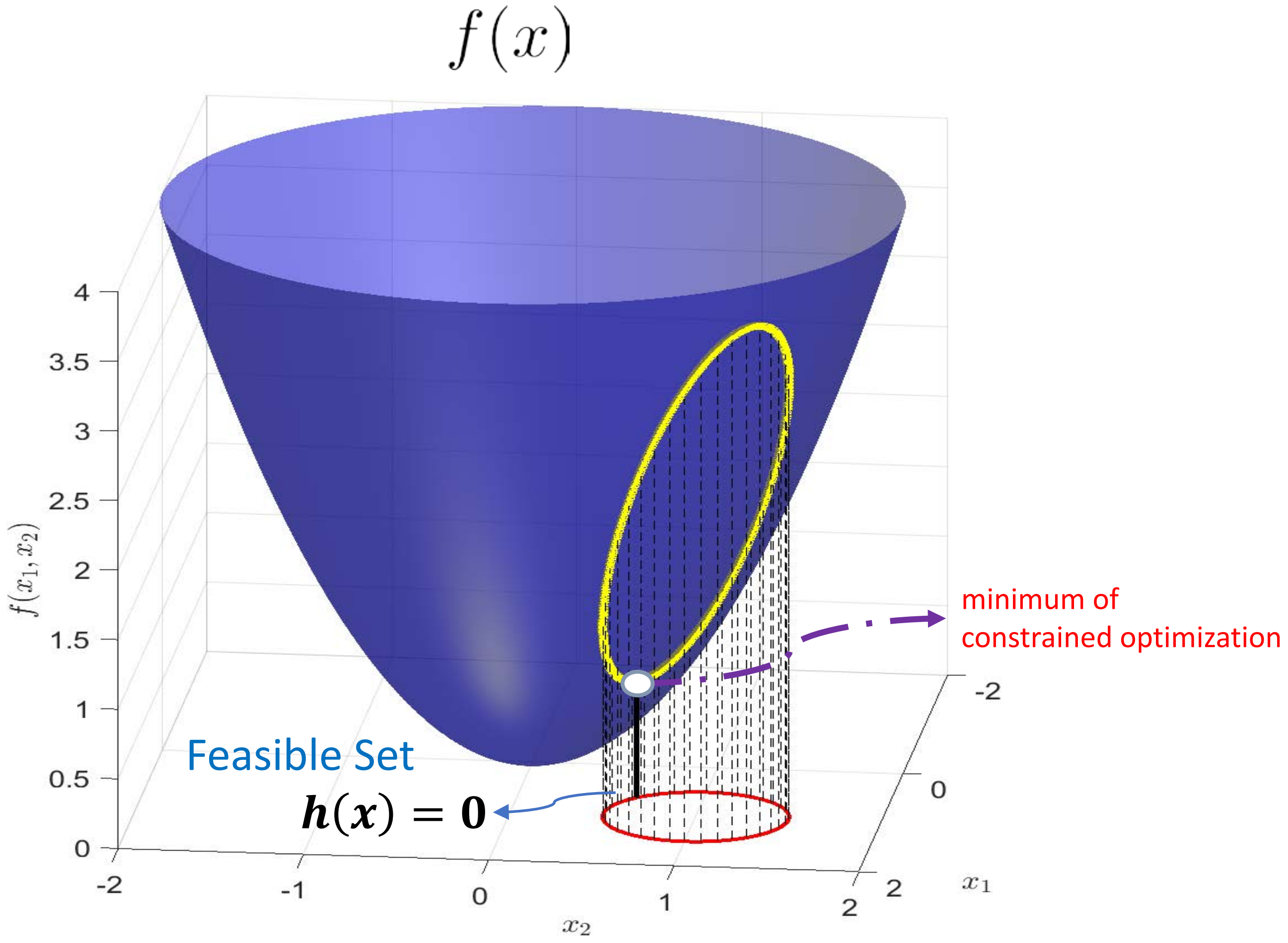


2) Optimality Conditions: Optimization with "**Equality**" Constraints

2) Optimality Conditions: Optimization with "**Equality**" Constraints

Optimization with "**Equality**" Constraint:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned}$$



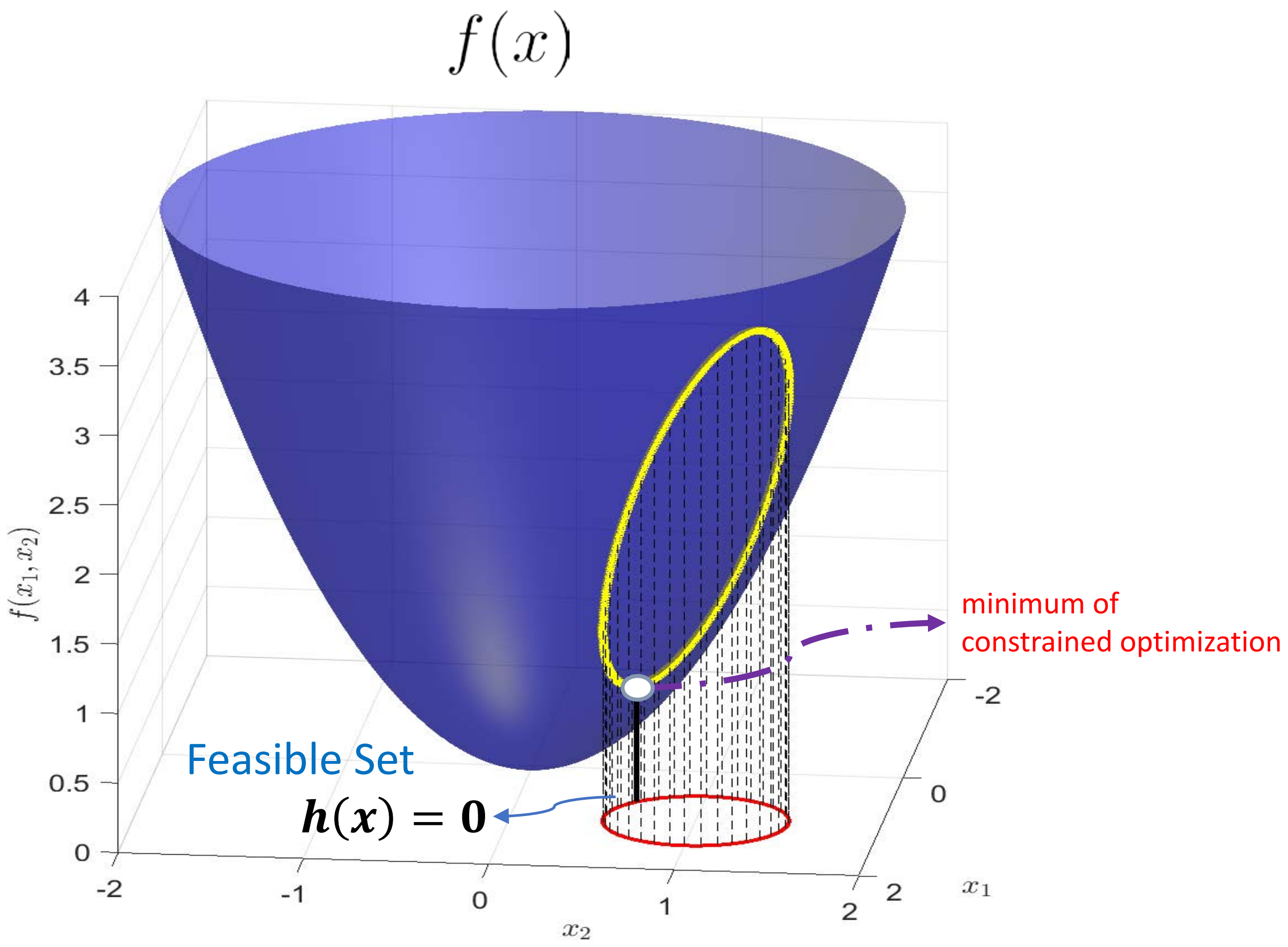
2) Optimality Conditions: Optimization with "Equality" Constraints

Optimization with "Equality" Constraint:

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{minimize}} && f(x) \\ &\text{subject to} && h(x) = 0 \end{aligned}$$

➢ We obtain the optimality condition

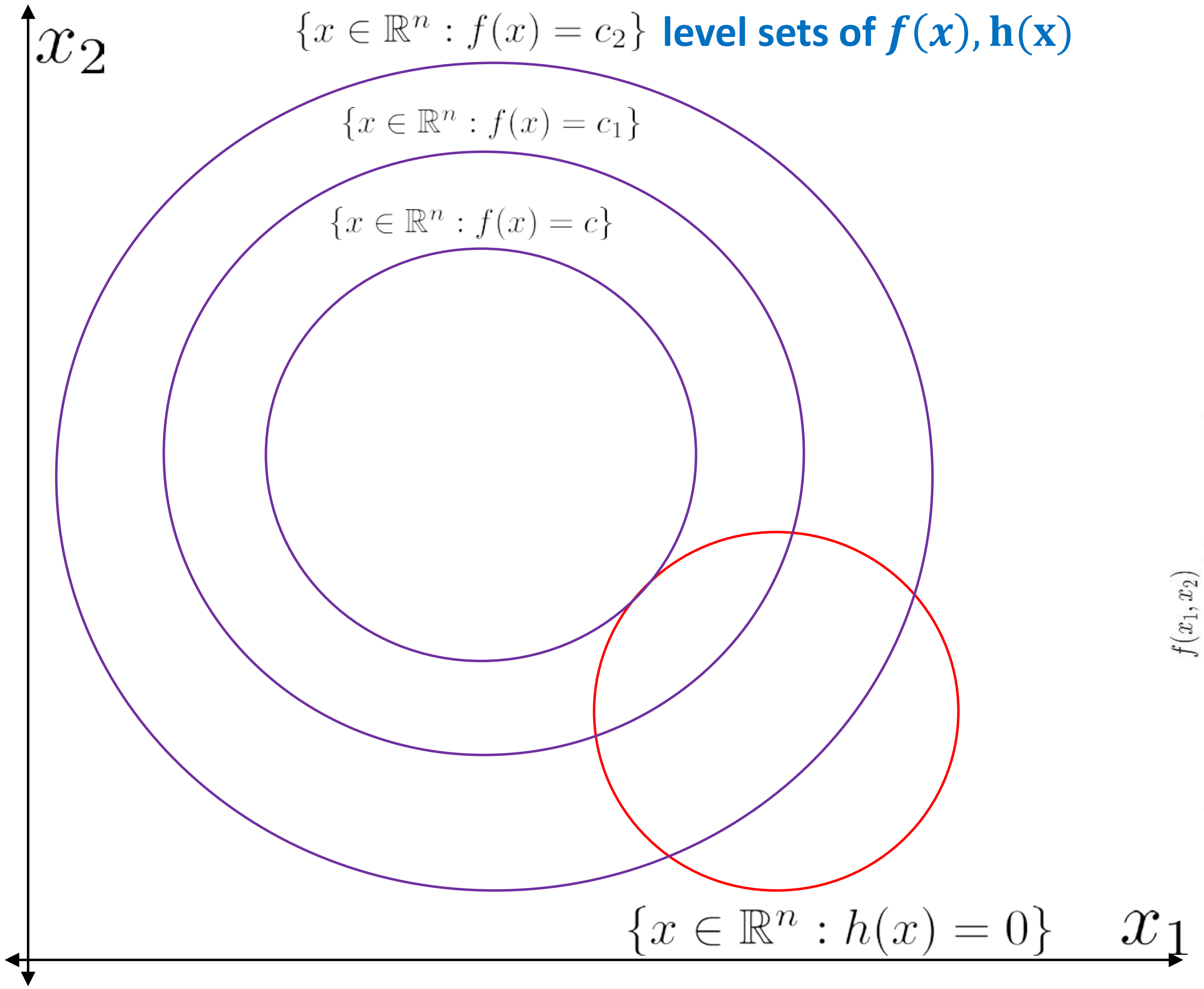
by looking at the level sets of $f(x), \mathrm{h}(\mathrm{x})$.

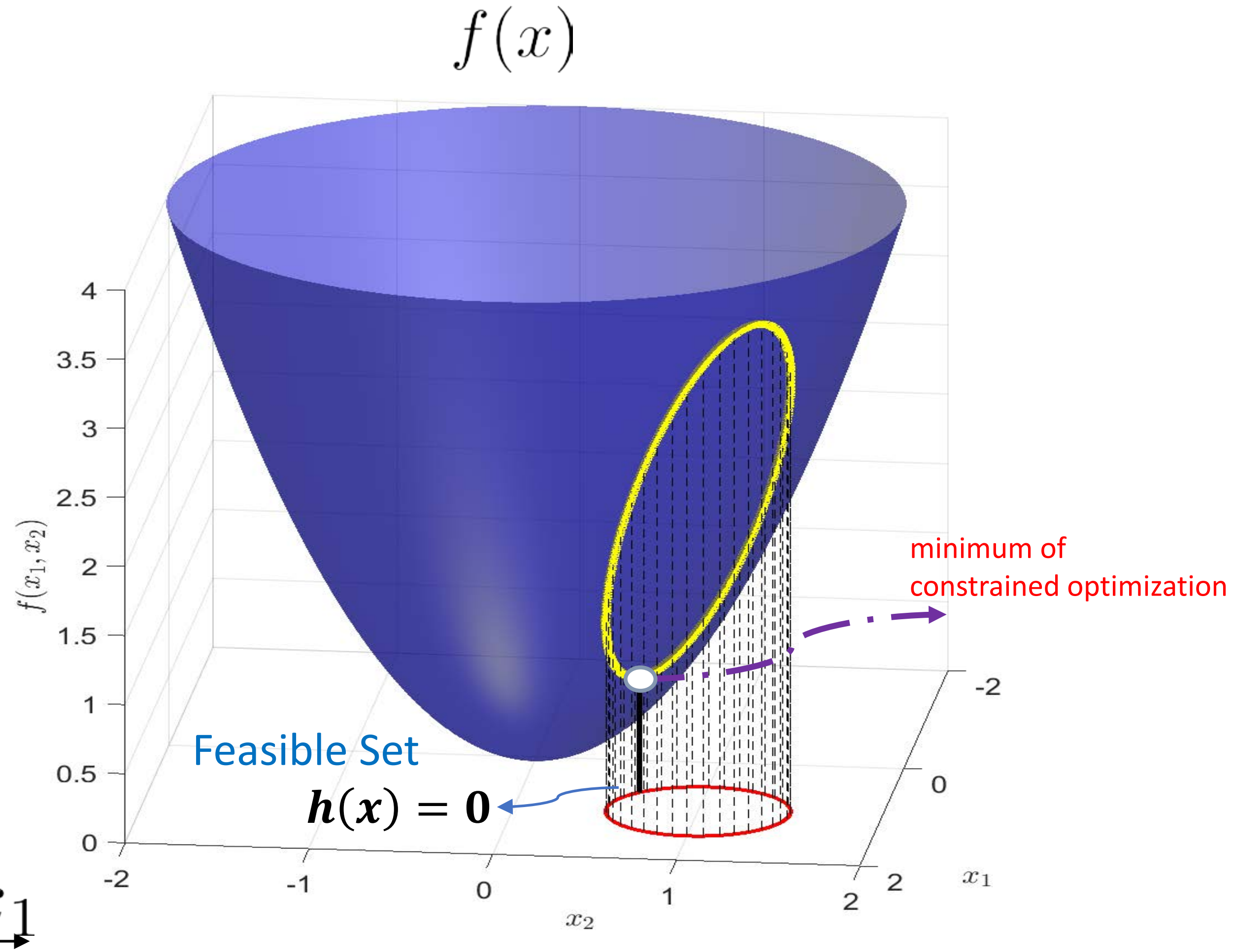


2) Optimality Conditions: Optimization with "Equality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

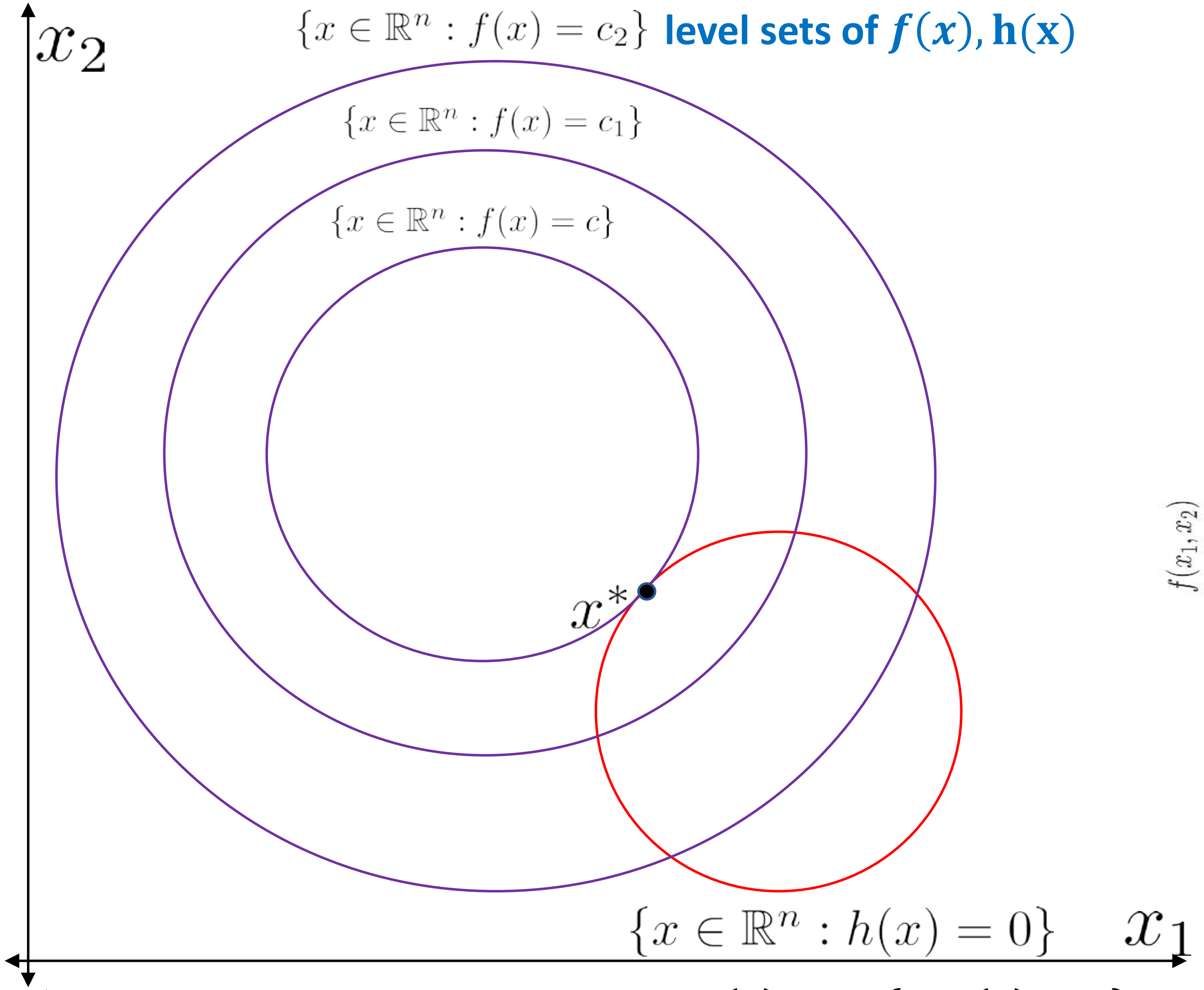
$$\text{subject to} \quad h(x) = 0$$

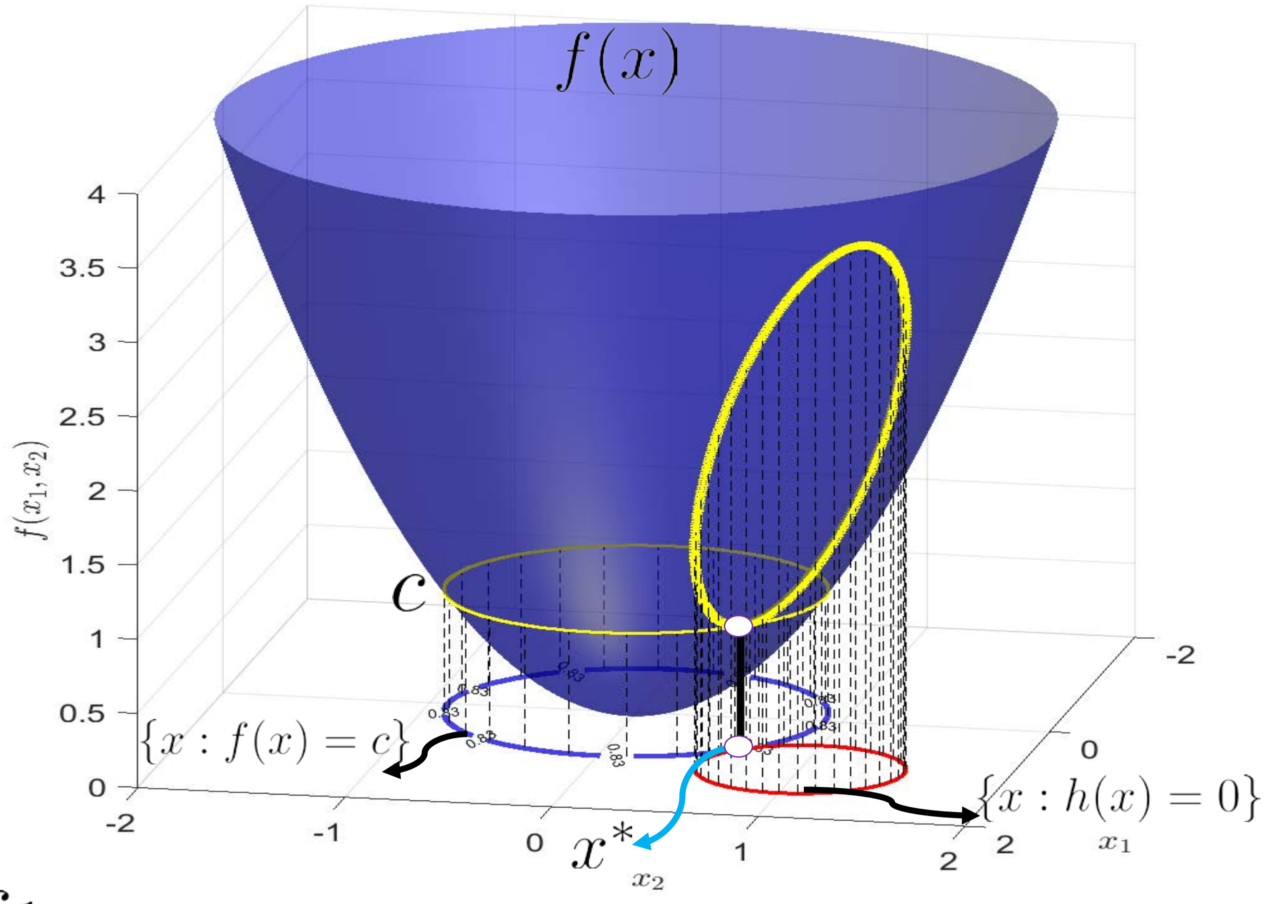




2) Optimality Conditions: Optimization with "Equality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$



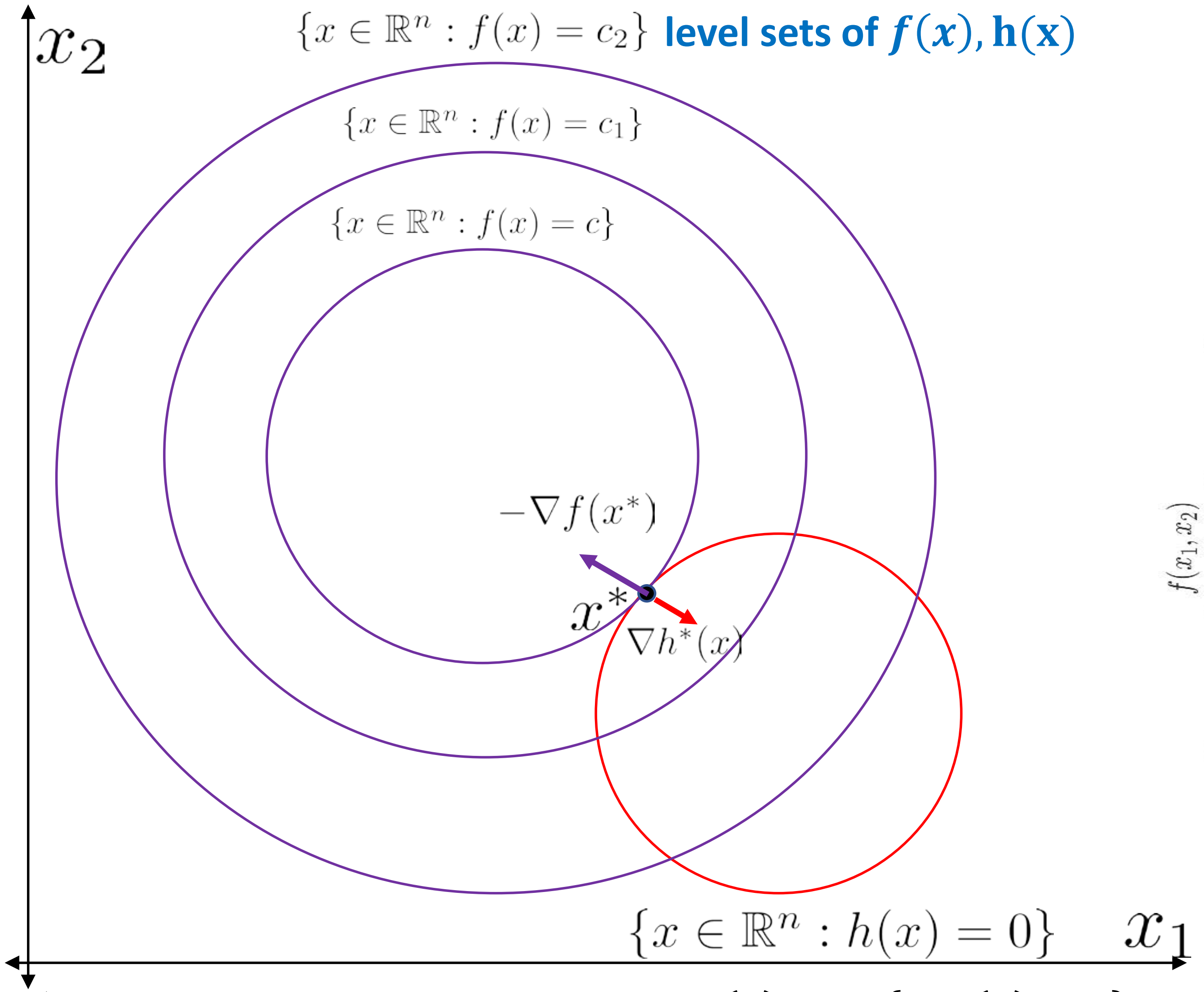


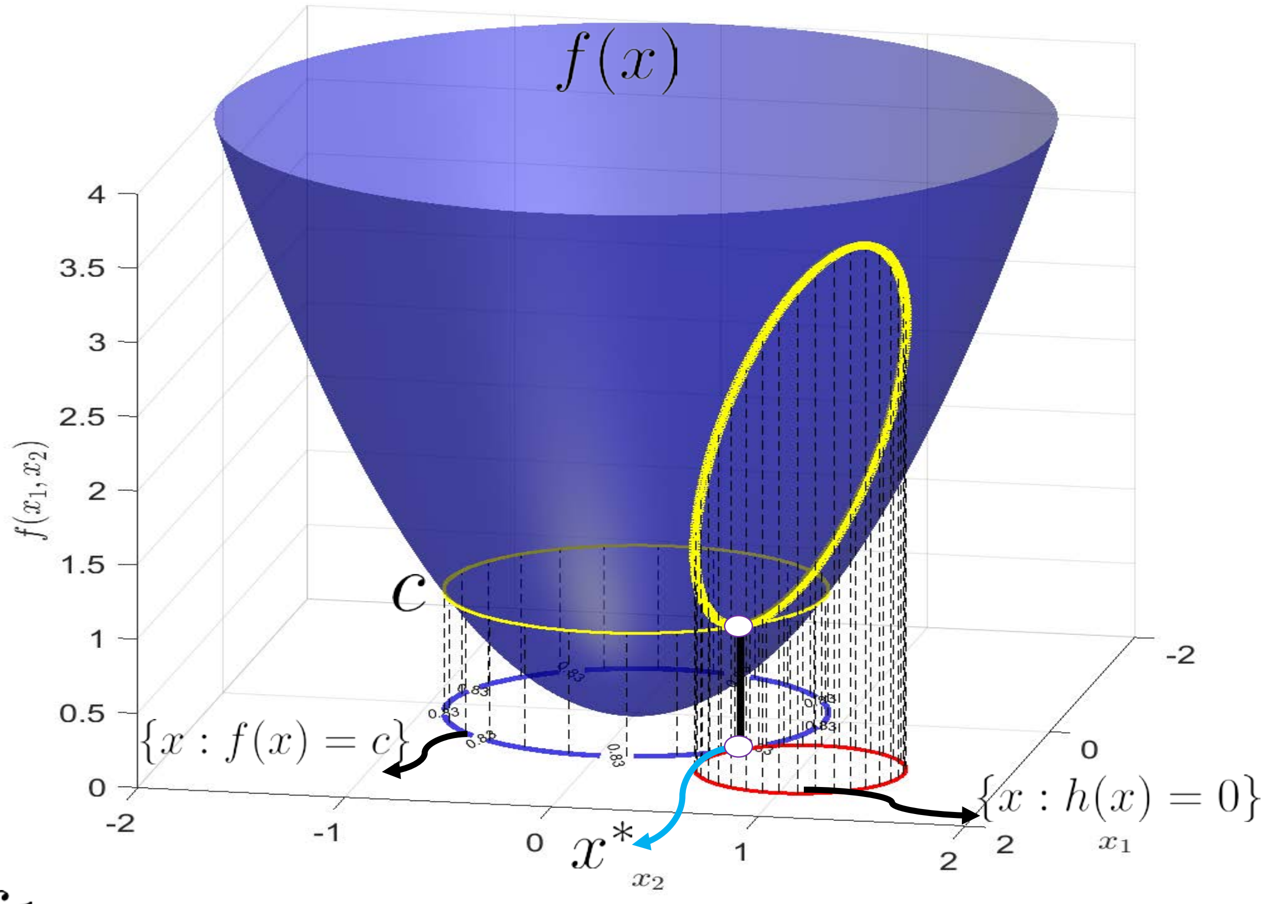
Tangent level sets

- There exist a level set of function $f(x), i.e., \{x: f(x) = c\}$, such that it is tangent to the constraint $\{x: h(x) = 0\}$.
- The tangent point $\boldsymbol{x^*}$ where $\{x: f(x) = c\}$ and $\{x: h(x) = 0\}$ coincide is minimum point of constrained optimization.

2) Optimality Conditions: Optimization with "Equality" Constraints

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned}$$

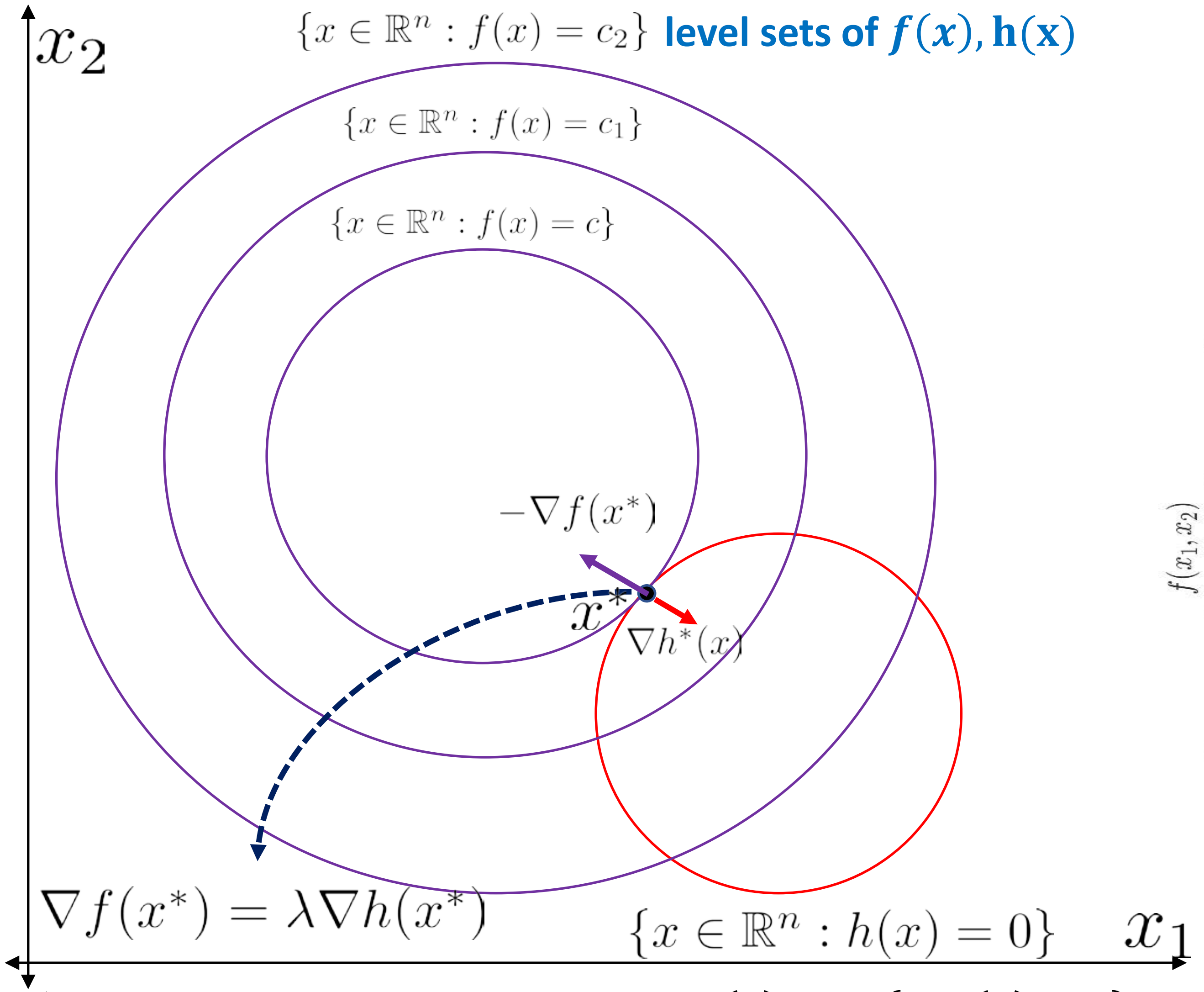


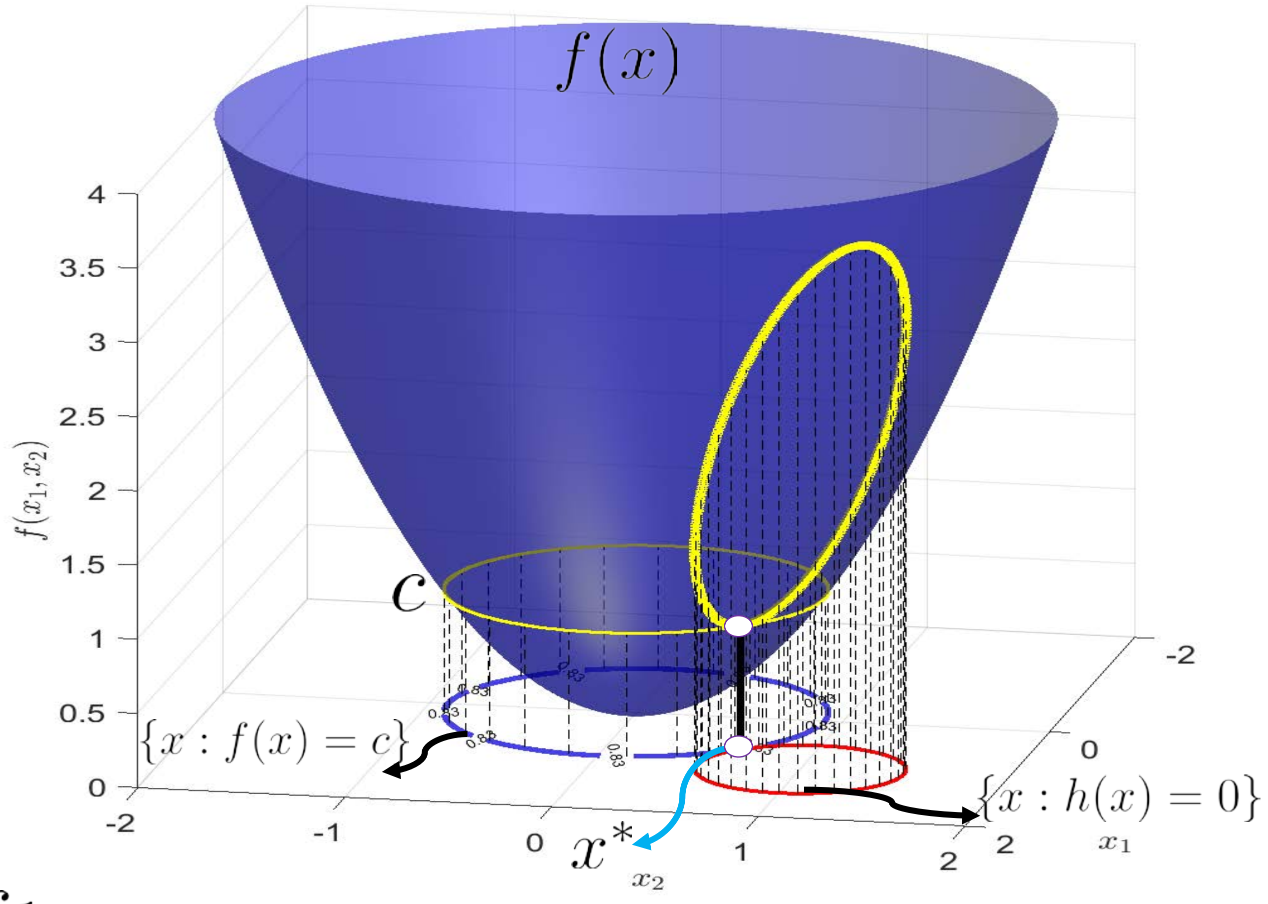


- There exist a level set of function $f(x)$, $i.e.$, $\{x: f(x) = c\}$, such that it is tangent to the constraint $\{x: h(x) = 0\}$.
- The tangent point $\boldsymbol{x}^*$ where $\{x: f(x) = c\}$ and $\{x: h(x) = 0\}$ coincide is minimum point of constrained optimization.

2) Optimality Conditions: Optimization with "Equality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$

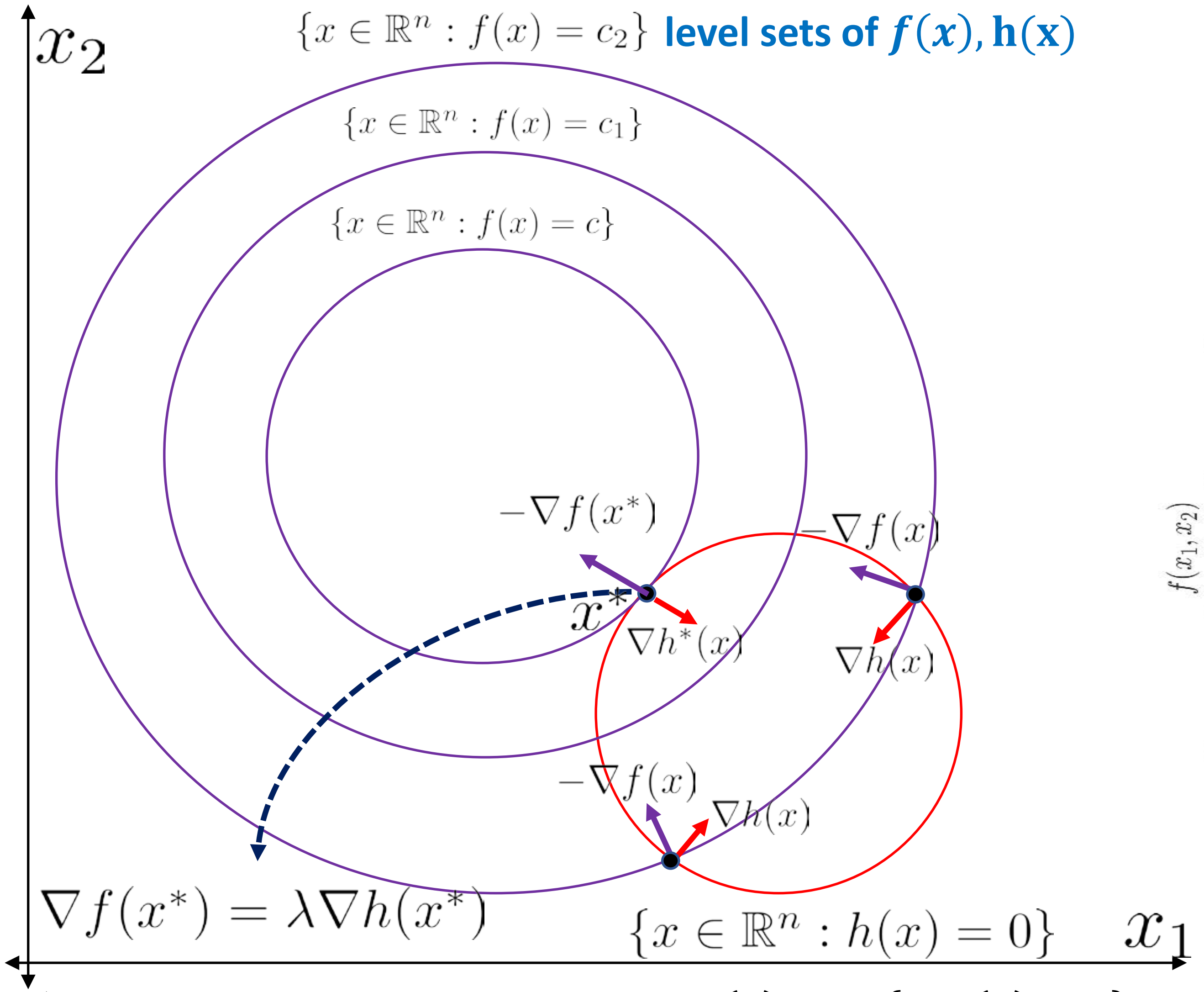


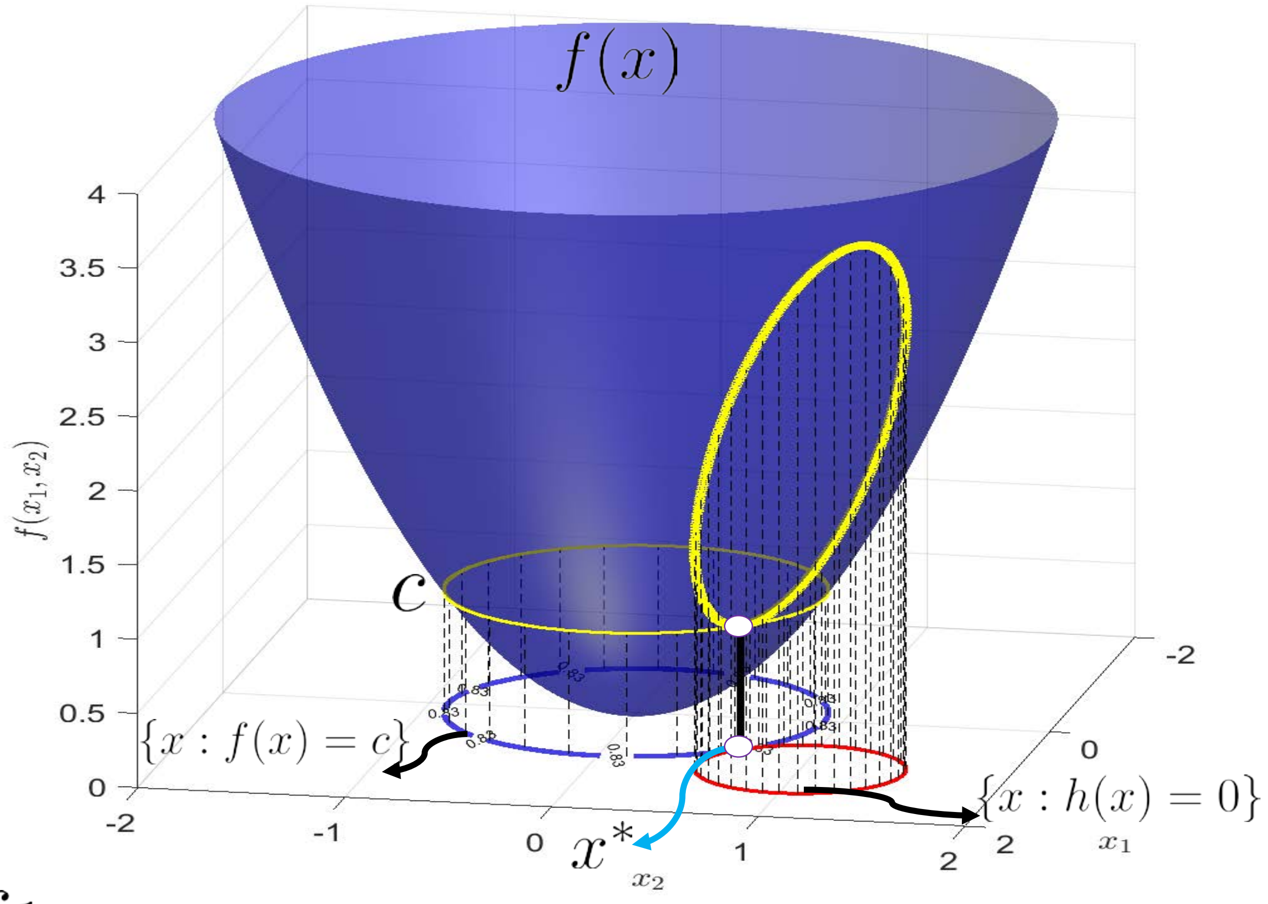


- There exist a level set of function $f(x), i.e., \{x : f(x) = c\}$, such that it is tangent to the constraint $\{x : h(x) = 0\}$.
- The tangent point $\boldsymbol{x}^*$ where $\{x : f(x) = c\}$ and $\{x : h(x) = 0\}$ coincide is minimum point of constrained optimization.

2) Optimality Conditions: Optimization with "Equality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$





Tangent level sets

- There exist a level set of function $f(x)$, $i.e.$, $\{x: f(x) = c\}$, such that it is tangent to the constraint $\{x: h(x) = 0\}$.
- The tangent point $\boldsymbol{x}^*$ where $\{x: f(x) = c\}$ and $\{x: h(x) = 0\}$ coincide is minimum point of constrained optimization.

2) Optimality Conditions: Optimization with "**Equality**" Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned}$$

Optimality Condition: $h(x^*) = 0$ and $\nabla f(x^*) = \lambda \nabla h(x^*)$

2) Optimality Conditions: Optimization with "**Equality**" Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \end{aligned}$$

Optimality Condition: $h(x^*) = 0$ and $\nabla f(x^*) = \lambda \nabla h(x^*)$

Standard Format:

$$L(x,\lambda) = f(x) - \lambda h(x) \xrightarrow{\text{Optimality Cond.}} \nabla_{x,\lambda} L(x,\lambda) = 0 \quad \begin{cases} \nabla_x L(x,\lambda) = 0 \longrightarrow \nabla f(x^*) = \lambda \nabla h(x^*) \\ \nabla_\lambda L(x,\lambda) = 0 \longrightarrow -h(x^*) = 0 \end{cases}$$

Lagrange function *Lagrange multiplier*

2) Optimality Conditions: Optimization with "**Equality**" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$

Optimality Condition: $h(x^*) = 0$ and $\nabla f(x^*) = \lambda \nabla h(x^*)$

Standard Format:

$$L(x, \lambda) = f(x) - \lambda h(x) \xrightarrow{\text{Optimality Cond.}} \nabla_{x,\lambda} L(x, \lambda) = 0 \quad \begin{cases} \nabla_x L(x, \lambda) = 0 \longrightarrow \nabla f(x^*) = \lambda \nabla h(x^*) \\ \nabla_\lambda L(x, \lambda) = 0 \longrightarrow -h(x^*) = 0 \end{cases}$$

Lagrange function *Lagrange multiplier*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h_i(x) = 0, \quad i = 1, \ldots, n_h$$

$$L(x, \lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x) \xrightarrow{\text{Optimality Cond.}} \begin{matrix} \nabla_x L(x, \lambda) = 0 \\ \nabla_\lambda L(x, \lambda) = 0 \end{matrix}$$

Lagrange function

2) Optimality Conditions: Optimization with "**Equality**" Constraints

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h(x) = 0$$

Optimality Condition: $h(x^*) = 0$ and $\nabla f(x^*) = \lambda \nabla h(x^*)$

Standard Format:

$$L(x,\lambda) = f(x) - \lambda h(x) \xrightarrow{\text{Optimality Cond.}} \nabla_{x,\lambda} L(x,\lambda) = 0 \quad \begin{cases} \nabla_x L(x,\lambda) = 0 \longrightarrow \nabla f(x^*) = \lambda \nabla h(x^*) \\ \nabla_\lambda L(x,\lambda) = 0 \longrightarrow -h(x^*) = 0 \end{cases}$$

Lagrange function *Lagrange multiplier*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h_i(x) = 0, \quad i = 1, \ldots, n_h$$

$$L(x,\lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x) \xrightarrow{\text{Optimality Cond.}^1} \begin{cases} \nabla_x L(x,\lambda) = 0 \\ \nabla_\lambda L(x,\lambda) = 0 \end{cases}$$

Lagrange function

.[1]Note that λ could be +/-. Hence, Lagrange function could also take the following form: $L(x,\lambda) = f(x) + \sum_{i=1}^{n_h} \lambda_i h_i(x)$

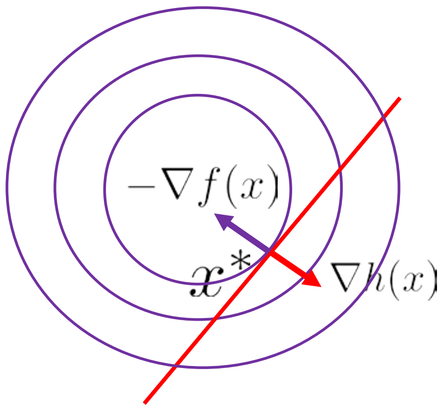
Chapter 3.1: Necessary Conditions For Equality Constraints, "Nonlinear Programming" Dimitri Bertsekas, MIT,

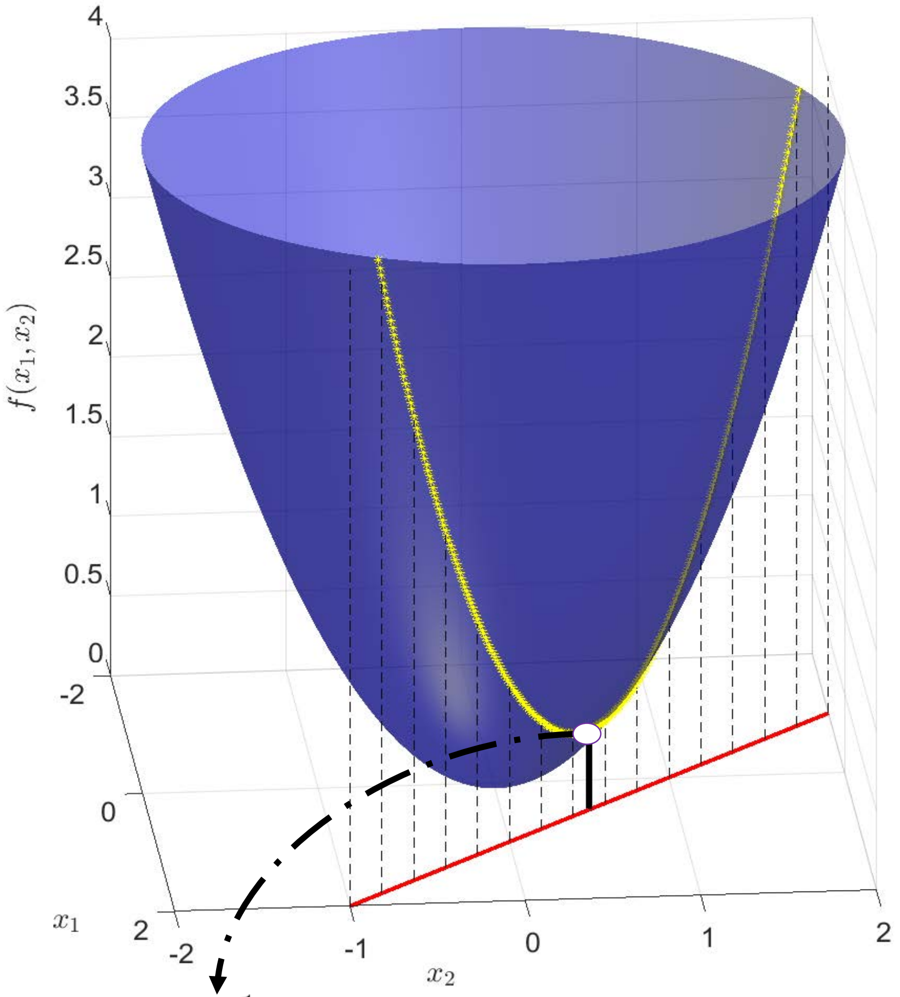
2) Optimality Conditions: Optimization with “Equality” Constraints

Example

$$\begin{aligned} \underset{x\in\mathbb{R}^2}{\text{minimize}} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 1 = 0 \end{aligned}$$

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





Optimality Cond.

$$\nabla_{x,\lambda} L(x,\lambda) = 0 \quad \begin{cases} \nabla_x L(x,\lambda) = 0 \longrightarrow & \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \nabla_\lambda L(x,\lambda) = 0 \longrightarrow & x_1^* + x_2^* - 1 = 0 \end{cases}$$

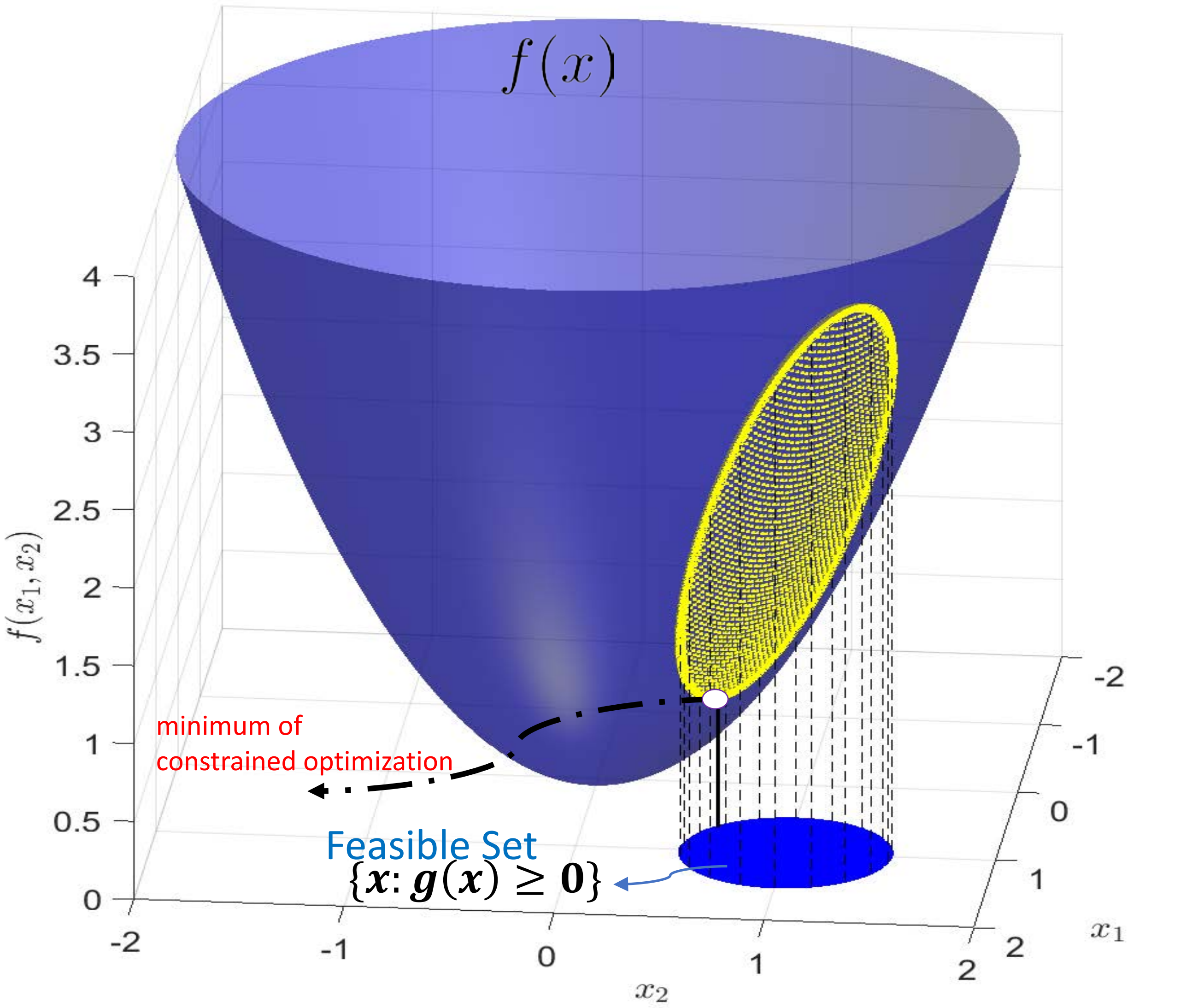
$$x_1^* = x_2^* = \frac{1}{2}, \quad \lambda^* = 1$$

3) Optimality Conditions: Optimization with "Inequality" Constraints

3) Optimality Conditions: Optimization with "Inequality" Constraints

Optimization with **"Inequality"** Constraint:

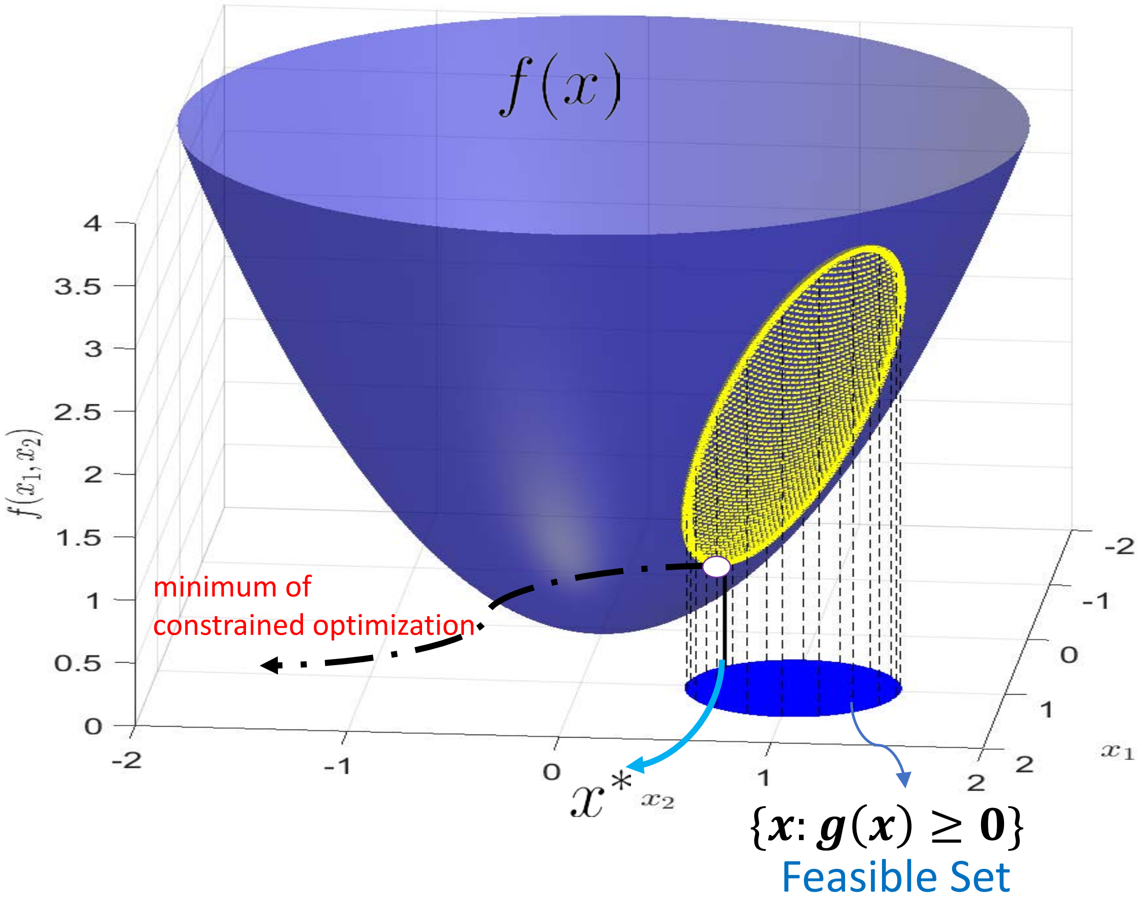
$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$



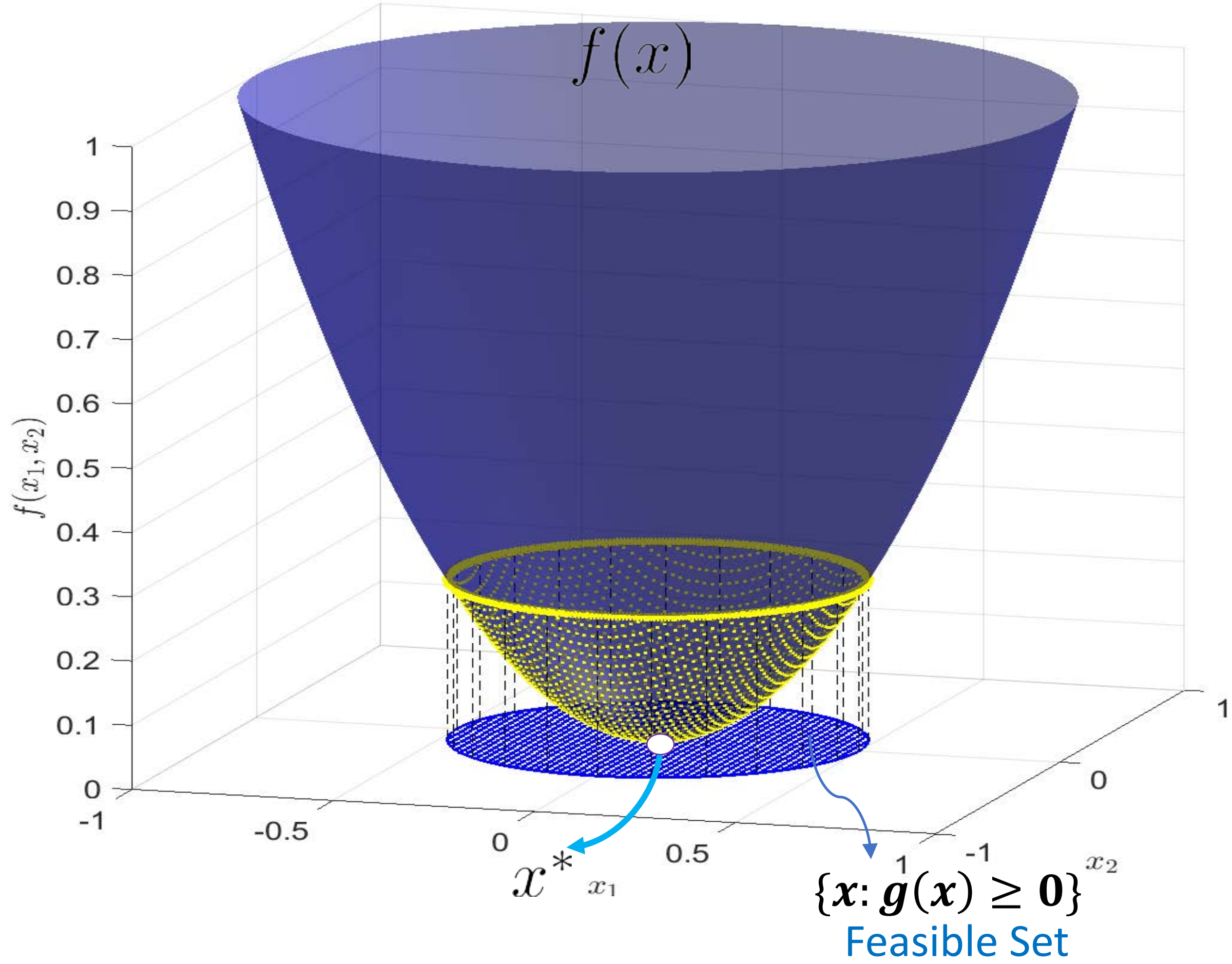
3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1: x^* is on the boundary of the feasible region.



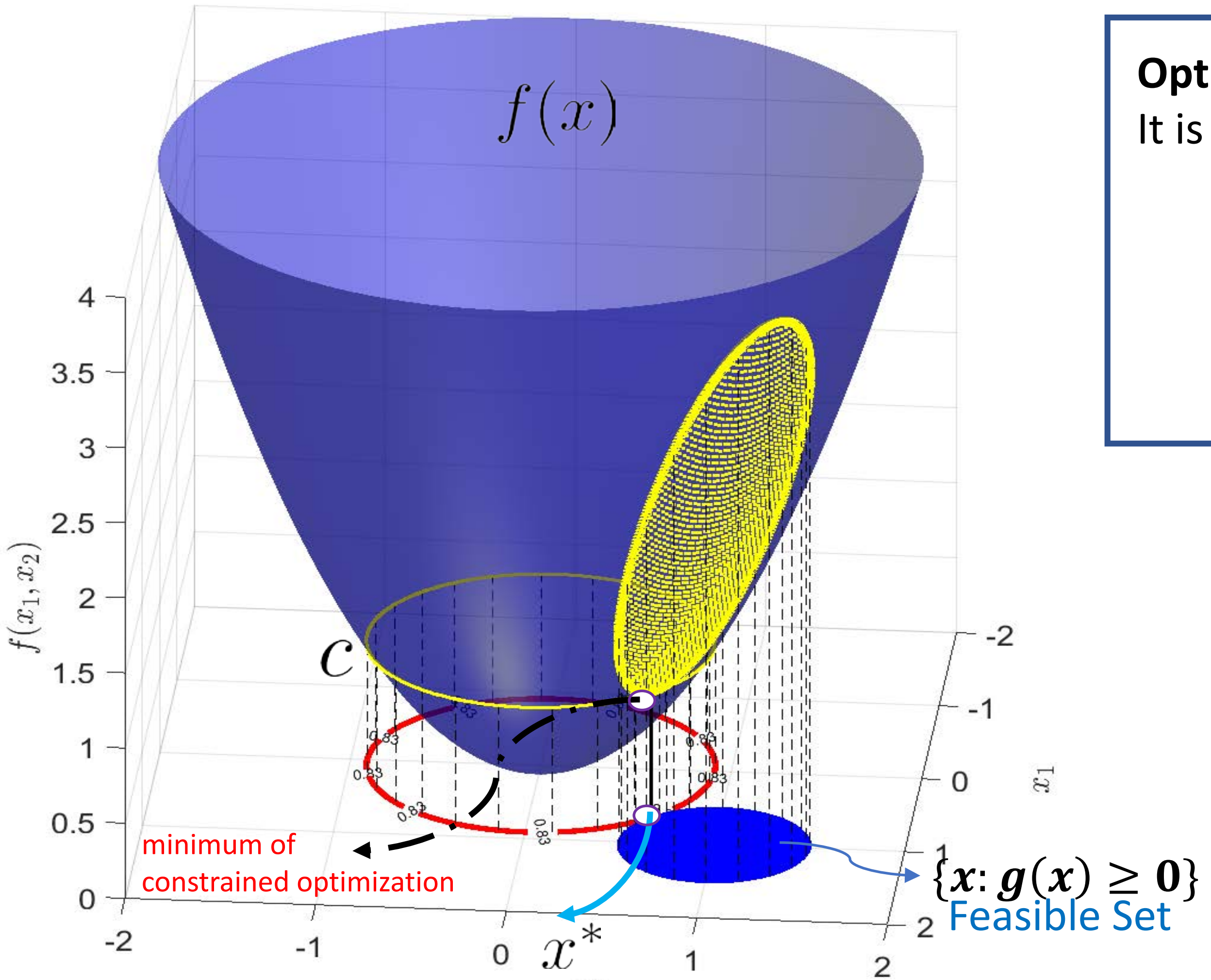
Case 2: x^* is inside the feasible region.



3) Optimality Conditions: Optimation with "Inequality" Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 1: x^* is on the boundary of the feasible region.



Optimality Condition:

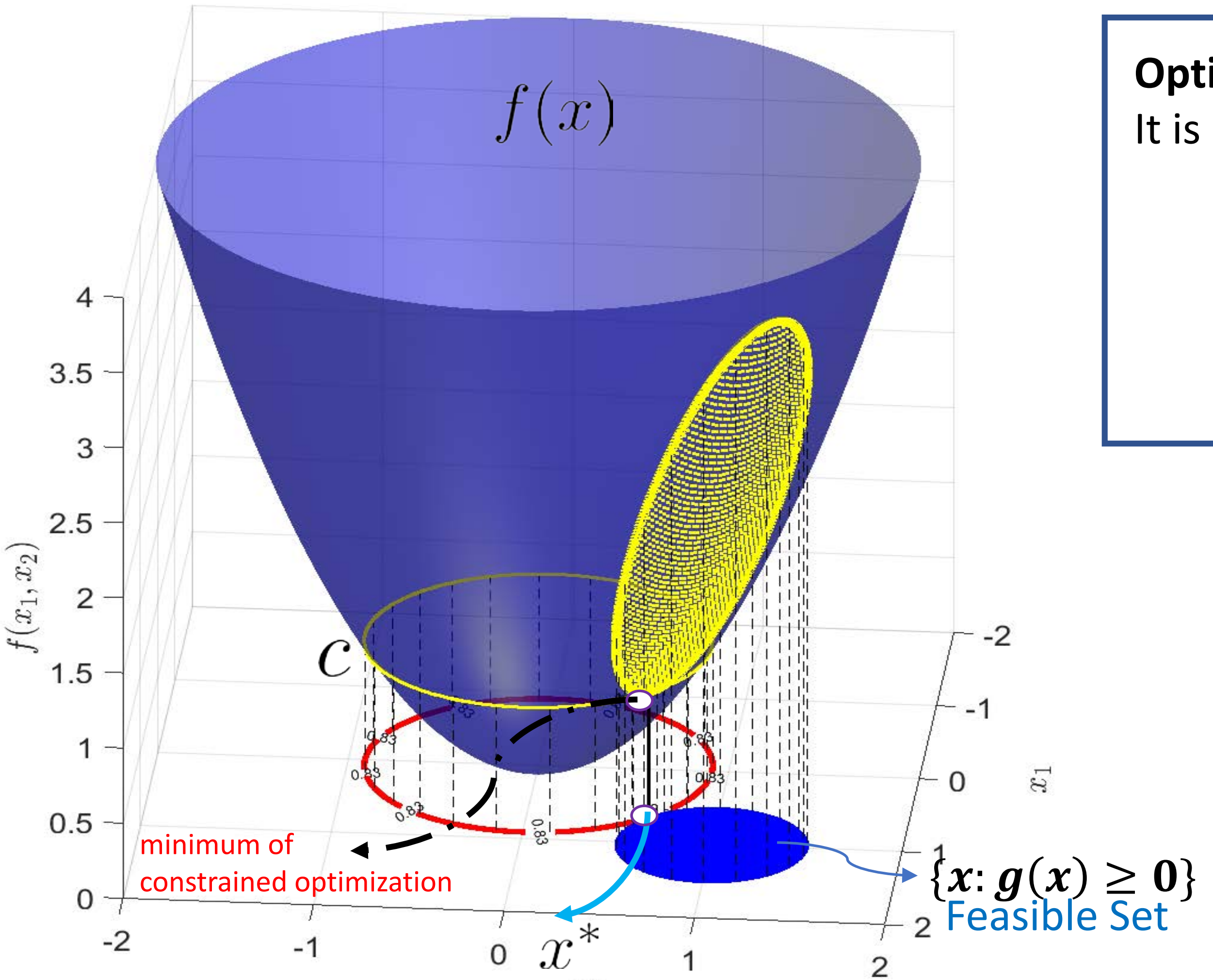
It is **similar** to constrained optimization with **equality constraint**.

$$g(x^*) = 0 \quad \text{and} \quad \nabla f(x^*) = \mu \nabla g(x^*)$$

3) Optimality Conditions: Optimization with "Inequality" Constraints

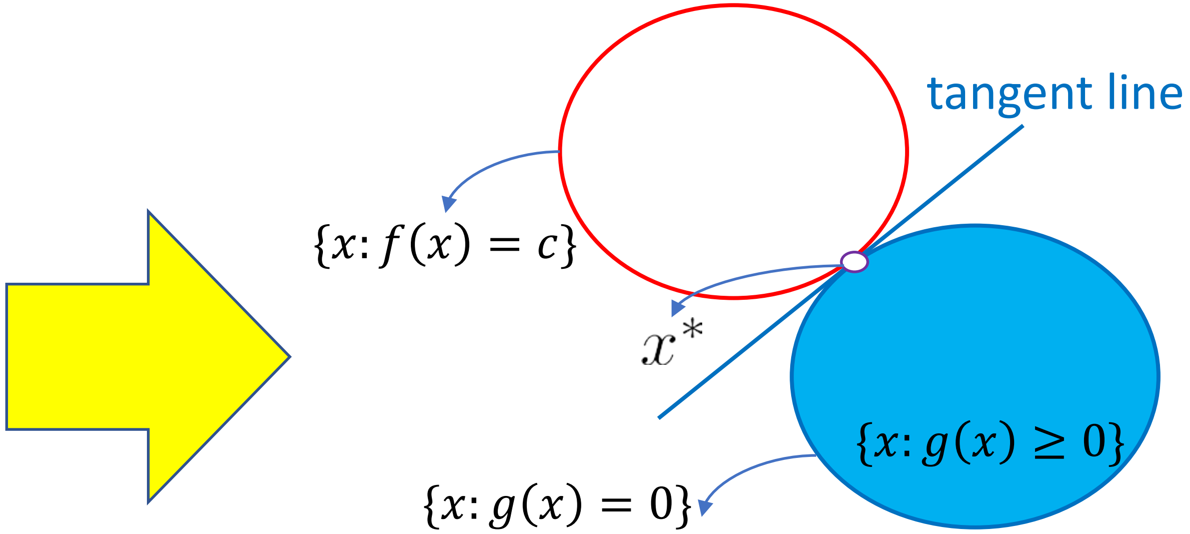
$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 1: x^* is on the boundary of the feasible region.



Optimality Condition:
It is similar to constrained optimization with equality constraint.

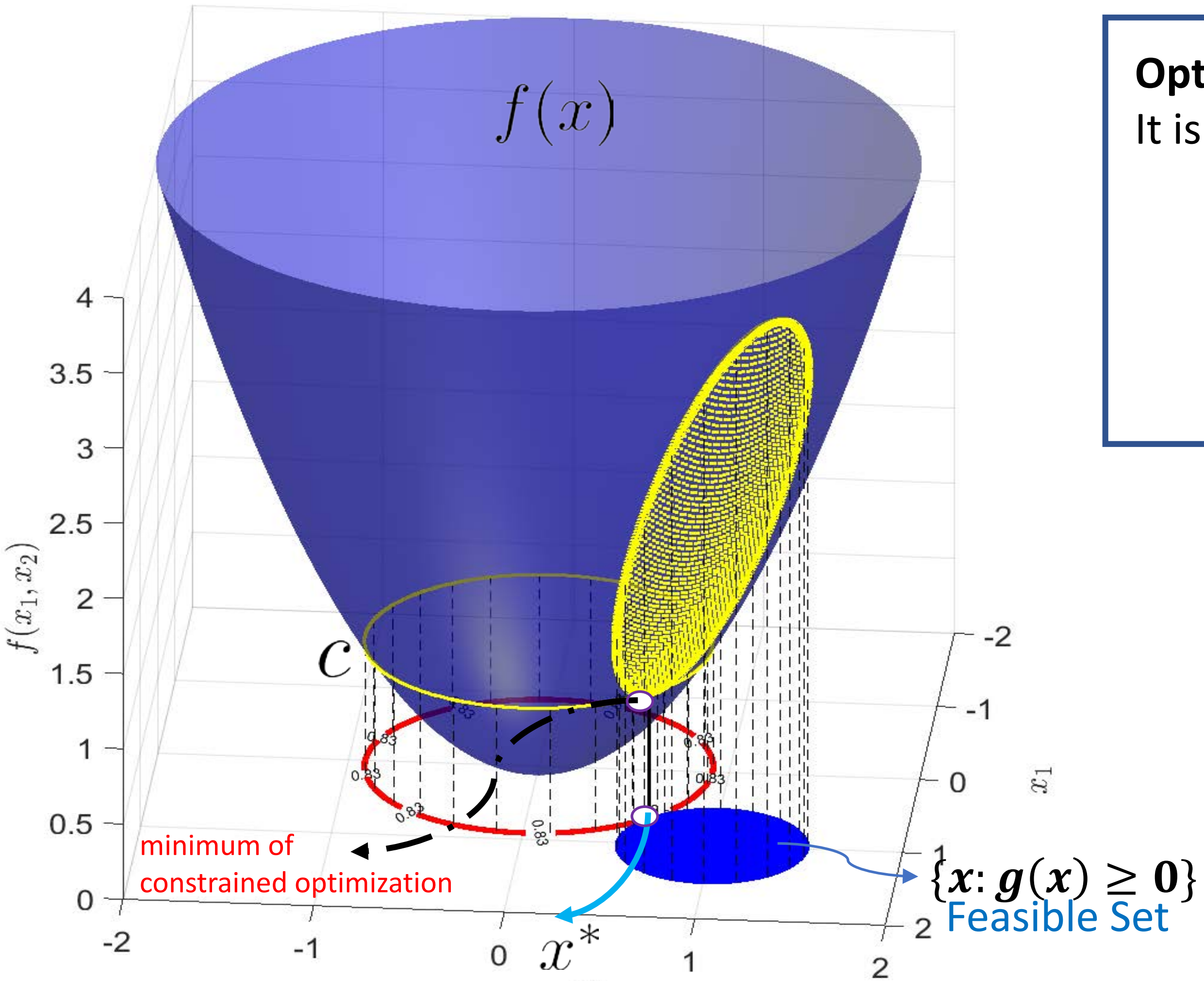
$$g(x^*) = 0 \quad \text{and} \quad \nabla f(x^*) = \mu \nabla g(x^*)$$



3) Optimality Conditions: Optimization with “Inequality” Constraints

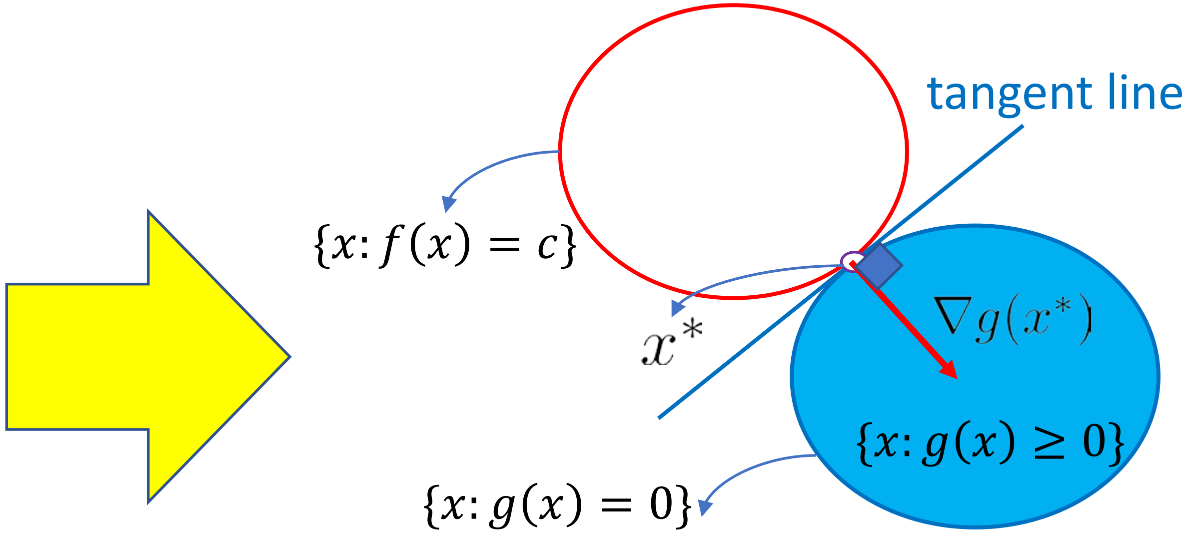
$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 1: x^* is on the boundary of the feasible region.



Optimality Condition:
It is similar to constrained optimization with equality constraint.

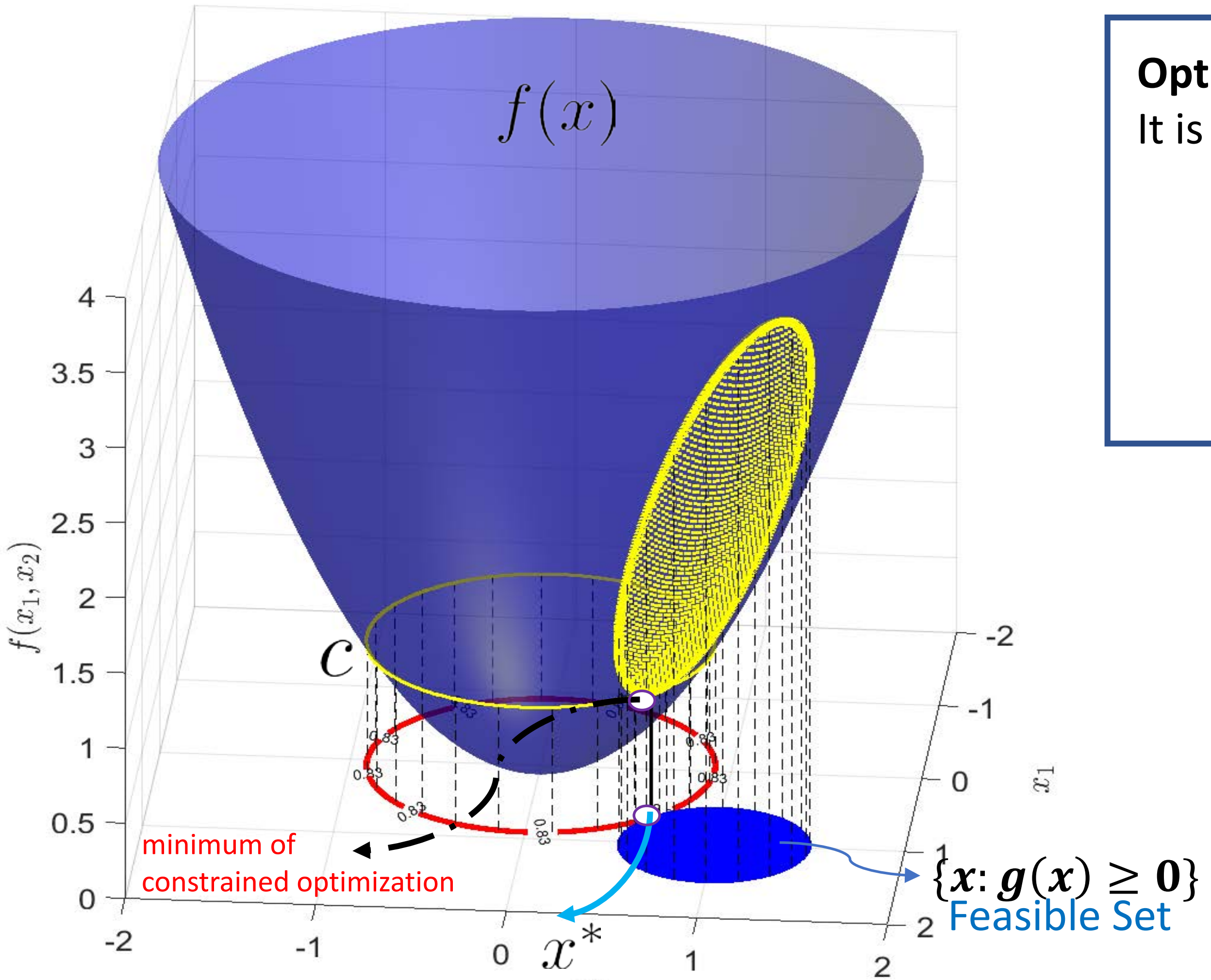
$$g(x^*) = 0 \quad \text{and} \quad \nabla f(x^*) = \mu \nabla g(x^*)$$



3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

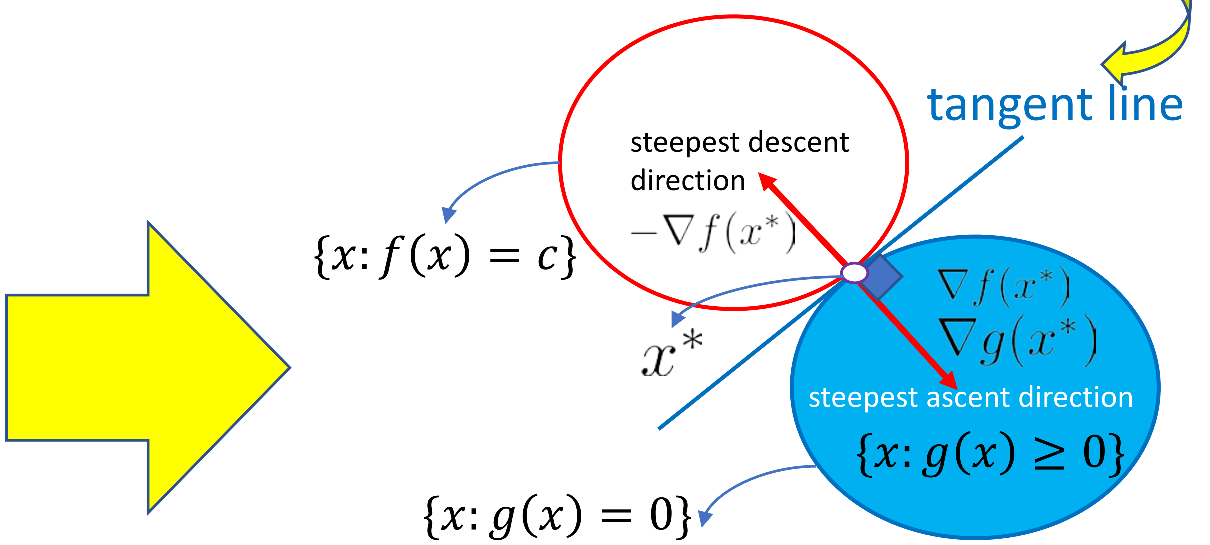
Case 1: x^* is on the boundary of the feasible region.



Optimality Condition:
It is similar to constrained optimization with equality constraint.

$$g(x^*) = 0 \quad \text{and} \quad \nabla f(x^*) = \mu \nabla g(x^*)$$

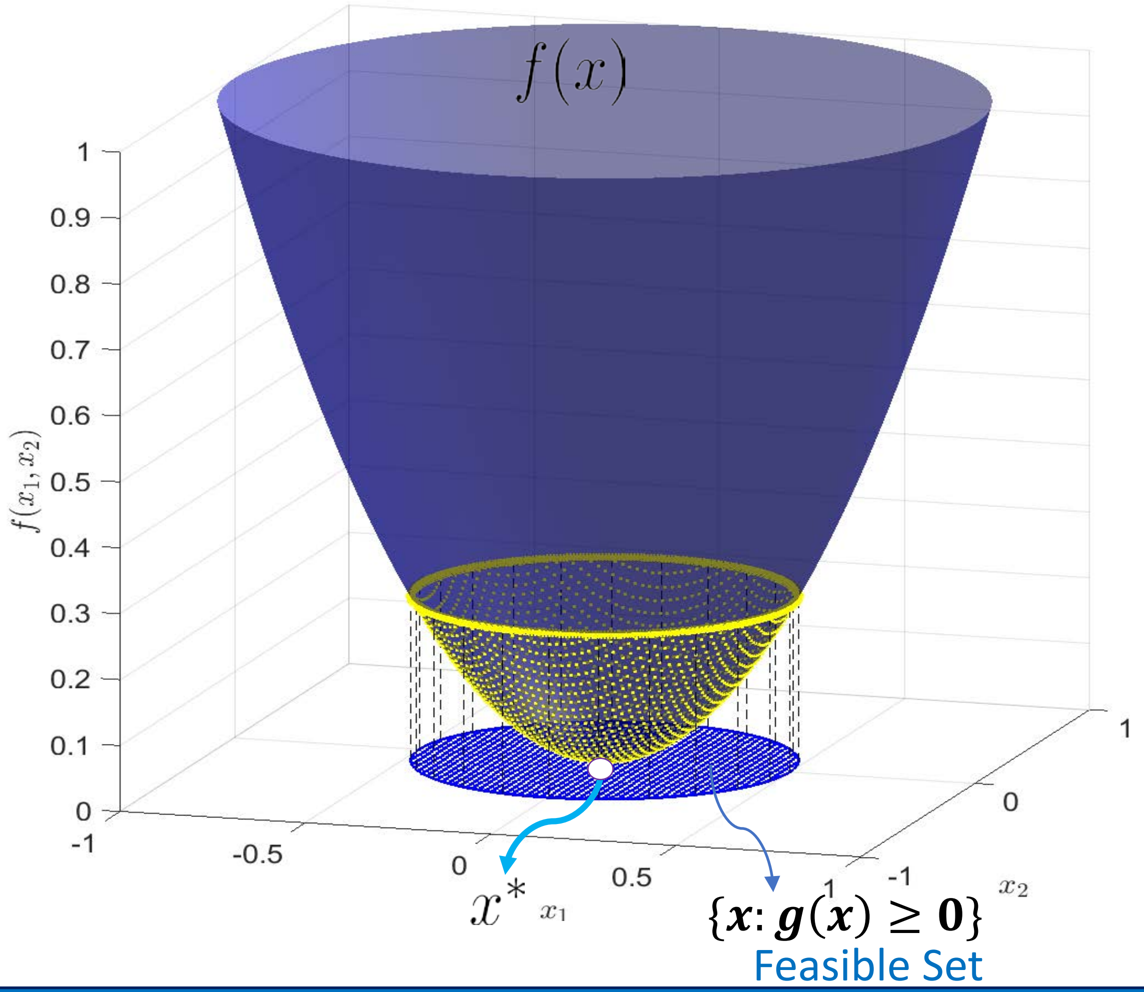
$$\mu \geq 0$$



3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 2: x^* is inside the feasible region.



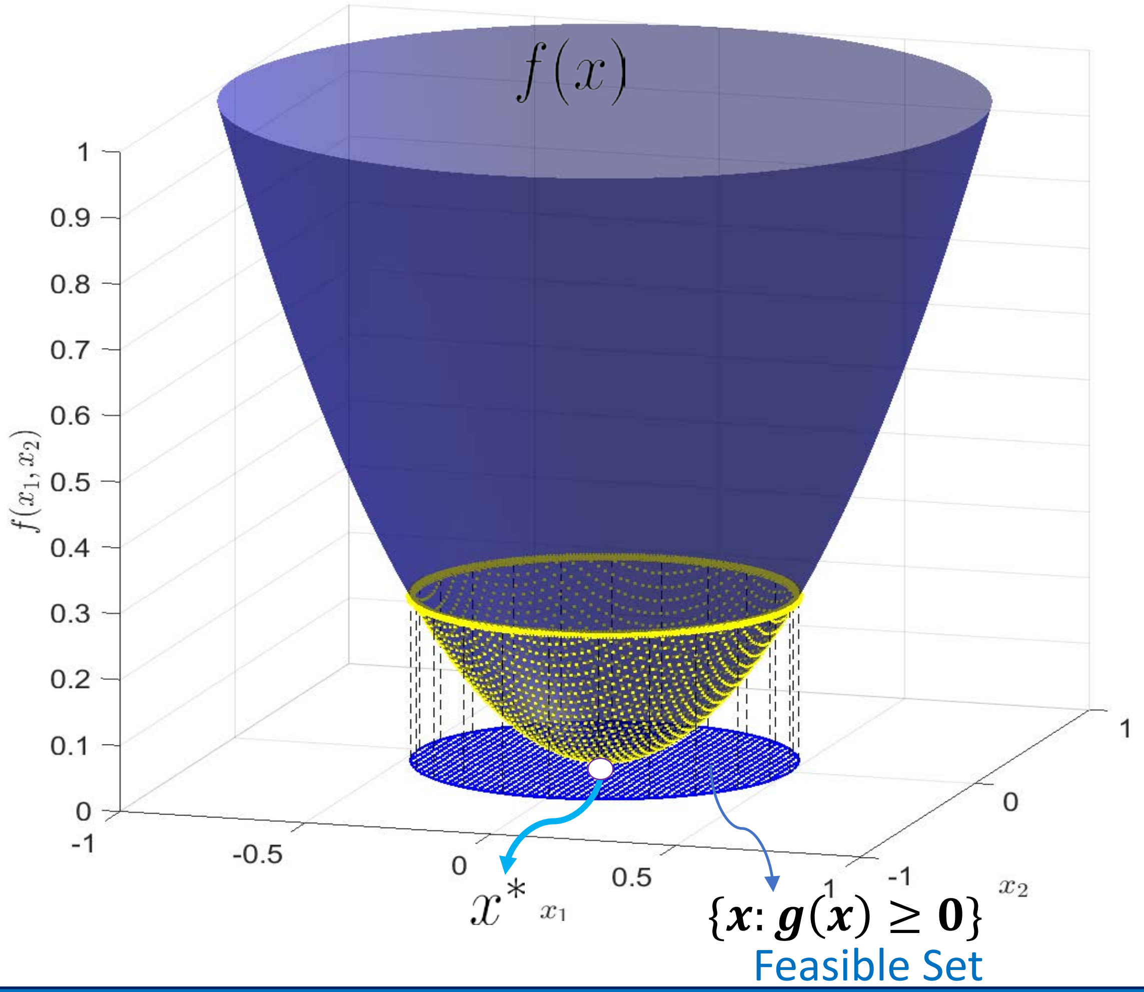
Optimality Condition:
It is similar to "*unconstrained*" optimization.

$$g(x^*) > 0 \quad \text{and} \quad \boxed{\nabla f(x^*) = 0}$$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 2: x^* is inside the feasible region.



Optimality Condition:
It is similar to "*unconstrained*" optimization.

$$g(x^*) > 0 \quad \text{and} \quad \boxed{\nabla f(x^*) = 0}$$

$$\nabla f(x^*) = \mu^* \nabla g(x^*)$$
$$\mu^* = 0$$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* \geq 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* \geq 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$$g(x^*) \geq 0$$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* \geq 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$g(x^*) \geq 0$ and $\mu^* \geq 0$

3) Optimality Conditions: Optimization with “Inequality” Constraints

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g(x) \geq 0 \end{aligned}$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* \geq 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$g(x^*) \geq 0$ and $\mu^* \geq 0$ and $\mu^* g(x^*) = 0$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* \geq 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$g(x^*) \geq 0$ and $\mu^* \geq 0$ and $\mu^* g(x^*) = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* > 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$g(x^*) \geq 0$ and $\mu^* \geq 0$ and $\mu^* g(x^*) = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Standard Format:

$L(x, \mu) = f(x) - \mu g(x)$ **Optimality Cond.** $\nabla_x L(x^*, \mu^*) = 0$

$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$

Lagrange function *Lagrange multiplier*

3) Optimality Conditions: Optimization with "Inequality" Constraints

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g(x) \geq 0$$

Case 1) Optimality Condition:
It is similar to constrained optimization with equality constraint.

$g(x^*) = 0$ and $\mu^* > 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Case 2) Optimality Condition:
It is similar to unconstrained optimization.

$g(x^*) > 0$ and $\mu^* = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

$g(x^*) \geq 0$ and $\mu^* \geq 0$ and $\mu^* g(x^*) = 0$ and $\nabla f(x^*) = \mu^* \nabla g(x^*)$

Standard Format:

$L(x, \mu) = f(x) - \mu g(x)$ **Optimality Cond.** $\nabla_x L(x^*, \mu^*) = 0$

$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$

Lagrange function *Lagrange multiplier*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \geq 0, \; i = 1, \ldots, n_g$$

Lagrange function

$L(x, \lambda) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x)$ **Optimality Cond.**

$\nabla_x L(x^*, \mu^*) = 0 \qquad g_i(x^*) \geq 0$

$\mu_i^* g_i(x^*) = 0 \qquad \mu_i^* \geq 0$

2) Optimality Conditions: Optimization with "Inequality" Constraints

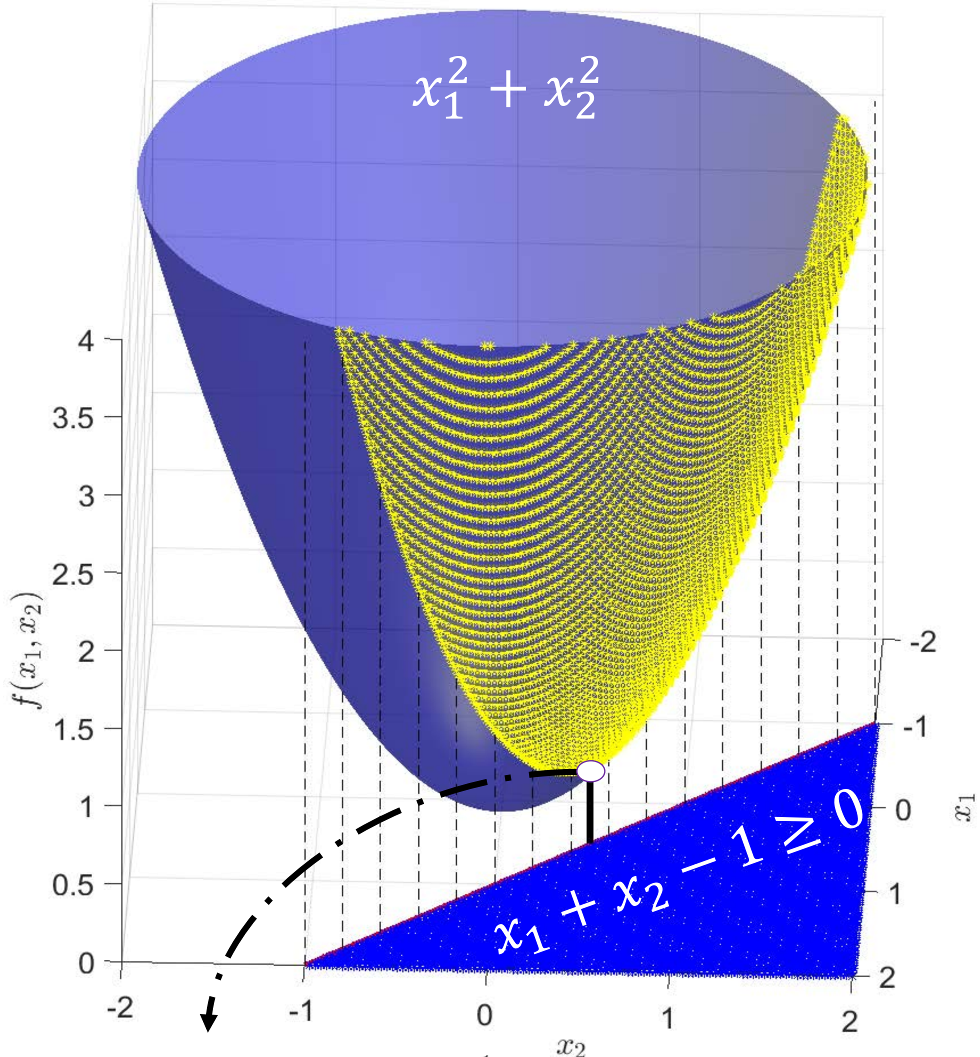
Example 1

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)$

Optimality Cond.

$$\nabla_x L(x^*, \mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0$$
$$\mu^* g(x^*) = 0$$



$$x_1^* = x_2^* = \frac{1}{2}, \quad \mu^* = 1$$

2) Optimality Conditions: Optimization with "Inequality" Constraints

Example 1

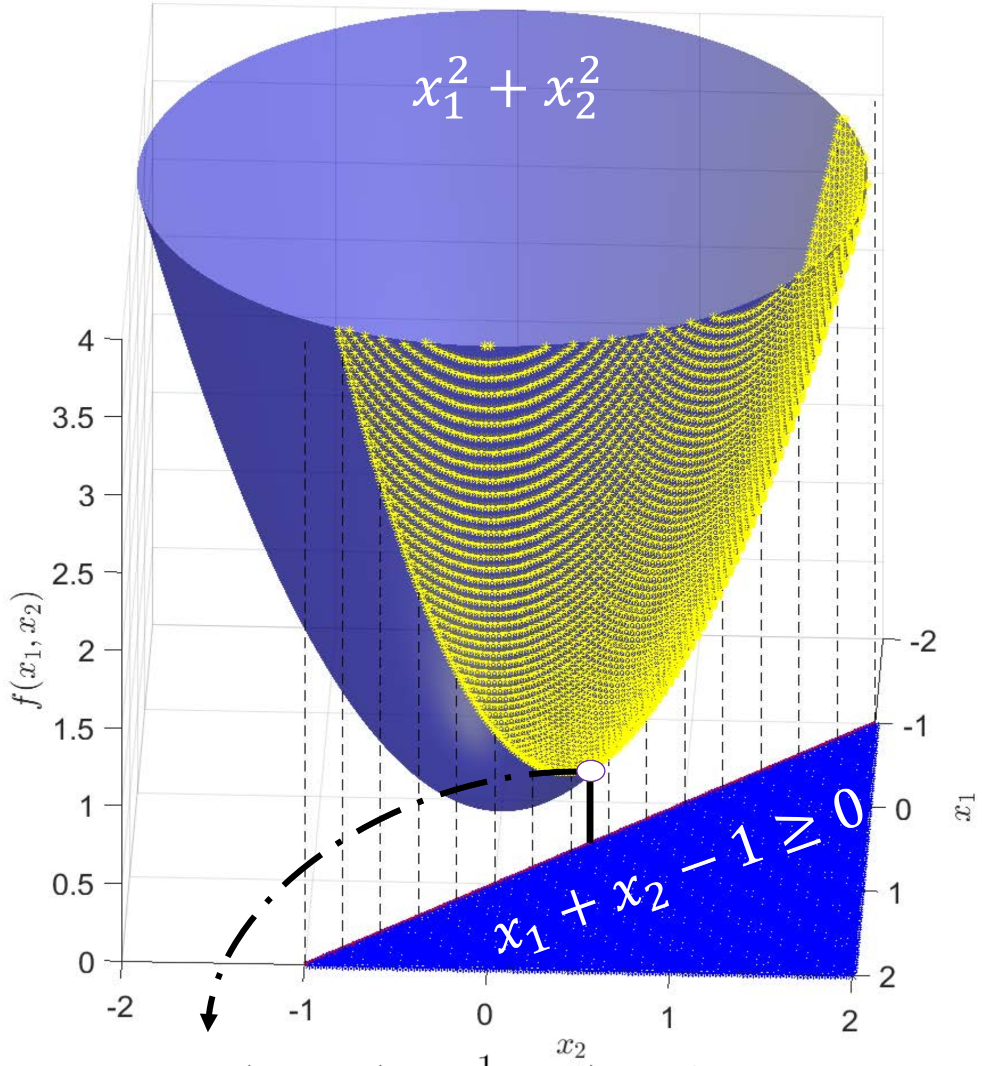
$$\min_x \; x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)$

Optimality Cond.

$$\boxed{\nabla_x L(x^*,\mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \quad \mu^* g(x^*) = 0}$$

$$\begin{cases} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \quad x_1^* + x_2^* - 1 \geq 0 \end{cases}$$



$$x_1^* = x_2^* = \frac{1}{2}, \quad \mu^* = 1$$

2) Optimality Conditions: Optimization with "Inequality" Constraints

Example 1

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)$

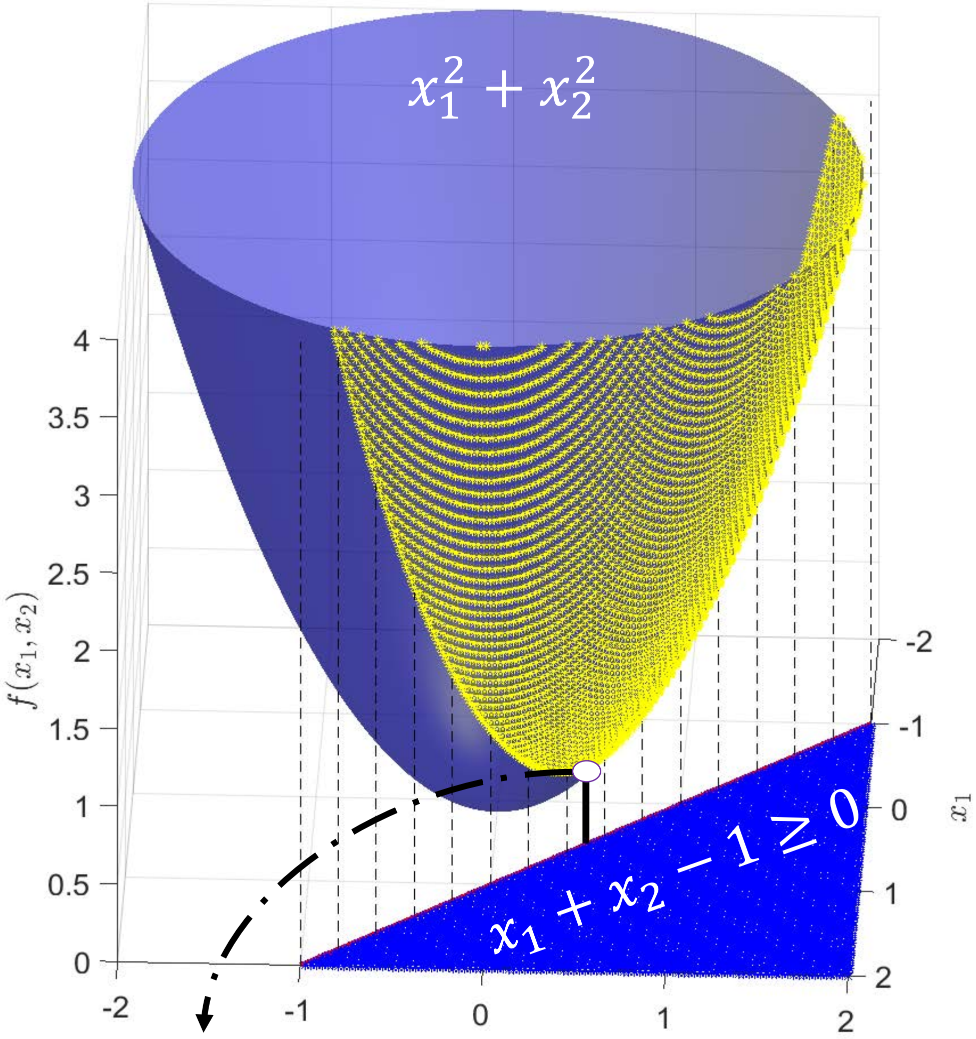
Optimality Cond.

$$\boxed{\nabla_x L(x^*,\mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \quad \mu^* g(x^*) = 0}$$

$$\begin{cases} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \quad x_1^* + x_2^* - 1 \geq 0 \end{cases}$$

Case 1

$$\begin{cases} \mu^* = 0 \Rightarrow x_1^* = x_2^* = 0 \\ x_1^* + x_2^* - 1 \geq 0 \end{cases}$$



$x_1^* = x_2^* = \frac{1}{2}, \; \mu^* = 1$

2) Optimality Conditions: Optimization with “Inequality” Constraints

Example 1

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)$

Optimality Cond.

$$\boxed{\begin{array}{l} \nabla_x L(x^*,\mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \\ \mu^* g(x^*) = 0 \end{array}}$$

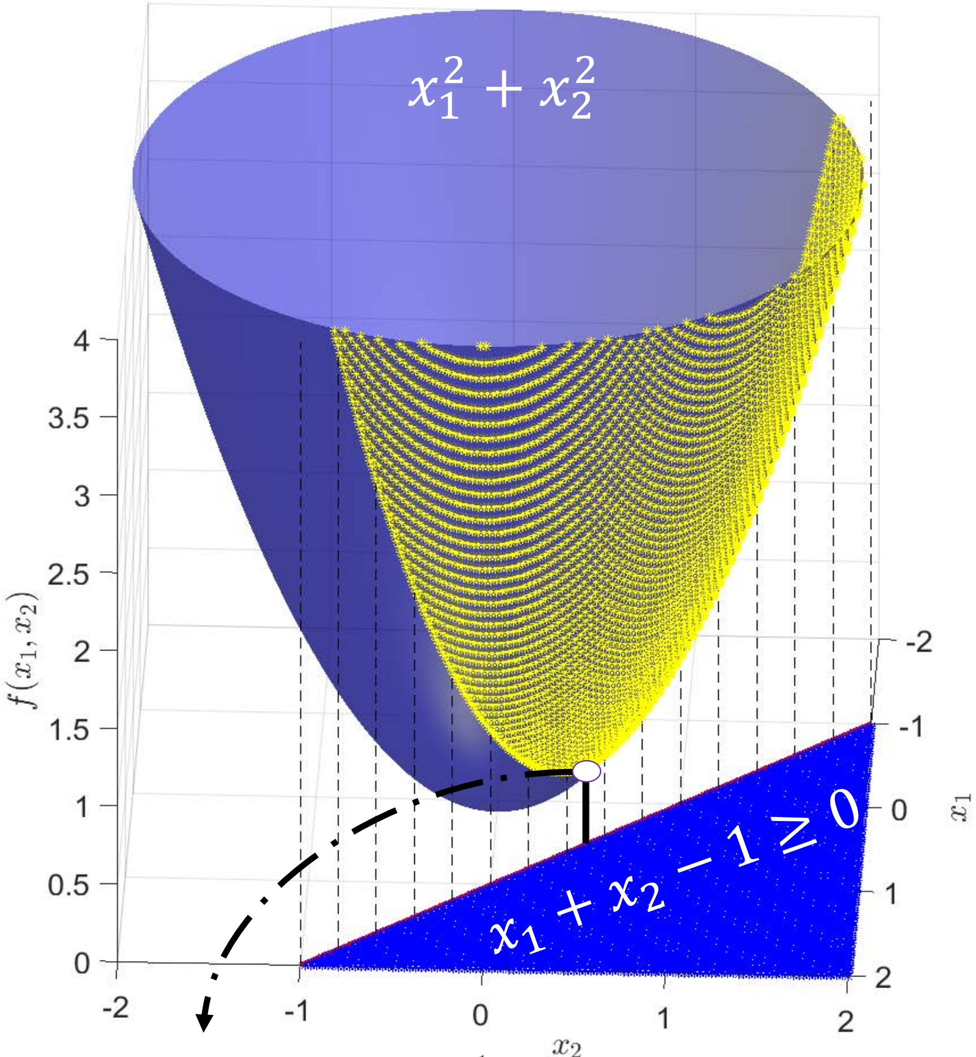
$$\left\{ \begin{array}{l} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \quad x_1^* + x_2^* - 1 \geq 0 \end{array} \right.$$

Case 1

$$\left\{ \begin{array}{l} \mu^* = 0 \Rightarrow x_1^* = x_2^* = 0 \\ x_1^* + x_2^* - 1 \geq 0 \end{array} \right.$$

Case 2

$$\boxed{\left\{ \begin{array}{l} x_1^* + x_2^* - 1 = 0 \\ \Downarrow \\ x_1^* = x_2^* = 1/2, \ \mu^* = 1 \end{array} \right.}$$



$$x_1^* = x_2^* = \frac{1}{2}, \quad \mu^* = 1$$

2) Optimality Conditions: Optimization with “Inequality” Constraints

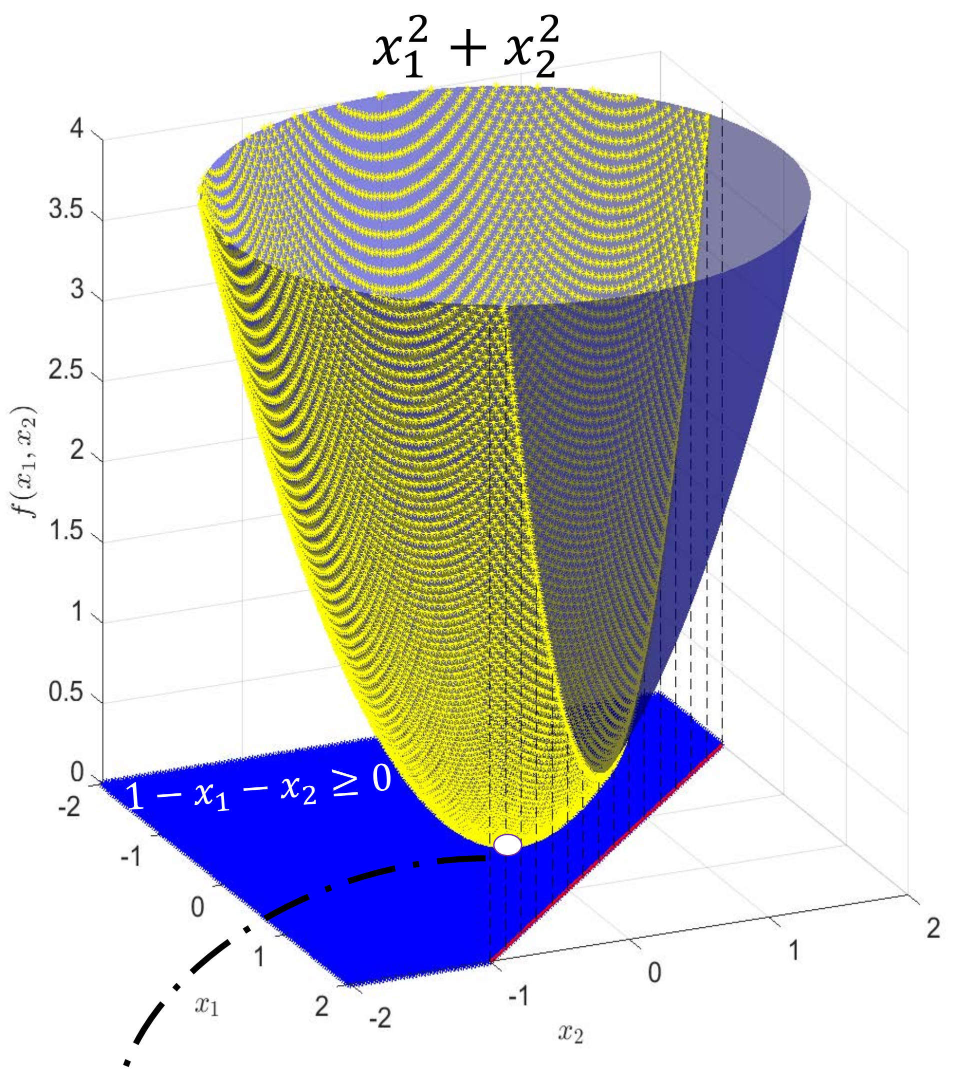
Example 2

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad 1 - x_1 - x_2 \geq 0$$

Lagrange function: $L(x, \mu) = (x_1^2 + x_2^2) - \mu(1 - x_1 - x_2 - 1)$

Optimality Cond.

$$\boxed{\nabla_x L(x^*, \mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \quad \mu^* g(x^*) = 0}$$



$$x_1^* = x_2^* = 0, \ \mu^* = 0$$

2) Optimality Conditions: Optimization with "Inequality" Constraints

Example 2

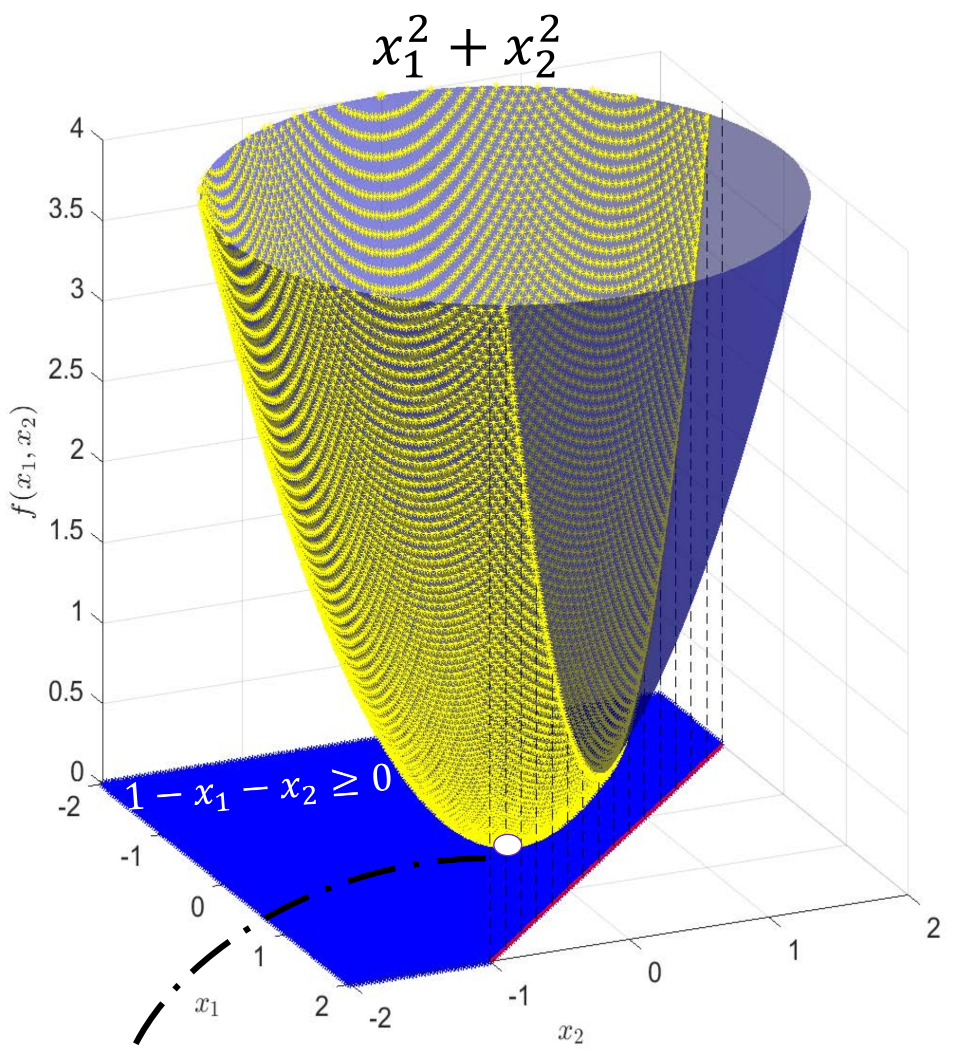
$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad 1 - x_1 - x_2 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(1 - x_1 - x_2 - 1)$

Optimality Cond.

$$\boxed{\nabla_x L(x^*,\mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0}$$

$$\begin{cases} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(1 - x_1^* - x_2^*) = 0 \\ \mu^* \geq 0 \quad 1 - x_1^* - x_2^* \geq 0 \end{cases}$$



$$x_1^* = x_2^* = 0, \quad \mu^* = 0$$

2) Optimality Conditions: Optimization with “Inequality” Constraints

Example 2

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad 1 - x_1 - x_2 \geq 0$$

Lagrange function: $L(x,\mu) = (x_1^2 + x_2^2) - \mu(1 - x_1 - x_2 - 1)$

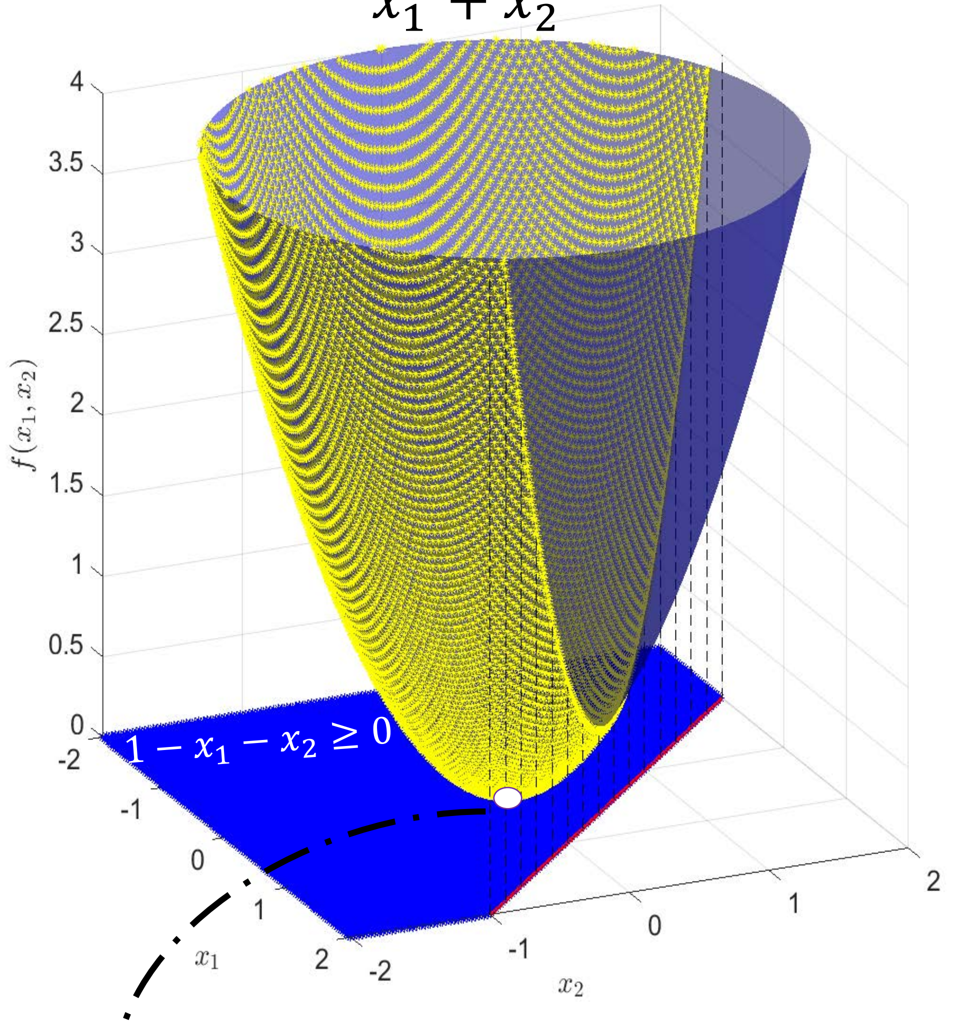
Optimality Cond.

$$\boxed{\nabla_x L(x^*, \mu^*) = 0 \quad g(x^*) \geq 0 \quad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0}$$

$$\left\{\begin{array}{l} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(1 - x_1^* - x_2^*) = 0 \\ \mu^* \geq 0 \quad 1 - x_1^* - x_2^* \geq 0 \end{array}\right.$$

$$\left\{\begin{array}{l} \mu^* = 0 \Rightarrow x_1^* = x_2^* = 0 \\ 1 - x_1^* - x_2^* \geq 0 \end{array}\right.$$

$x_1^2 + x_2^2$



$$x_1^* = x_2^* = 0, \quad \mu^* = 0$$

Optimization:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \;\; i = 1, \ldots, n_g \\ & h_i(x) = 0, \;\; i = 1, .., n_h \end{aligned}$$

Optimization:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0,\ \ i = 1, ..., n_g \\ & h_i(x) = 0,\ \ i = 1, .., n_h \end{aligned}$$

Lagrange function $L(x, \mu, \lambda) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$

Lagrange multiplier

Optimality Cond.

$$\nabla_x L(x, \mu, \lambda) = 0$$

$(\nabla_{\lambda_i} L(x, \lambda) = 0) \rightarrow$ $-h_i(x^*) = 0,\ \ i = 1, ..., n_h$

$$g_i(x^*) \geq 0, \quad i = 1, ..., n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, ..., n_g$$

Optimization:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \; i = 1, ..., n_g \\ & h_i(x) = 0, \; i = 1, .., n_h \end{aligned}$$

KKT (Karush-Kuhn-Tucker) Necessary Optimality Condition:

Lagrange function $L(x, \mu, \lambda) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$

Lagrange multiplier

Optimality Cond.

$\nabla_x L(x, \mu, \lambda) = 0$ *Stationarity*

$-h_i(x^*) = 0, \; i = 1, ..., n_h$

$g_i(x^*) \geq 0, \quad i = 1, ..., n_g$ *Primal Feasibility*

$\mu_i \geq 0$ *Dual Feasibility*

$\mu_i^* g_i(x^*) = 0, \quad i = 1, ..., n_g$ *Dual Complementary Slackness*

Optimization:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \; i = 1, ..., n_g \\ & h_i(x) = 0, \; i = 1, .., n_h \end{aligned}$$

KKT (Karush-Kuhn-Tucker) Necessary Optimality Condition:

Lagrange function $L(x,\mu,\lambda) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$

Lagrange multiplier

Note: we might also see the following form of the Lagrange function:

$$L(x,\mu,\lambda) = f(x) + \sum_{i=1}^{n_g} \mu_i g_i(x) + \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

i) Because inequality constraints of the original problem is in he form of $g_i(x) \leq 0$

ii) $\lambda : +/-$

Chapter 3.3: Inequality Constraints, "Nonlinear Programming" Dimitri Bertsekas, MIT,

Optimization

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0,\ i = 1, ..., n_g \\ & h_i(x) = 0,\ i = 1, .., n_h \end{aligned}$$

Optimality

Constrained Optimization:

$$\nabla_x L(x, \mu, \lambda) = 0$$

$$-h_i(x^*) = 0,\ i = 1, ..., n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, ..., n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, ..., n_g$$

system of nonlinear equations and inequalities

Unconstrained Optimization:

$$\nabla_x f(x) = 0$$

system of nonlinear equations

To find x^*, we need to solve a system of nonlinear equations and inequalities.

Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

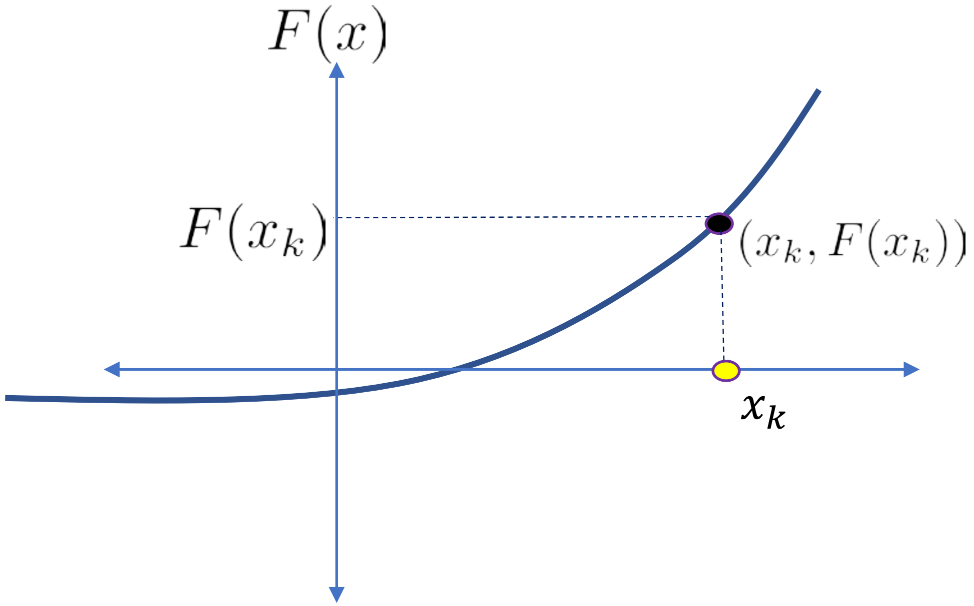
- We **can not** solve arbitrary *"nonlinear equations"* very easily.
- We **can** solve *"linear equations"* using the techniques from linear algebra.

Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary "*nonlinear equations*" very easily.
- We **can** solve "*linear equations*" using the techniques from linear algebra.
- We linearize the system of "*nonlinear equations*" with a first-order Taylor and solve the obtained "*linear system*", and then iterate.

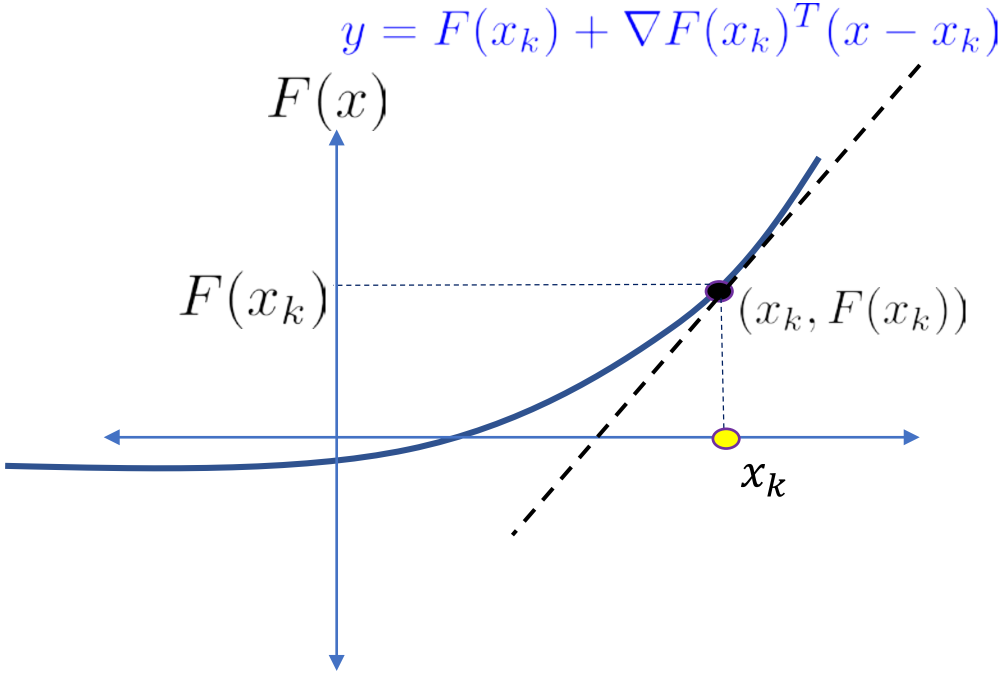
Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary *"nonlinear equations"* very easily.
- We **can** solve *"linear equations"* using the techniques from linear algebra.
- We linearize the system of *"nonlinear equations"* with a first-order Taylor and solve the obtained *"linear system"*, and then iterate.



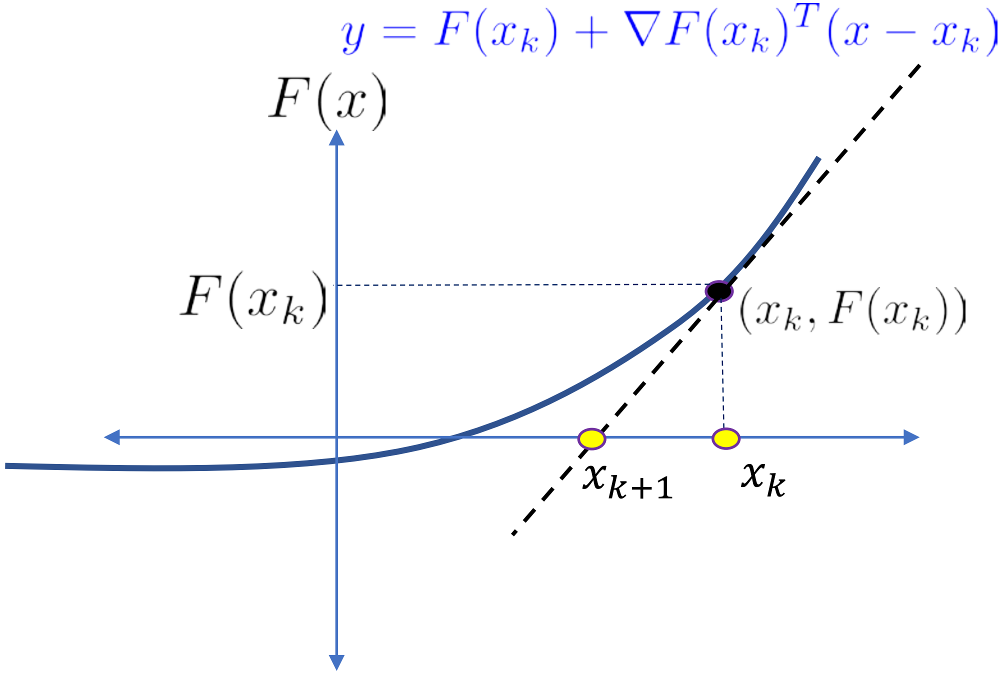
Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary "*nonlinear equations*" very easily.
- We **can** solve "*linear equations*" using the techniques from linear algebra.
- We linearize the system of "*nonlinear equations*" with a first-order Taylor and solve the obtained "*linear system*", and then iterate.



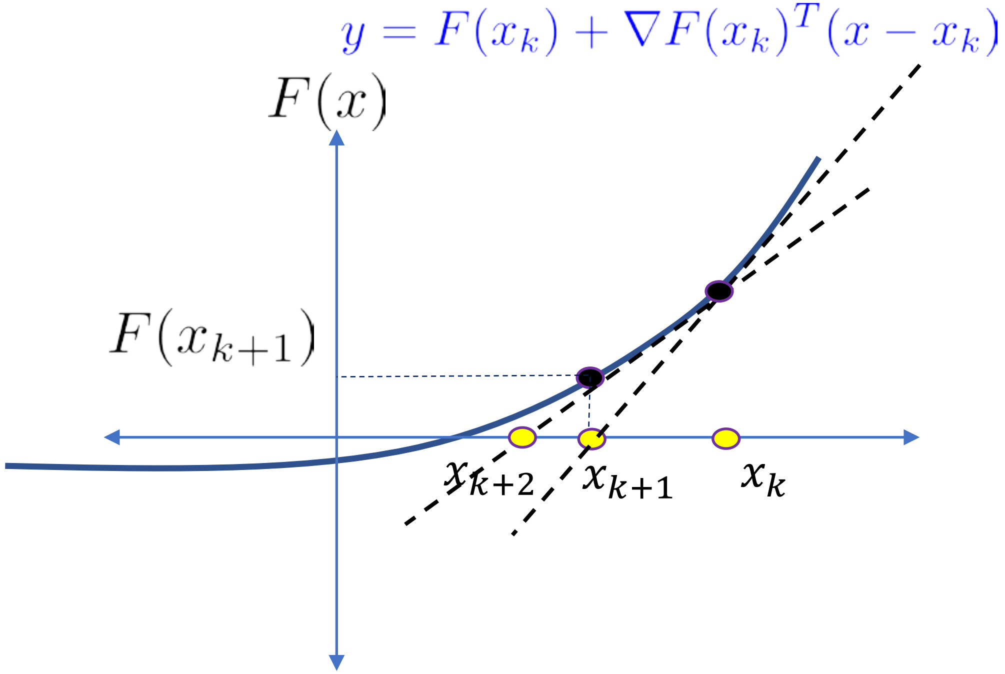
Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary "*nonlinear equations*" very easily.
- We **can** solve "*linear equations*" using the techniques from linear algebra.
- We linearize the system of "*nonlinear equations*" with a first-order Taylor and solve the obtained "*linear system*", and then iterate.



Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary "*nonlinear equations*" very easily.
- We **can** solve "*linear equations*" using the techniques from linear algebra.
- We linearize the system of "*nonlinear equations*" with a first-order Taylor and solve the obtained "*linear system*", and then iterate.



Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary *"nonlinear equations"* very easily.
- We **can** solve *"linear equations"* using the techniques from linear algebra.
- We linearize the system of *"nonlinear equations"* with a first-order Taylor and solve the obtained *"linear system"*, and then iterate.

(0) *k=0,*

(1) *Linear approximation at point* x_k:

$$y = F(x_k) + \nabla F(x_k)^T (x - x_k)$$

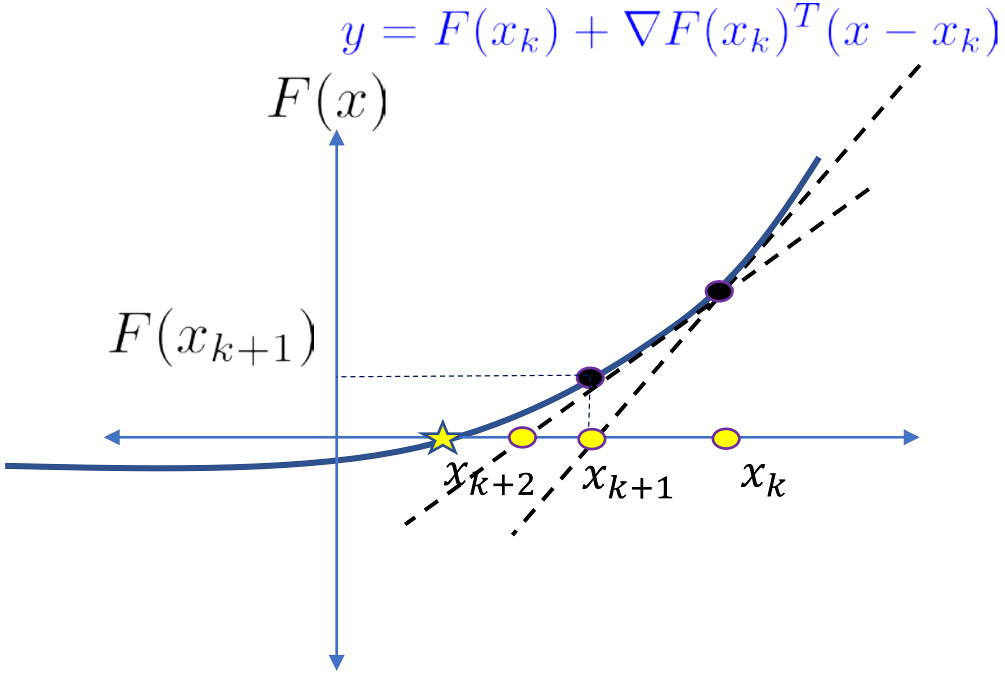
(2) *Solve linear system:*

$$y = 0 \longrightarrow F(x_k) + \nabla F(x_k)'(x - x_k) = 0$$

$$\Rightarrow (x - x_k) = -(\nabla F(x_k))^{-1} F(x_k)$$

$$\Rightarrow \boxed{x_{k+1} = x_k - (\nabla F(x_k))^{-1} F(x_k)}$$

(3) *Go to Step* (1)



Newton's Method: Solving a system of nonlinear equations $F(x) = 0$

- We **can not** solve arbitrary "*nonlinear equations*" very easily.
- We **can** solve "*linear equations*" using the techniques from linear algebra.
- We linearize the system of "*nonlinear equations*" with a first-order Taylor and solve the obtained "*linear system*", and then iterate.

(0) *k=0,*

(1) *Linear approximation at point* x_k:

$$y = F(x_k) + \nabla F(x_k)^T (x - x_k)$$

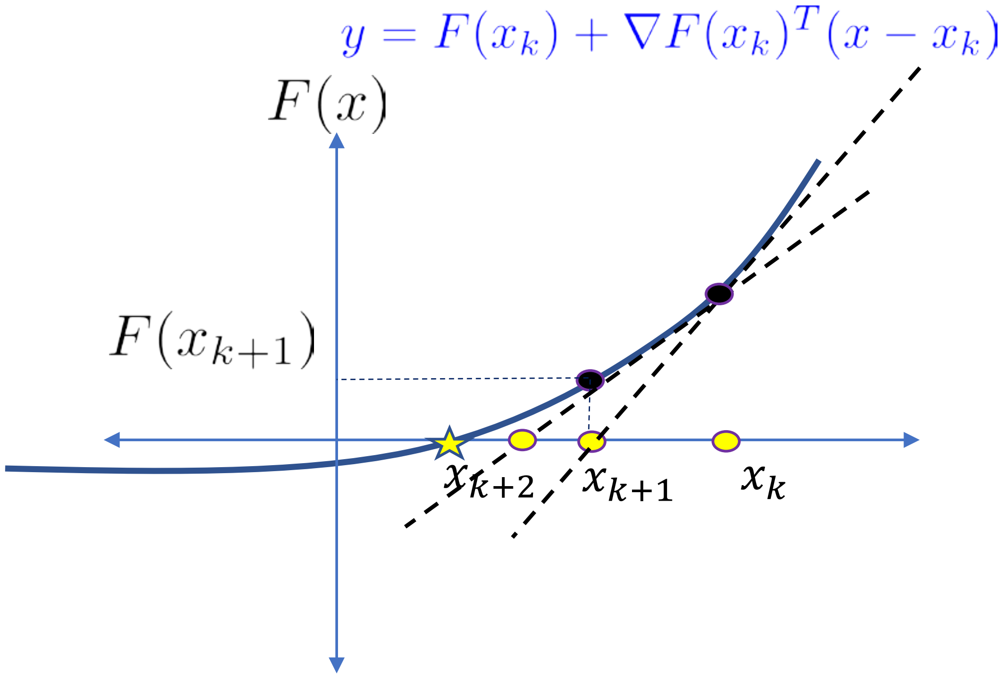
(2) *Solve linear system:*

$$y = 0 \longrightarrow F(x_k) + \nabla F(x_k)'(x - x_k) = 0$$

⇨ (1) *Linear approximation at point* x_k:

⇨ $$\boxed{x_{k+1} = x_k - (\nabla F(x_k))^{-1} F(x_k)}$$

(3) *Go to Step* (1)



system of nonlinear equations $F(x) = 0$ ***Update at each iteration:*** $x_{k+1} = x_k - \mathbf{J}_F^{-1} F(x_k)$

Jacobian matrix of $\boldsymbol{F(x)}$

1) Newton's Method for Unconstrained optimization

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

- **Optimality Condition:** $F(x) = 0 \longrightarrow \nabla f(x) = 0$

  system of nonlinear equations

- **Newton's method:** $\longrightarrow$

$$x_{k+1} = x_k - (\nabla F(x_k))^{-1} F(x_k)$$

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

$$\boxed{x_{k+1} = x_k - \mathbf{H}_f^{-1} \nabla f(x_k)}$$

1) Newton's Method for Unconstrained optimization

Example

$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_1 x_2 + \frac{1}{2}x_2^2$$

$$F(x) = \nabla f(x) = \begin{bmatrix} 2x_1 + x_2 \\ x_2 + x_1 \end{bmatrix}$$

1) Newton's Method for Unconstrained optimization

Example $$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_1 x_2 + \frac{1}{2}x_2^2$$

$$F(x) = \nabla f(x) = \begin{bmatrix} 2x_1 + x_2 \\ x_2 + x_1 \end{bmatrix}$$

$$\left\{ \begin{array}{l} x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \\ \\ \nabla f(x_k) = \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix} \qquad \nabla^2 f(x_k) = \mathbf{H}_f = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \end{array} \right\} \Rightarrow$$

$$\begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix}$$

$$\begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix}$$

1) Newton's Method for Unconstrained optimization

Example

$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_1x_2 + \frac{1}{2}x_2^2$$

$$F(x) = \nabla f(x) = \begin{bmatrix} 2x_1 + x_2 \\ x_2 + x_1 \end{bmatrix} = 0$$

$$\left\{ \begin{array}{l} x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1}\nabla f(x_k) \\[2ex] \nabla f(x_k) = \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix} \qquad \nabla^2 f(x_k) = \mathbf{H}_f = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \end{array} \right\} \Rightarrow$$

$$\begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix}$$

$$\begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2x_{1k} + x_{2k} \\ x_{2k} + x_{1k} \end{bmatrix}$$

$$\begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \xrightarrow{\text{Step 1}} \begin{bmatrix} x_{1k+1} \\ x_{2k+1} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \longrightarrow \left\{ \begin{array}{l} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\[2ex] F(x) = \nabla f(x) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{array} \right.$$

2) Newton's Method for Constrained optimization with Equality Constraints

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0,\ i = 1, \ldots, n_h \end{aligned}$$

- **Optimality Condition:** $F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ \nabla_\lambda L(x,\lambda) \end{bmatrix} = 0$

system of nonlinear equations

($(n + n_h)$equations, $(n + n_h)$variables)

2) Newton's Method for Constrained optimization with Equality Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \; i = 1, \ldots, n_h \end{aligned}$$

- **Optimality Condition:** $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

system of nonlinear equations
(($n + n_h$)equations, ($n + n_h$)variables)

Lagrange function

$$L(x, \lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

2) Newton's Method for Constrained optimization with Equality Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \ i = 1, \dots, n_h \end{aligned}$$

- **Optimality Condition:** $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

system of nonlinear equations
($(n + n_h)$equations, $(n + n_h)$variables)

Lagrange function

$$L(x, \lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x) = f(x) - \lambda^T h(x)$$

$$h(x) = [h_1(x), \dots, h_{n_h}(x)]^T$$

2) Newton's Method for Constrained optimization with Equality Constraints

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \; i = 1, \dots, n_h \end{aligned}$$

- **Optimality Condition:** $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

system of nonlinear equations
($(n + n_h)$equations, $(n + n_h)$variables)

Lagrange function

$$L(x, \lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x) = f(x) - \lambda^T h(x)$$

$$h(x) = [h_1(x), \dots, h_{n_h}(x)]^T$$

- **Newton's method:** $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$

2) Newton's Method for Constrained optimization with Equality Constraints

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h_i(x) = 0, \; i = 1, \dots, n_h$$

- **Optimality Condition:** $F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ \nabla_\lambda L(x,\lambda) \end{bmatrix} = 0$

  system of nonlinear equations

  ($(n + n_h)$equations, $(n + n_h)$variables)

Lagrange function

$$L(x,\lambda) = f(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x) = f(x) - \lambda^T h(x)$$

$$h(x) = [h_1(x), \dots, h_{n_h}(x)]^T$$

- **Newton's method:** $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$

$$F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ \nabla_\lambda L(x,\lambda) \end{bmatrix} = \begin{bmatrix} \nabla_x f(x) - \mathbf{J}_h^T \lambda \\ -h(x) \end{bmatrix}$$

$$\mathbf{J}_F = \begin{bmatrix} [\nabla F_1(x)]^T \\ [\nabla F_2(x)]^T \end{bmatrix} = \begin{bmatrix} [\nabla(\nabla_x L(x,\lambda))]^T \\ [\nabla(-h(x))]^T \end{bmatrix} = \begin{bmatrix} [\nabla_{xx}^2 L(x,\lambda) & -\mathbf{J}_h^T] \\ [-\mathbf{J}_h & 0] \end{bmatrix}$$

$$\boxed{\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \begin{bmatrix} \nabla_{xx}^2 L(x,\lambda) & -\mathbf{J}_h^T \\ -\mathbf{J}_h & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_x L(x,\lambda) \\ -h(x) \end{bmatrix}}$$

2) Newton's Method for Constrained optimization with Equality Constraints

Example

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 1 = 0 \end{aligned}$$

Optimality Condition: $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

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

2) Newton's Method for Constrained optimization with Equality Constraints

Example

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_2^2$$
$$\text{subject to} \quad x_1 + x_2 - 1 = 0$$

Optimality Condition: $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

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

- **Newton's method:** $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$

2) Newton's Method for Constrained optimization with Equality Constraints

Example

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \quad x_1^2 + x_2^2$$
$$\text{subject to} \quad x_1 + x_2 - 1 = 0$$

Optimality Condition: $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ \nabla_\lambda L(x,\lambda) \end{bmatrix} = 0$

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

- **Newton's method:** $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$

$$F(x,\lambda) = \begin{bmatrix} \nabla_x L(x,\lambda) \\ h(x) \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix} \\ -(x_1 + x_2 - 1) \end{bmatrix}$$

$$\mathbf{J}_F = \begin{bmatrix} [\nabla F_1(x)]^T \\ [\nabla F_2(x)]^T \end{bmatrix} = \begin{bmatrix} [\nabla \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix}]^T \\ [\nabla(-x_1 - x_2 + 1)]^T \end{bmatrix} = \begin{bmatrix} [2 & 0 & -1] \\ [0 & 2 & -1] \\ [-1 & -1 & 0] \end{bmatrix}$$

$$\begin{bmatrix} x_{1\,k+1} \\ x_{2\,k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_{1\,k} \\ x_{2\,k} \\ \lambda_k \end{bmatrix} - \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \begin{bmatrix} 2x_{1\,k} - \lambda_k \\ 2x_{2\,k} - \lambda_k \end{bmatrix} \\ -(x_{1\,k} + x_{2\,k} - 1) \end{bmatrix}$$

2) Newton's Method for Constrained optimization with Equality Constraints

Example

$$\begin{aligned} \underset{x \in \mathbb{R}^2}{\text{minimize}} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 + x_2 - 1 = 0 \end{aligned}$$

Optimality Condition: $F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{bmatrix} = 0$

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

➢ **Newton's method:** $\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$

$$F(x, \lambda) = \begin{bmatrix} \nabla_x L(x, \lambda) \\ h(x) \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix} \\ -(x_1 + x_2 - 1) \end{bmatrix}$$

$$\mathbf{J}_F = \begin{bmatrix} [\nabla F_1(x)]^T \\ [\nabla F_2(x)]^T \end{bmatrix} = \begin{bmatrix} [\nabla \begin{bmatrix} 2x_1 - \lambda \\ 2x_2 - \lambda \end{bmatrix}]^T \\ [\nabla(-x_1 - x_2 + 1)]^T \end{bmatrix} = \begin{bmatrix} [2 & 0 & -1] \\ [0 & 2 & -1] \\ [-1 & -1 & 0] \end{bmatrix}$$

$$\boxed{\begin{bmatrix} x_{1k+1} \\ x_{2k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_{1k} \\ x_{2k} \\ \lambda_k \end{bmatrix} - \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \begin{bmatrix} 2x_{1k} - \lambda_k \\ 2x_{2k} - \lambda_k \end{bmatrix} \\ -(x_{1k} + x_{2k} - 1) \end{bmatrix}}$$

$$\begin{bmatrix} x_{1k} \\ x_{2k} \\ \lambda_k \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \xrightarrow{\text{Step 1}} \begin{bmatrix} x_{1k+1} \\ x_{2k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{4} & -\frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \longrightarrow$$

$$\boxed{\begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \quad F(x, \lambda) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = 0}$$

3) Newton's Method for Optimization with Inequality Constraints

- Newton's method deal with solving a system of nonlinear equations $F(x) = 0$

3) Newton's Method for Optimization with Inequality Constraints

- Newton's method deal with solving a system of nonlinear equations $F(x) = 0$

- Optimality condition for **constrained optimization** (involving inequality constraint) is a "*system of nonlinear equations and inequalities*".

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad \boxed{g_i(x) \geq 0, \ i = 1, \dots, n_g} \longrightarrow$$
$$h_i(x) = 0, \ i = 1, .., n_h$$

KKT:

$$\nabla_x L(x, \mu, \lambda) = 0$$
$$-h_i(x^*) = 0, \ i = 1, \dots, n_h$$
$$\boxed{\begin{array}{l} g_i(x^*) \geq 0, \quad i = 1, \dots, n_g \\ \mu_i \geq 0 \end{array}}$$
$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \dots, n_g$$

3) Newton's Method for Optimization with Inequality Constraints

- Newton's method deal with solving a system of nonlinear equations $F(x) = 0$

- Optimality condition for **constrained optimization** (involving inequality constraint) is a "*system of nonlinear equations and inequalities*".

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & \boxed{g_i(x) \geq 0, \; i = 1, ..., n_g} \\ & h_i(x) = 0, \; i = 1, .., n_h \end{aligned}$$

KKT:

$$\begin{aligned} & \nabla_x L(x, \mu, \lambda) = 0 \\ & -h_i(x^*) = 0, \; i = 1, ..., n_h \\ & \boxed{\begin{aligned} & g_i(x^*) \geq 0, \quad i = 1, ..., n_g \\ & \mu_i \geq 0 \end{aligned}} \\ & \mu_i^* g_i(x^*) = 0, \quad i = 1, ..., n_g \end{aligned}$$

Interior Point Method

- Reformulate optimization problem with Inequality Constraints as optimization with only equality constraints.
- Apply Newton's method to optimality *system of nonlinear equations* .

Interior Point Method

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \;\; i = 1, \dots, n_g \\ & h_i(x) = 0, \;\; i = 1, \dots, n_h \end{aligned}$$

Slack variables "s_i"

$$\begin{aligned} \underset{x\in\mathbb{R}^n, s\in\mathbb{R}^{n_g}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) - s_i = 0, \;\; i = 1, \dots, n_g \\ & s_i \geq 0, \;\; i = 1, \dots, n_g \\ & h_i(x) = 0, \;\; i = 1, \dots, n_h \end{aligned}$$

Interior Point Method

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \;\; i = 1, \ldots, n_g \\ & h_i(x) = 0, \;\; i = 1, \ldots, n_h \end{aligned}$$

Slack variables "s_i"

$$\begin{aligned} \underset{x \in \mathbb{R}^n, s \in \mathbb{R}^{n_g}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) - s_i = 0, \; i = 1, \ldots, n_g \\ & s_i \geq 0, \; i = 1, \ldots, n_g \\ & h_i(x) = 0, \;\; i = 1, \ldots, n_h \end{aligned}$$

$$z = [s_i|_{i=1}^{n_g}, x_i|_{i=1}^{n}]^T$$

$$\bar{f}(z) = f(x)$$

$$\bar{h}(z) = [g(x)_i - s_i \;|_{i=1}^{n_g}, h_i(x) \;|_{i=1}^{n_h}]^T$$

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) \\ \text{subject to} \quad & \bar{h}(z) = 0 \\ & z_i \geq 0, \;\; i = 1, \ldots, n_g \end{aligned}$$

Interior Point Method

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0, \;\; i = 1, ..., n_g \\ & h_i(x) = 0, \;\; i = 1, ..., n_h \end{aligned}$$

Slack variables "s_i"

$$\begin{aligned} \underset{x\in\mathbb{R}^n, s\in\mathbb{R}^{n_g}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) - s_i = 0, \;\; i = 1, ..., n_g \\ & s_i \geq 0, \;\; i = 1, ..., n_g \\ & h_i(x) = 0, \;\; i = 1, ..., n_h \end{aligned}$$

$$z = [s_i|_{i=1}^{n_g}, x_i|_{i=1}^{n}]^T$$

$$\bar{f}(z) = f(x)$$

$$\bar{h}(z) = [g(x)_i - s_i \;|_{i=1}^{n_g}, h_i(x) \;|_{i=1}^{n_h}]^T$$

$$\begin{aligned} \underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) \\ \text{subject to} \quad & \bar{h}(z) = 0 \\ & z_i \geq 0, \;\; i = 1, ..., n_g \end{aligned}$$

Optimization with equality constraints

$$\begin{aligned} \underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) + \text{Penalty Function}(z_i) \\ \text{subject to} \quad & \bar{h}(z) = 0 \end{aligned}$$

Use Newton's Method

Interior Point Method

$$
\begin{aligned}
\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) \\
\text{subject to} \quad & \bar{h}(z) = 0 \\
& z_i \geq 0, \;\; i = 1, ..., n_g
\end{aligned}
$$

Optimization with equality constraints

$$
\begin{aligned}
\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) + \text{Penalty Function}(z_i) \\
\text{subject to} \quad & \bar{h}(z) = 0
\end{aligned}
$$

$$
\begin{aligned}
\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) + \sum_{i=1}^{n_g} I_i(z_i) \\
\text{subject to} \quad & \bar{h}(z) = 0
\end{aligned}
$$

Interior Point Method

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) \\ \text{subject to} \quad & h(z) = 0 \\ & z_i \geq 0, \; i = 1, \dots, n_g \end{aligned}$$

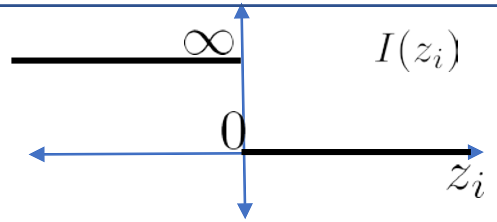
Optimization with equality constraints

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) + \text{Penalty Function}(z_i) \\ \text{subject to} \quad & h(z) = 0 \end{aligned}$$

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) + \sum_{i=1}^{n_g} I_i(z_i) \\ \text{subject to} \quad & \bar{h}(z) = 0 \end{aligned}$$

Penalty function for constraint $z_i \geq 0$

Indicator function $I(z_i) = \begin{cases} 0 & z_i \geq 0 \\ \infty & z_i < 0 \end{cases}$



Interior Point Method

$$\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z)$$
$$\text{subject to} \quad h(z) = 0$$
$$z_i \geq 0, \; i = 1, \ldots, n_g$$

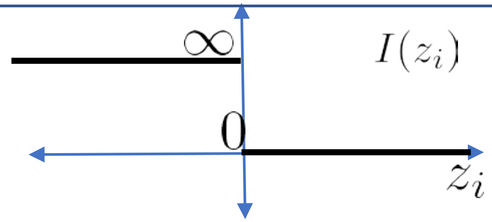
Optimization with equality constraints

$$\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) + \text{Penalty Function}(z_i)$$
$$\text{subject to} \quad h(z) = 0$$

$$\underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) + \sum_{i=1}^{n_g} I_i(z_i)$$
$$\text{subject to} \quad \bar{h}(z) = 0$$

Penalty function for constraint $z_i \geq 0$

Indicator function $I(z_i) = \begin{cases} 0 & z_i \geq 0 \\ \infty & z_i < 0 \end{cases}$



- Indicator function not smooth and therefore not differentiable.

Interior Point Method

$$\underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z)$$
$$\text{subject to} \quad h(z) = 0$$
$$z_i \geq 0,\ \ i = 1, \ldots, n_g$$

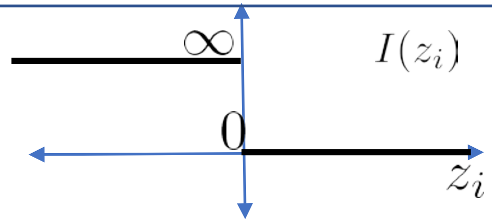
Optimization with equality constraints

$$\underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) + \text{Penalty Function}(z_i)$$
$$\text{subject to} \quad h(z) = 0$$

$$\underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) + \sum_{i=1}^{n_g} I_i(z_i)$$
$$\text{subject to} \quad \bar{h}(z) = 0$$

Penalty function for constraint $z_i \geq 0$

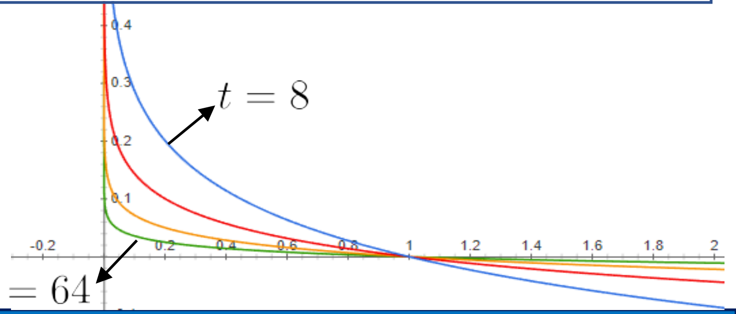
Indicator function $I(z_i) = \begin{cases} 0 & z_i \geq 0 \\ \infty & z_i < 0 \end{cases}$



- Indicator function not smooth and therefore not differentiable.
- ➢ It is approximated by the logarithmic barrier function.

$$-\frac{1}{t}ln(z_i) \rightarrow I(z_i)$$
$$t \rightarrow \infty$$

$$-\frac{1}{t}ln(z_i)$$



Interior Point Method

Optimization with equality constraints

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) - \sum_{i=1}^{n_g} \frac{1}{t} ln(z_i) \\ \text{subject to} \quad & \bar{h}(z) = 0 \end{aligned}$$

Interior Point Method

Optimization with equality constraints

$$\underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) - \sum_{i=1}^{n_g} \frac{1}{t} ln(z_i)$$
$$\text{subject to} \quad \bar{h}(z) = 0$$

Use Newton's Method

(0) *k=0, feasible* z_k , t_k

(1) Inner Loop: *Compute* $z^*\ (k+1)$ *by solving constrained optimization using newton's method.*

Lagrange function $L(z,\lambda) = \bar{f}(z) - \frac{1}{t}\sum_{i=1}^{n_g} ln(z_i) - \lambda^T \bar{h}(z)$ → *Lagrange multipliers*

Optimality Condition: $F(z,\lambda) = \begin{bmatrix} \nabla_z L(z,\lambda) \\ \nabla_\lambda L(z,\lambda) \end{bmatrix} = 0 \longrightarrow \boxed{\begin{bmatrix} z_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} z_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(z_k, \lambda_k)}$

(2) *Outer Loop:* *Increase* $t_{k+1} = \beta t_k$

(3) $k \leftarrow k+1$, Go to step (1)

Interior Point Method

Optimization with equality constraints

$$\begin{aligned} \underset{z \in \mathbb{R}^{n+n_g}}{\text{minimize}} \quad & \bar{f}(z) - \sum_{i=1}^{n_g} \frac{1}{t} ln(z_i) \\ \text{subject to} \quad & \bar{h}(z) = 0 \end{aligned}$$

Use Newton's Method

(0) *k=0, feasible* z_k , t_k

(1) Inner Loop: *Compute* $z^*\ (k+1)$ *by solving constrained optimization using newton's method.*

Lagrange function $L(z,\lambda) = \bar{f}(z) - \frac{1}{t}\sum_{i=1}^{n_g} ln(z_i) - \lambda^T \bar{h}(z)$ *Lagrange multipliers*

Optimality Condition: $F(z,\lambda) = \begin{bmatrix} \nabla_z L(z,\lambda) \\ \nabla_\lambda L(z,\lambda) \end{bmatrix} = 0 \longrightarrow \boxed{\begin{bmatrix} z_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} z_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(z_k,\lambda_k)}$

(2) *Outer Loop:* *Increase* $t_{k+1} = \beta t_k$

(3) $k \leftarrow k+1$, Go to step (1)

- Every limit point of a sequence $\{z_k\}$ generated by a barrier method solves the original constrained problem.
- For more information and convergence analysis see:
  - Stephen J. Wright, "On the convergence of the Newton/log-barrier method", Mathematical Programming , Volume 90, Issue 1, pp 71–100, 2001.
  - A. Forsgren, P. E. Gill, M. H. Wright "Interior Methods for Nonlinear Optimization", SIAM REVIEW, Vol. 44, No. 4, pp. 525–597, 2002.
  - Chapter 4.1: Barrier and Interior Point Methods, "Nonlinear Programming" Dimitri Bertsekas, MIT.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Optimality Condition

$$\nabla_x f(x) = 0$$

Newton's Method

$$x_{k+1} = x_k - \mathbf{H}_f^{-1} \nabla f(x_k)$$

1

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Optimality Condition

$$\nabla_x f(x) = 0$$

Newton's Method

$$x_{k+1} = x_k - \mathbf{H}_f^{-1} \nabla f(x_k)$$

2

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\text{subject to} \quad h_i(x) = 0,\ i = 1, \ldots, n_h$$

Optimality Condition

$$F = \begin{cases} \nabla_x L(x, \mu, \lambda) = 0 \\ -h_i(x^*) = 0,\ i = 1, \ldots, n_h \end{cases}$$

Lagrange Function

$$L(x, \lambda) = f(x) - \lambda^T h(x)$$

Newton's Method

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$$

$$\Downarrow$$

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \begin{bmatrix} \nabla^2_{xx} L(x, \lambda) & -\mathbf{J}_h^T \\ -\mathbf{J}_h & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_x L(x, \lambda) \\ -h(x) \end{bmatrix}$$

1

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

Optimality Condition

$$\nabla_x f(x) = 0$$

Newton's Method Step

$$x_{k+1} = x_k - \mathbf{H}_f^{-1}\nabla f(x_k)$$

2

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad h_i(x) = 0, \; i = 1, \ldots, n_h$$

Optimality Condition

$$F = \begin{cases} \nabla_x L(x,\mu,\lambda) = 0 \\ -h_i(x^*) = 0, \; i = 1, \ldots, n_h \end{cases}$$

Lagrange Function

$$L(x,\lambda) = f(x) - \lambda^T h(x)$$

Newton's Method Step

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \mathbf{J}_F^{-1} F(x_k, \lambda_k)$$

$$\Downarrow$$

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} - \begin{bmatrix} \nabla^2_{xx} L(x,\lambda) & -\mathbf{J}_h^T \\ -\mathbf{J}_h & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla_x L(x,\lambda) \\ -h(x) \end{bmatrix}$$

3

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad g_i(x) \geq 0, \; i = 1, \ldots, n_g$$
$$h_i(x) = 0, \; i = 1, \ldots, n_h$$

***Optimality Condition* (KKT)**

$$\nabla_x L(x,\mu,\lambda) = 0 \qquad -h_i(x^*) = 0, \; i = 1, \ldots, n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, \ldots, n_g$$

$$\mu_i^* \geq 0 \quad \mu_i^* g_i(x^*) = 0, \quad i = 1, \ldots, n_g$$

Optimization with equality constraints

$$z = [s_i|_{i=1}^{n_g}, x_i|_{i=1}^{n}]^T \qquad \bar{f}(z) = f(x)$$

$$\bar{h}(z) = [g(x)_i - s_i \;|_{i=1}^{n_g}, h_i(x)\;|_{i=1}^{n_h}]^T$$

$$\underset{z\in\mathbb{R}^{n+n_g}}{\text{minimize}} \quad \bar{f}(z) - \sum_{i=1}^{n_g} \frac{1}{t} ln(z_i)$$
$$\text{subject to} \quad \bar{h}(z) = 0$$

$$\Downarrow$$

2

Nonconvex Vs Convex Optimization

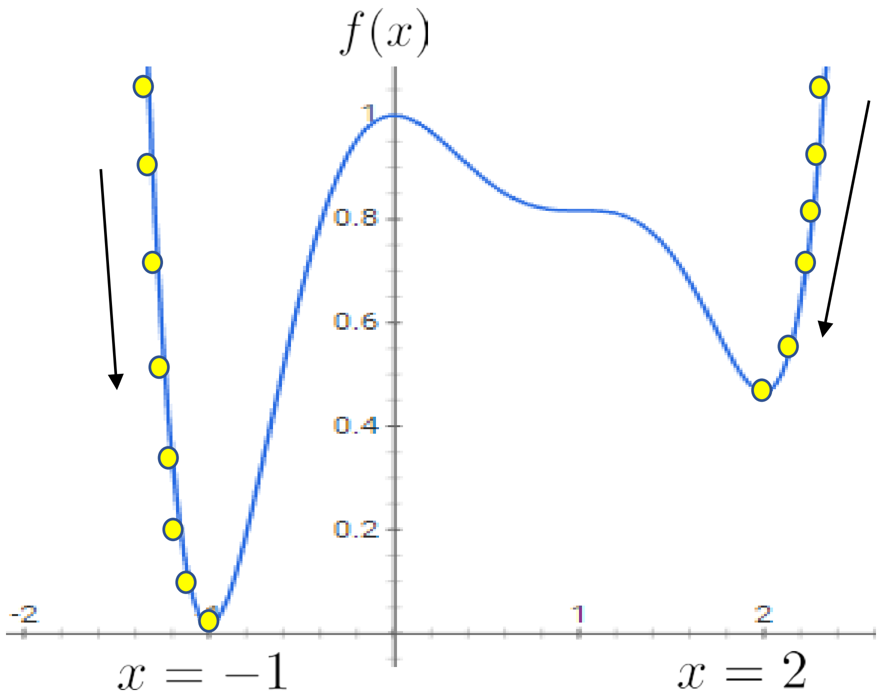
Example

$$\underset{x}{\text{minimize}} \quad 1 - x^2 + x^3 + \frac{1}{4}x^4 - \frac{3}{5}x^5 + \frac{1}{6}x^6$$
$$\text{subject to} \quad -3 \leq x \leq -3$$

$\boldsymbol{x_0 = -3}$ $\boldsymbol{f(x_0) = 252.5}$

$\boldsymbol{f(x_k) =}$

23.595
13.785
1.7330
1.0507
0.58202
0.48708
0.46764
0.46680
0.46669
0.46666



$\boldsymbol{x_0 = 3}$ $\boldsymbol{f(x_0) = 14.95}$

$\boldsymbol{f(x_k) =}$

0.4.669724
0.4.667130
0.4.666669
0.4.666667

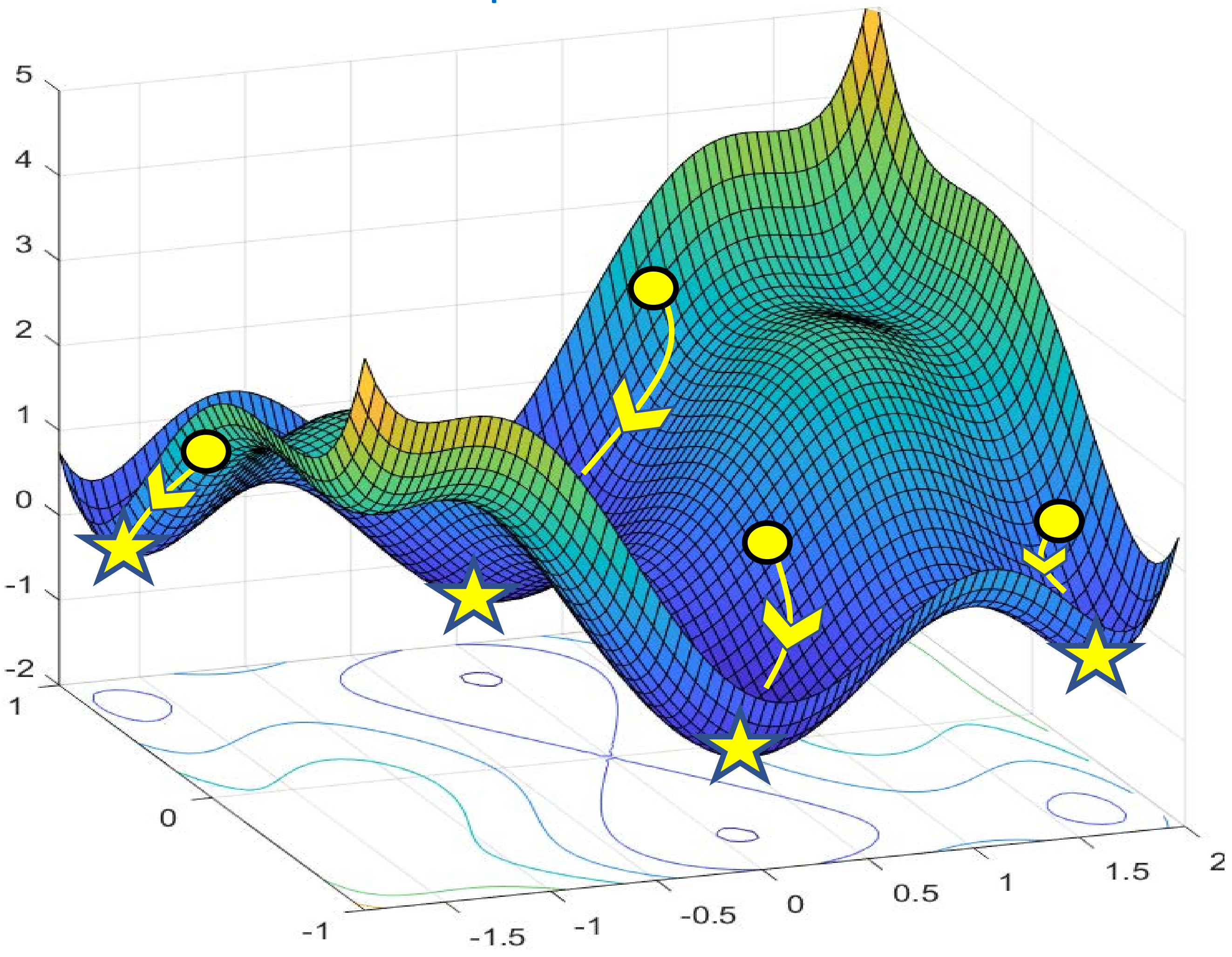
Optimality condition: $\nabla f(x) = 0, \quad \nabla^2 f(x) \succ 0$

Newton's Step: $x_{k+1} = x_k - \mathbf{H}_f^{-1} \nabla f(x_k)$

https://github.com/jasour/rarnop19/tree/master/Lecture2_NonlinearOptimization

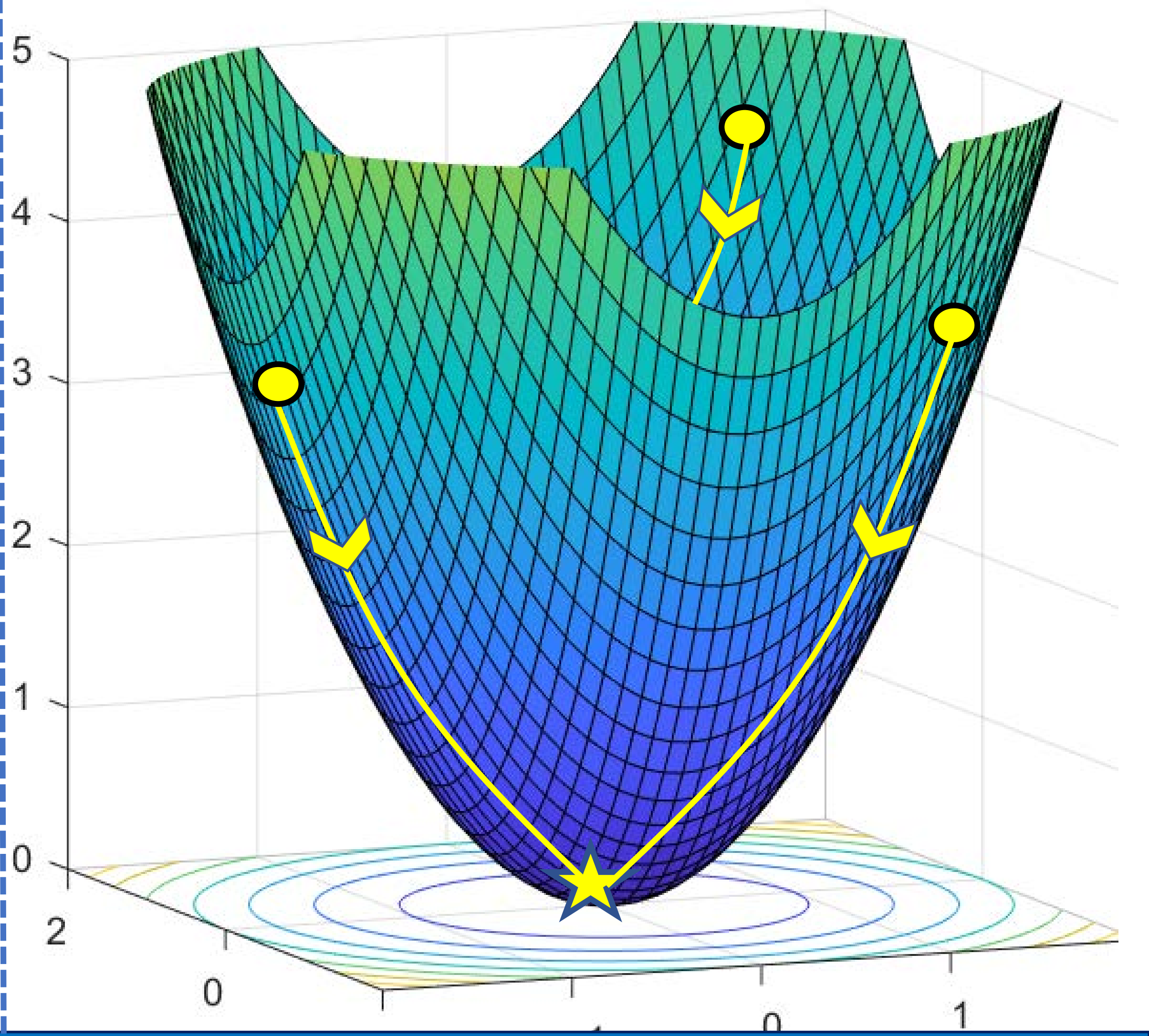
Nonconvex Optimization

- Multiple local minima
- Sensitive to initial point



Convex Optimization

- Unique minimum: global/local



Convex Optimization

Convex Optimization:

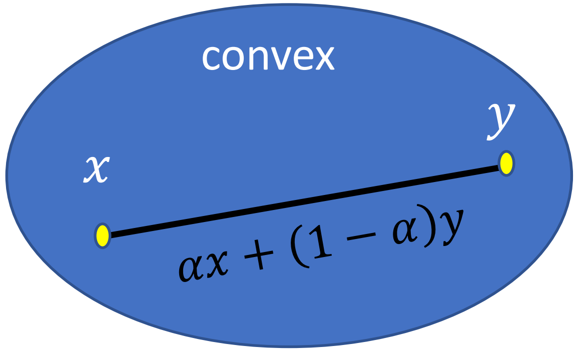
$$
\begin{aligned}
\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0, \;\; i = 1, ..., n_g \\
& h_i(x) = 0, \;\; i = 1, .., n_h
\end{aligned}
$$

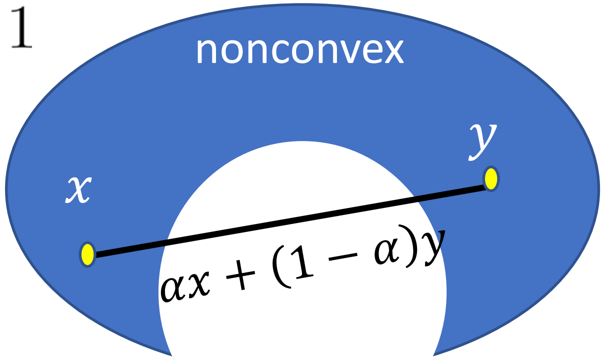
$$
\left.
\begin{aligned}
& f : \mathbb{R}^n \rightarrow \mathbb{R} \\
& g_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i = 1, ..., n_g \\
& h_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i = 1, ..., n_h
\end{aligned}
\right\} \text{Convex Functions}
$$

Convex Set:

Set $C \in \mathbb{R}^n$ is a convex set if a line segment joining any two elements lies entirely in the set.

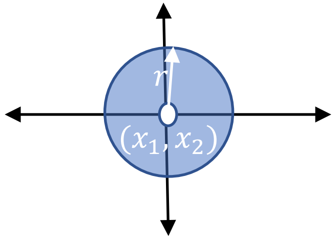
$$x, y \in C \rightarrow \alpha x + (1-\alpha)y \in C, \quad \forall 0 \leq \alpha \leq 1$$





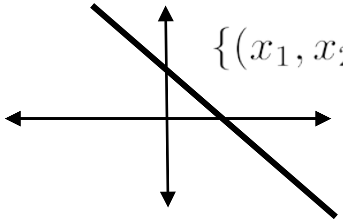
Examples:

Norm ball $\{x : ||x|| \leq r\}$ $||.||$:norm r:radius



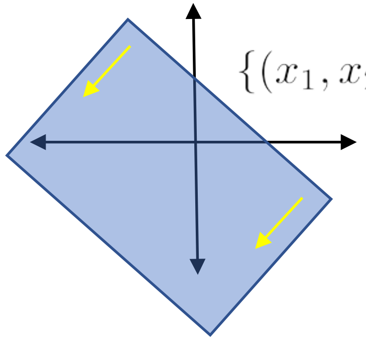
$\{(x_1, x_2) : \sqrt{x_1^2 + x_2^2} \leq r\}$

Hyperplane $\{x : a^T x = b\}$ a, b :vectors



$\{(x_1, x_2) : \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1\}$

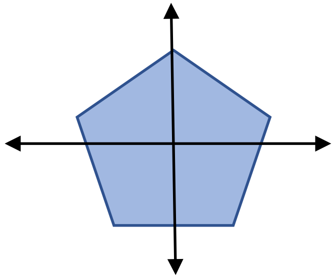
Halfspace $\{x : a^T x \leq b\}$ a, b :vectors



$\{(x_1, x_2) : \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq 1\}$

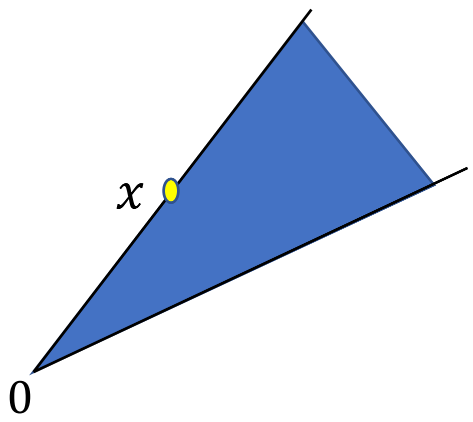
Affine Space $\{x : Ax = b\}$ A :matrix b:vectors

Polyhedron $\{x : Ax \leq b\}$



Cone:

A cone $C \in \mathbb{R}^n$ is a set such that $x \in C \rightarrow \alpha x \in C, \quad \forall \alpha \geq 0$



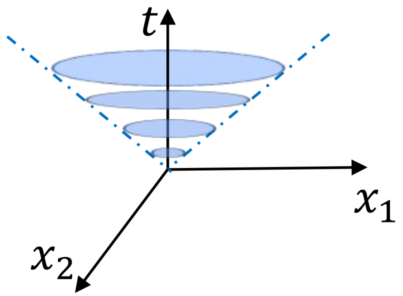
Convex Cone: $x, y \in C \rightarrow \alpha x + (1-\alpha)y \in C, \quad \forall 0 \leq \alpha \leq 1$

Examples:

Norm Cone $\{x : ||x|| \leq t\}$

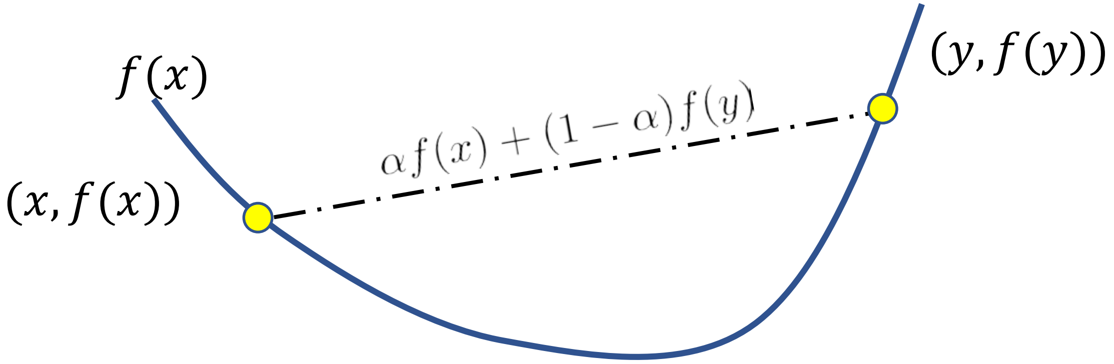
Under the l_2 norm, this is called a second-order cone.

$$\{(x_1, x_2) : \sqrt{x_1^2 + x_2^2} \leq t\}$$



Convex Function:

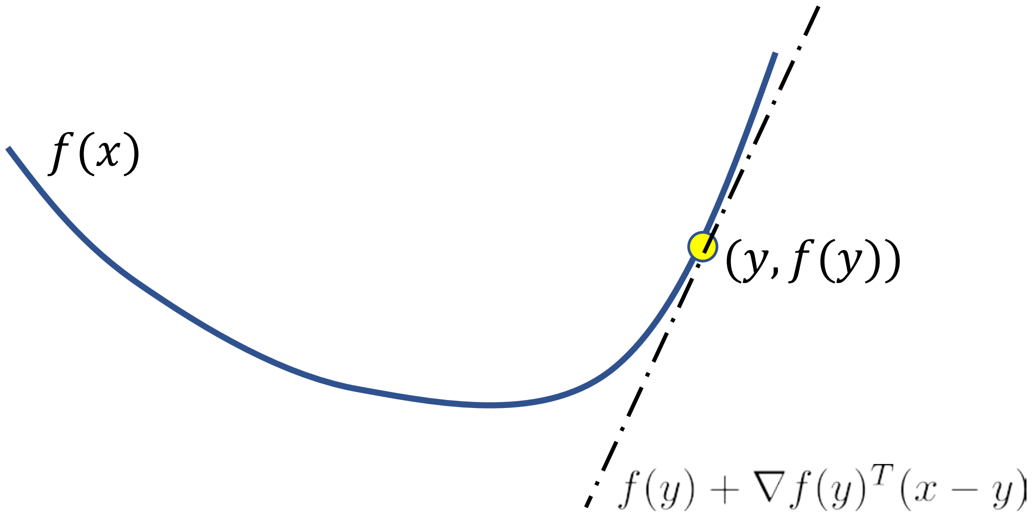
Function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if: $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \quad \forall 0 \leq \alpha \leq 1$



f lies below the line segment joining $f(x)$ and $f(y)$

- First order convexity condition:

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



- Second order convexity condition: $\nabla^2 f(x) \geq 0, \quad \forall x \in dom(f)$

- Domain of a convex function is a convex set.

Class of Convex Optimizations

Linear program (LP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Quadratic program (QP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2} x^T P_0 x + q_0^T x + c_0 \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

where $P_0 \succcurlyeq 0$

Quadratically Constrained Quadratic Program (QCQP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \frac{1}{2} x^T P_0 x + q_0^T x + c_0 \\ \text{subject to} \quad & \frac{1}{2} x^T P_i x + q_i^T x + c_i, \quad i = 1, \ldots, m \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

where $P_i \succcurlyeq 0, \ i = 0, \ldots, m$

Second-Order Cone Program (SOCP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & ||C_i x + d_i||_2 \leq e_i^T x + f_i, \ \ i = 1, \ldots, m \end{aligned}$$

Semidefinite Program (SDP)

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & C \bullet X \\ \text{subject to} \quad & A_i \bullet X = b_i \quad i = 1, \ldots, m. \\ & X \succcurlyeq 0. \end{aligned}$$

Cone Program (CP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \in K \end{aligned}$$

Primal and Dual Optimization

Primal and Dual Optimization

- Primal Optimization:

$$f^* = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\text{subject to} \quad g_i(x) \geq 0 \quad i = 1, \ldots, n_g$$

$$h_i(x) = 0 \quad i = 1, \ldots, n_h$$

Lagrange function:

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

KKT optimality condition:

$$\nabla_x L(x, \mu, \lambda) = 0$$

$$-h_i(x^*) = 0, \quad i = 1, \ldots, n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, \ldots, n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \ldots, n_g$$

Primal and Dual Optimization

➢ Primal Optimization:

$$\begin{aligned} f^* = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0 \quad i = 1, \ldots, n_g \\ & h_i(x) = 0 \quad i = 1, \ldots, n_h \end{aligned}$$

Lagrange function:

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

KKT optimality condition:

$$\nabla_x L(x, \mu, \lambda) = 0$$

$$-h_i(x^*) = 0, \; i = 1, \ldots, n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, \ldots, n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \ldots, n_g$$

Dual Optimization:

Instead of solving a nonconvex optimization, we want to solve a convex optimization and obtain a lower bound of the original optimization problem.

➢ Dual Optimization:

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}, \lambda \in \mathbb{R}^{n_h}}{\text{maximize}} \quad & d(\mu, \lambda) \\ \text{subject to} \quad & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{aligned}$$

$$\Rightarrow \quad d^* \leq f^*$$

Primal and Dual Optimization

- Primal Optimization:

$$
\begin{aligned}
f^* = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \geq 0 \quad i = 1, \ldots, n_g \\
& h_i(x) = 0 \quad i = 1, \ldots, n_h
\end{aligned}
$$

Lagrange function:

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

KKT optimality condition:

$$\nabla_x L(x, \mu, \lambda) = 0$$

$$-h_i(x^*) = 0, \; i = 1, \ldots, n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, \ldots, n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \ldots, n_g$$

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

- Lagrange function is a **lower bound** of the objective function

$$L(x, \lambda, \mu) = f(x) - \underbrace{\sum_{i=1}^{n_g} \mu_i g_i(x)}_{+} - \underbrace{\sum_{i=1}^{n_h} \lambda_i h_i(x)}_{0} \leq f(x)$$

$$\Rightarrow \boxed{L(x, \lambda, \mu) \leq f(x)}$$

Primal and Dual Optimization

➢ Primal Optimization:

$$
\begin{aligned}
f^* = \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \geq 0 \quad i = 1, \ldots, n_g \\
& h_i(x) = 0 \quad i = 1, \ldots, n_h
\end{aligned}
$$

Lagrange function:

$$L(x, \lambda, \mu) = f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)$$

KKT optimality condition:

$$\nabla_x L(x, \mu, \lambda) = 0$$

$$-h_i(x^*) = 0, \; i = 1, \ldots, n_h$$

$$g_i(x^*) \geq 0, \quad i = 1, \ldots, n_g$$

$$\mu_i \geq 0$$

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \ldots, n_g$$

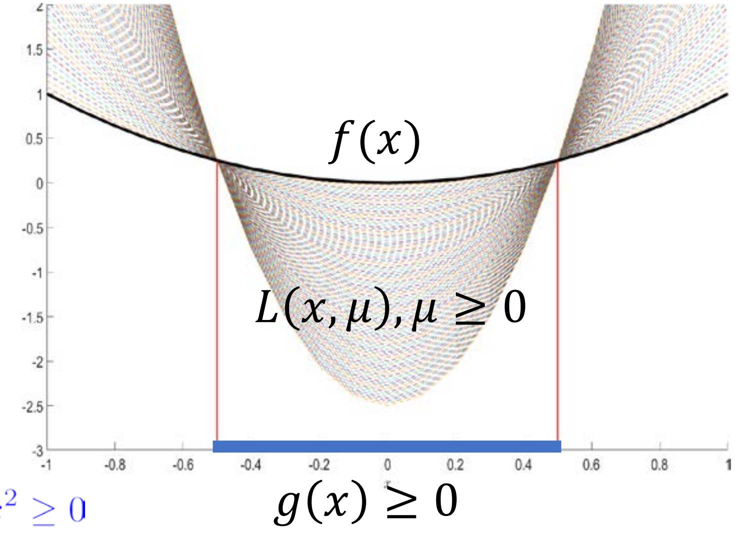
For feasible x (i.e., $h_i(x) = 0, \; g_i(x) \geq 0$) and $\mu \geq 0$

- Lagrange function is a **lower bound** of the objective function

$$L(x, \lambda, \mu) = f(x) - \underbrace{\sum_{i=1}^{n_g} \mu_i g_i(x)}_{+} - \underbrace{\sum_{i=1}^{n_h} \lambda_i h_i(x)}_{0} \leq f(x)$$

$$\Rightarrow \boxed{L(x, \lambda, \mu) \leq f(x)}$$

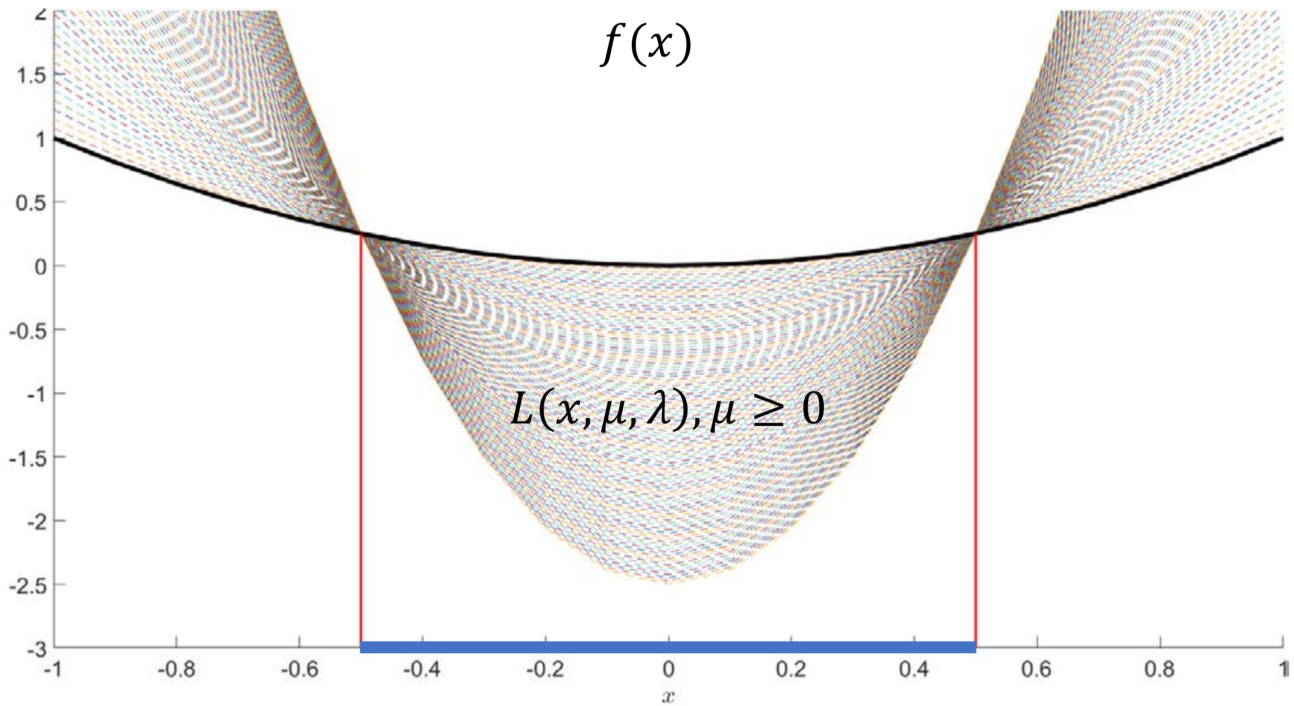
$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & x^2 \\
\text{subject to} \quad & 0.5^2 - x^2 \geq 0
\end{aligned}
$$



Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$



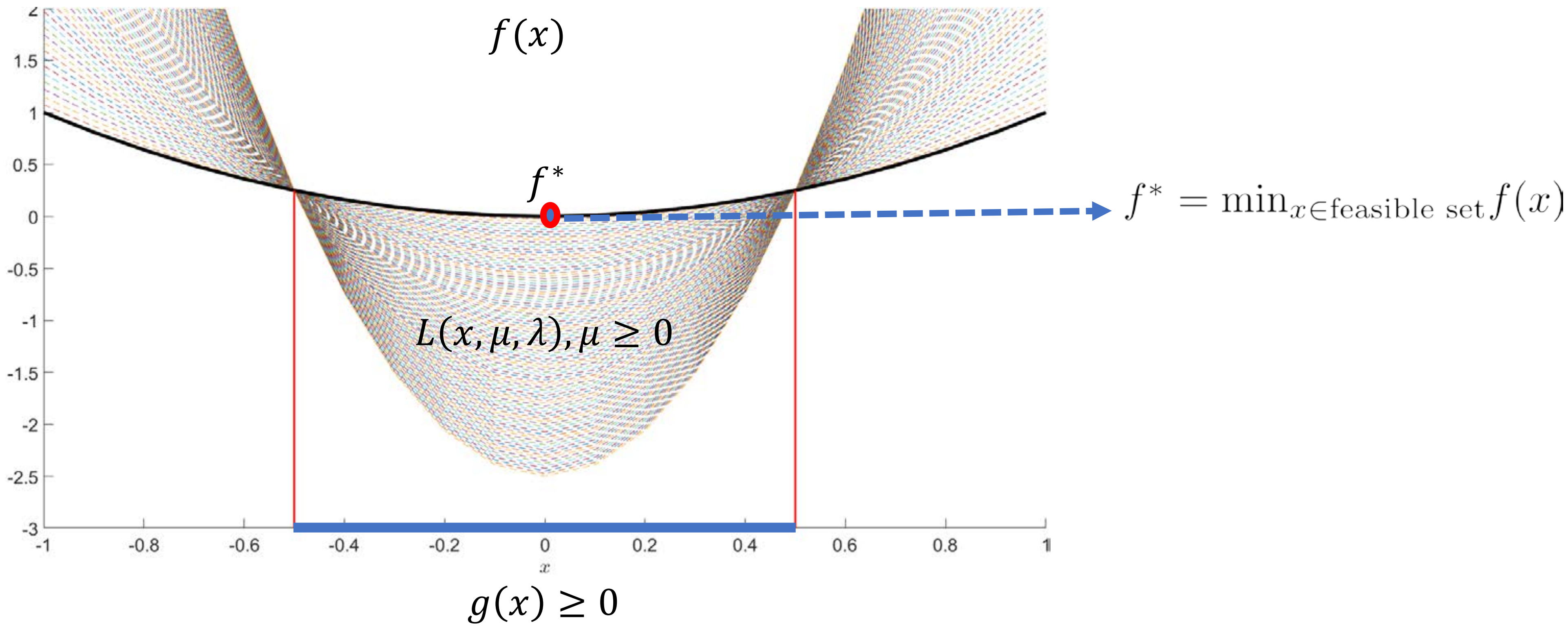
$g(x) \geq 0$

Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the best lower bound of f^* by looking at Lagrange function

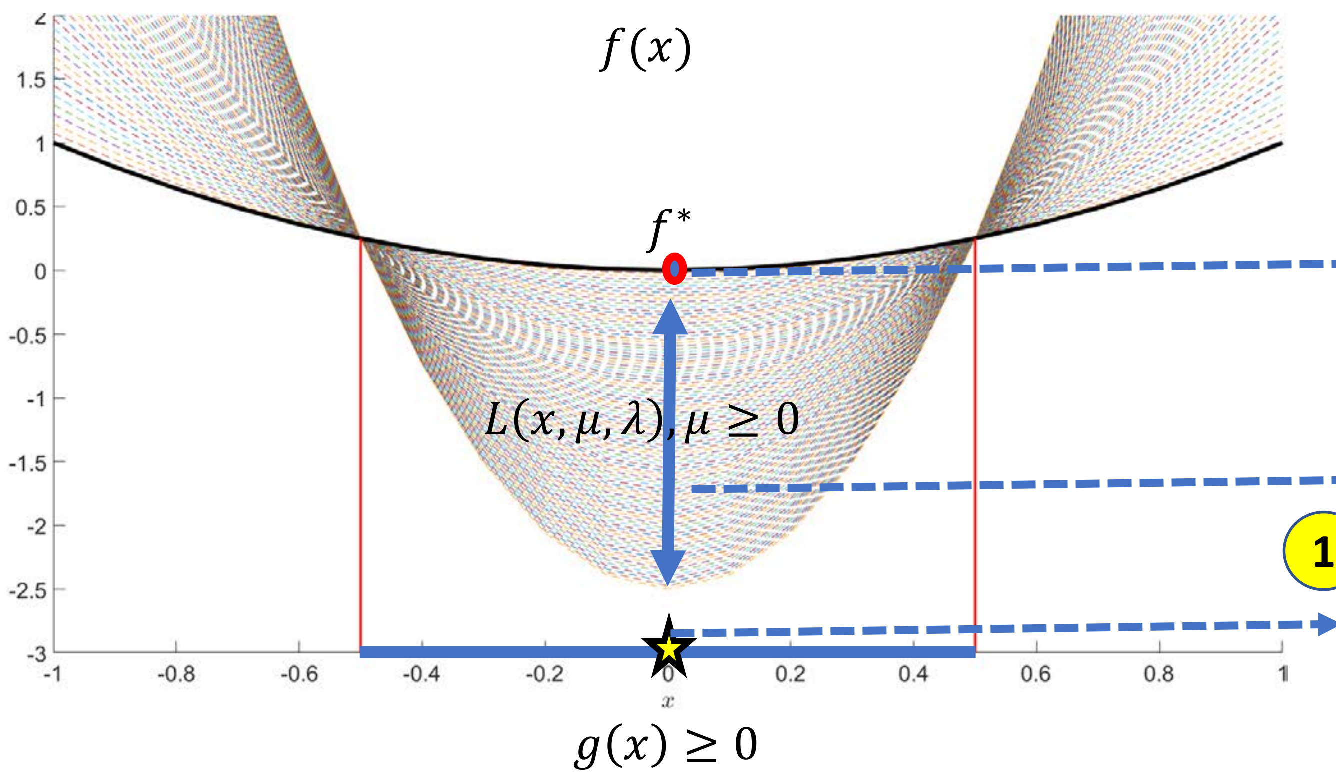


Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the **best lower bound of f^*** by looking at **Lagrange function**



$f^* = \min_{x \in \text{feasible set}} f(x)$

1

$d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$

$x^* = \text{argmin } d(\lambda, \mu)$

Primal and Dual Optimization

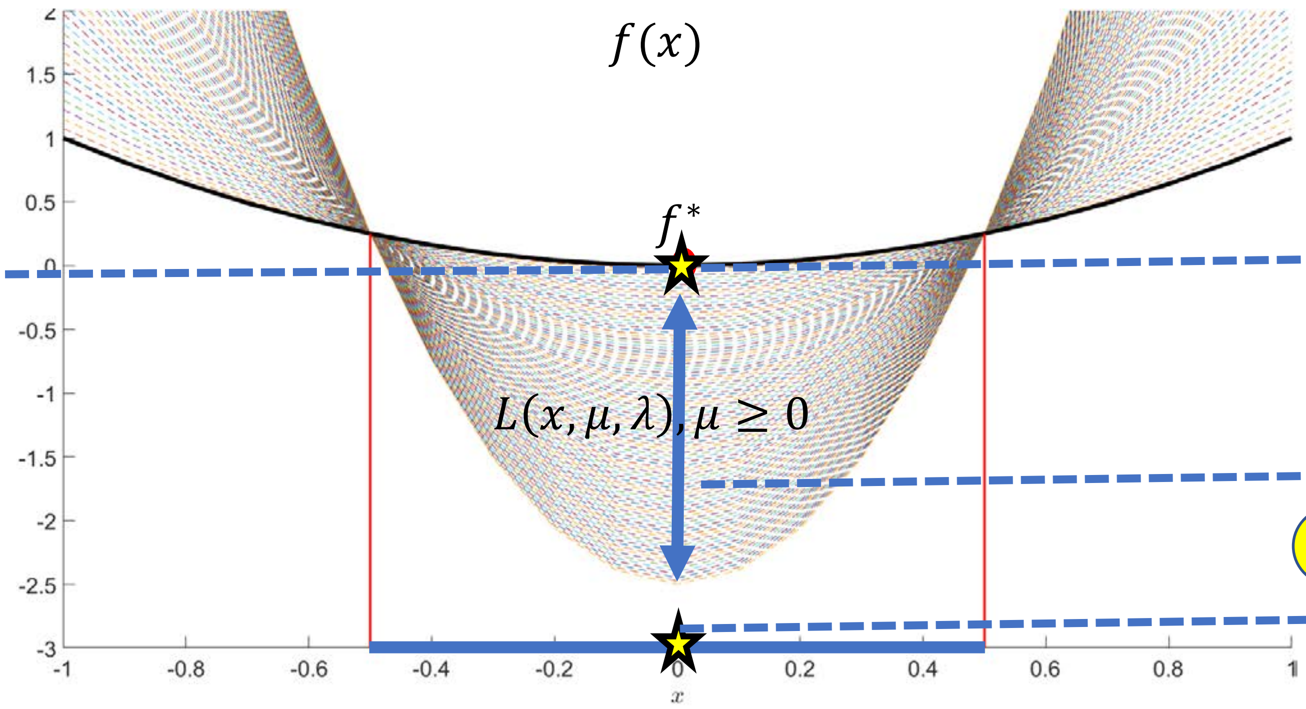
For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the best lower bound of f^* by looking at Lagrange function

2

$d^* = \max_{\mu \geq 0, \lambda} d(\mu, \lambda)$



$g(x) \geq 0$

$f^* = \min_{x \in \text{feasible set}} f(x)$

1

$d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$

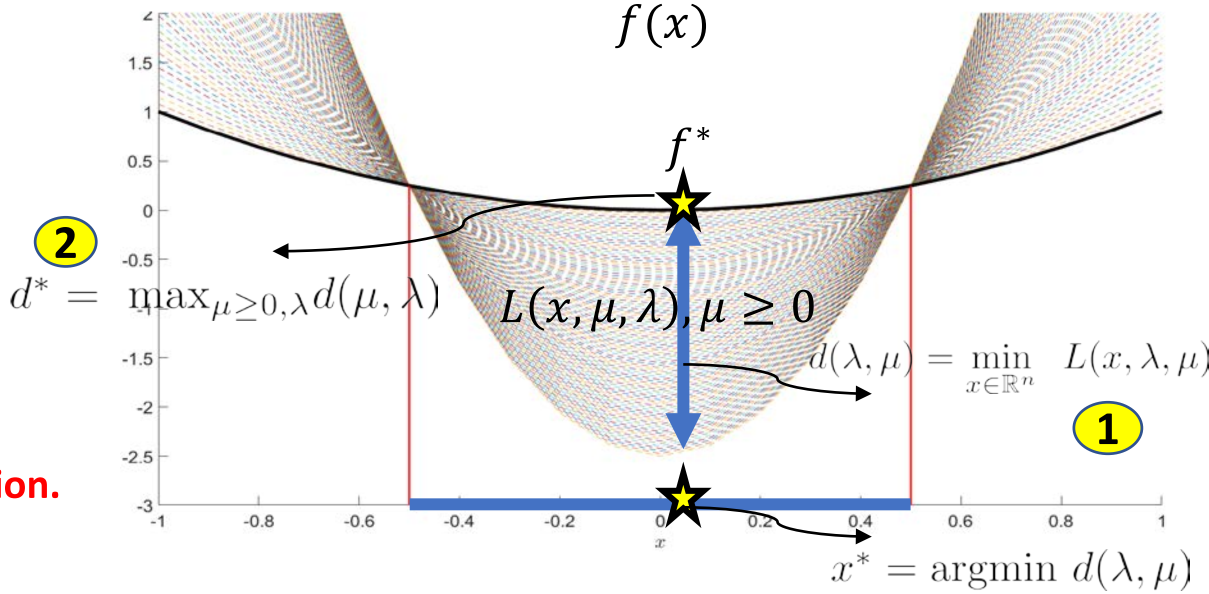
$x^* = \text{argmin } d(\lambda, \mu)$

Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the best lower bound of f^* by looking at Lagrange function.



Lagrange Dual function: $d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \; L(x, \lambda, \mu)$

Dual Optimization

$$\begin{aligned} d^* = \underset{\mu, \lambda}{\text{maximize}} \quad & d(\mu, \lambda) \\ \text{subject to} \quad & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{aligned}$$

Best Lower Bound

Weak Duality

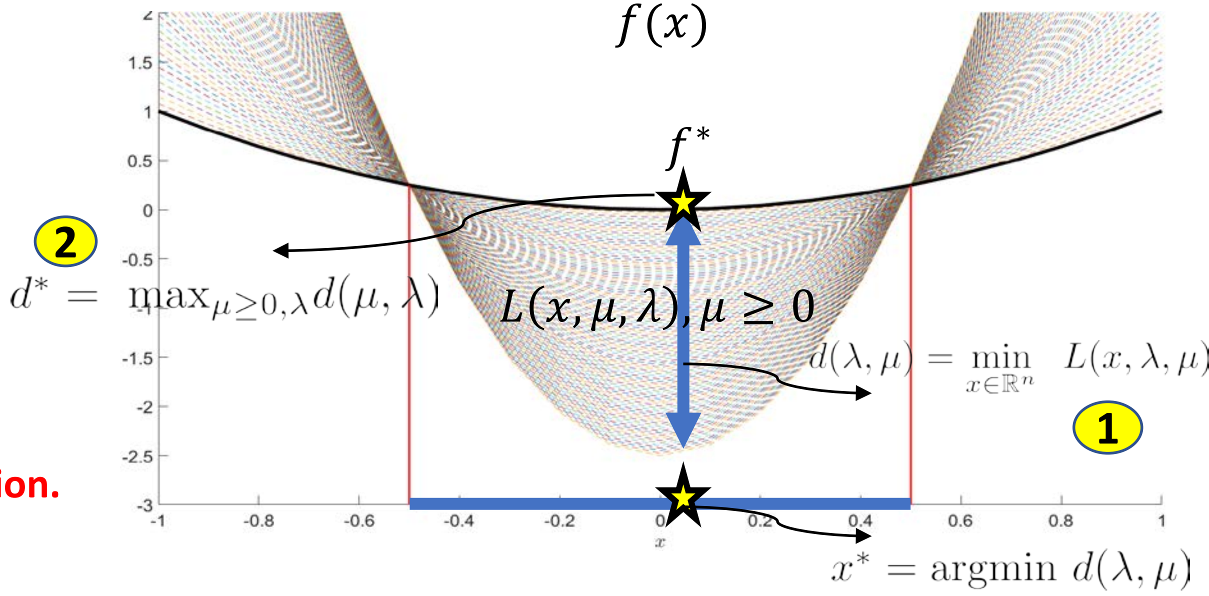
$$\boxed{d^* \leq f^*}$$

Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the best lower bound of f^* by looking at Lagrange function.



Lagrange Dual function: $d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$

Dual Optimization

$$\begin{aligned} d^* = \underset{\mu, \lambda}{\text{maximize}} \quad & d(\mu, \lambda) \\ \text{subject to} \quad & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{aligned}$$

Best Lower Bound

Weak Duality $\Rightarrow$ $\boxed{d^* \leq f^*}$

- **Dual Optimization is a convex problem.**
  - Linear constraints
  - Concave objective

$$d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \boxed{\min_{x \in \mathbb{R}^n} \left[f(x) - \sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)\right]}$$

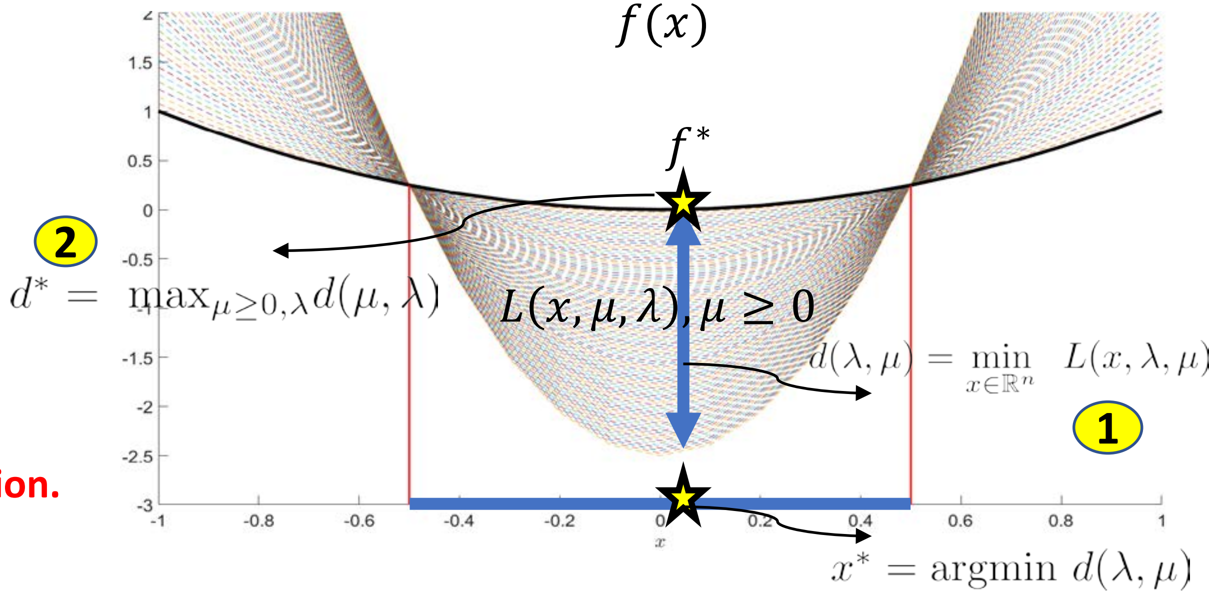
Concave function in μ, λ

Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$

We want to find the **best lower bound of f^*** by looking at **Lagrange function.**



Lagrange Dual function: $d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$

Dual Optimization

$$\begin{array}{ll} d^* = \underset{\mu, \lambda}{\text{maximize}} & d(\mu, \lambda) \\ \text{subject to} & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{array}$$

Best Lower Bound

Weak Duality $\Rightarrow$ $\boxed{d^* \leq f^*}$

- **Dual Optimization is a convex problem.**
  - Linear constraints
  - Concave objective

$$d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \min_{x \in \mathbb{R}^n} [f(x) - \textstyle\sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)]$$

Concave function in μ, λ

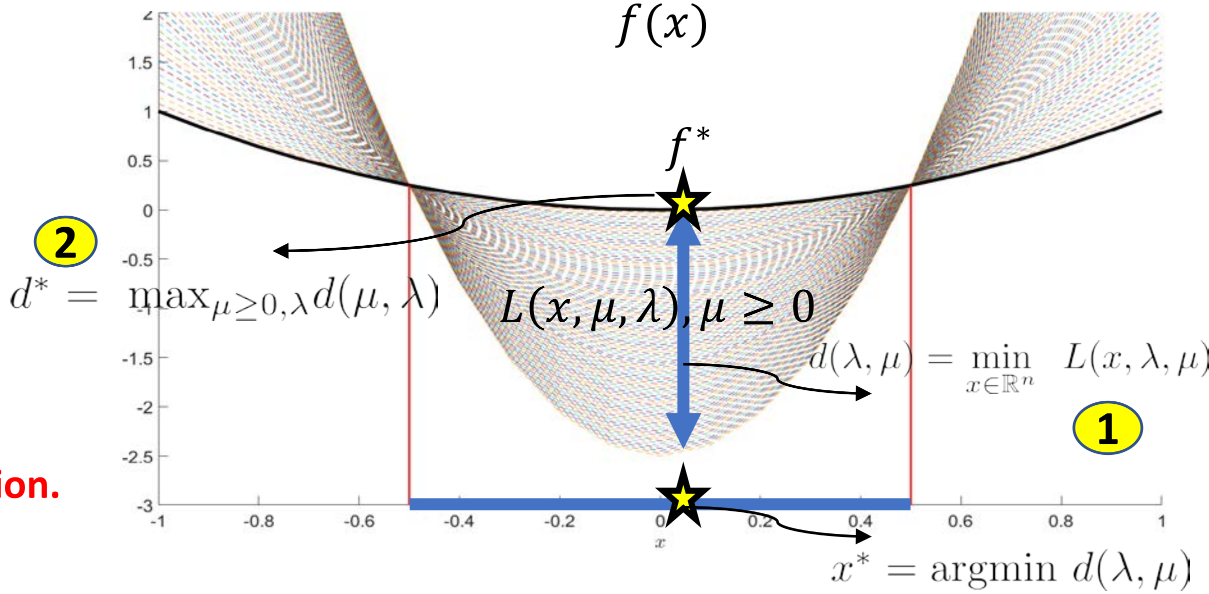
$$\min[f(x_1) - \textstyle\sum_{i=1}^{n_g} \mu_i g_i(x_1) - \sum_{i=1}^{n_h} \lambda_i h_i(x_1), f(x_2) - \sum_{i=1}^{n_g} \mu_i g_i(x_2) - \sum_{i=1}^{n_h} \lambda_i h_i(x_2), \ldots]$$

Minimum of concave functions is also concave.

Primal and Dual Optimization

For feasible x (i.e., $h_i(x) = 0, \quad g_i(x) \geq 0$) and $\mu \geq 0$

$$\boxed{L(x, \lambda, \mu) \leq f(x)}$$



We want to find the best lower bound of f^* by looking at Lagrange function.

Lagrange Dual function: $d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$

Dual Optimization

$$\begin{aligned} d^* = \underset{\mu, \lambda}{\text{maximize}} \quad & d(\mu, \lambda) \\ \text{subject to} \quad & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{aligned}$$

Best Lower Bound

Weak Duality $\Rightarrow$ $\boxed{d^* \leq f^*}$

- **Dual Optimization is a convex problem.**

- Linear constraints
- Concave objective

$$d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu) = \min_{x \in \mathbb{R}^n} [f(x) - \textstyle\sum_{i=1}^{n_g} \mu_i g_i(x) - \sum_{i=1}^{n_h} \lambda_i h_i(x)]$$

Concave function in μ, λ

Primal Problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \geq 0 \quad i = 1, \ldots, n_g \\ & h_i(x) = 0 \quad i = 1, \ldots, n_h \end{aligned}$$

Dual Problem

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}, \lambda \in \mathbb{R}^{n_h}}{\text{maximize}} \quad & d(\mu, \lambda) \\ \text{subject to} \quad & \mu_i \geq 0 \quad i = 1, \ldots, n_g \end{aligned}$$

Where, $d(\lambda, \mu) = \min_{x \in \mathbb{R}^n} \quad L(x, \lambda, \mu)$

Weak Duality

$$d^* \leq f^*$$

Strong Duality

Slater's Condition: If the primal problem is **convex**, and **strictly feasible** (feasible region must have an interior point), then

$$d^* = f^*$$

If x^* is a solution of primal optimization and $\mu^* \geq 0$ satisfies $d^*(\mu) = f^*(x^*)$

Then, x^* is a global solution of the original optimization problem.

Example

Primal Optimization:

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \quad \rightarrow \quad x_1^* = x_2^* = \mu^*/2$$

$$\mu^*(x_1^* + x_2^* - 1) = 0 \quad \rightarrow \quad x_1^* + x_2^* - 1 = 0$$

$$\mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0$$

$$x_1^* = x_2^* = 1/2, \quad \mu^* = 1$$

Example

Primal Optimization:

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\left\{\begin{array}{l} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \quad x_1^* + x_2^* - 1 \geq 0 \end{array}\right.$$

$$x_1^* = x_2^* = \mu^*/2$$

$$x_1^* + x_2^* - 1 = 0$$

$$x_1^* = x_2^* = 1/2, \ \mu^* = 1$$

Dual Optimization:

$$d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad d(\mu)$$
$$\text{subject to} \quad \mu \geq 0$$

Example

Primal Optimization:

$$
\begin{aligned}
&min_x \quad x_1^2 + x_2^2 \\
s.t. \quad &x_1 + x_2 - 1 \geq 0
\end{aligned}
$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$
\left\{
\begin{aligned}
&\begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\
&\mu^*(x_1^* + x_2^* - 1) = 0 \\
&\mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0
\end{aligned}
\right.
$$

$$x_1^* = x_2^* = \mu^*/2$$

$$x_1^* + x_2^* - 1 = 0$$

$$x_1^* = x_2^* = 1/2, \;\; \mu^* = 1$$

Dual Optimization:

$$
\begin{aligned}
d^* = &\underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad d(\mu) \\
&\text{subject to} \quad \mu \geq 0
\end{aligned}
$$

Lagrange Dual function:

$$
\begin{aligned}
d(\mu) &= \min_{x \in \mathbb{R}^n} \; L(x, \mu) \\
&= \min_{x \in \mathbb{R}^n} \; (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)
\end{aligned}
$$

Example

Primal Optimization:

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\left\{\begin{array}{l} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \quad x_1^* + x_2^* - 1 \geq 0 \end{array}\right.$$

$$x_1^* = x_2^* = \mu^*/2$$

$$x_1^* + x_2^* - 1 = 0$$

$$\Downarrow$$

$$x_1^* = x_2^* = 1/2, \;\; \mu^* = 1$$

Dual Optimization:

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad & d(\mu) \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

Lagrange Dual function:

$$\begin{aligned} d(\mu) &= \min_{x \in \mathbb{R}^n} \; L(x,\mu) \\ &= \min_{x \in \mathbb{R}^n} \; (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1) \end{aligned}$$

Optimality Condition of Unconstraint Opt:

$$\nabla_x((x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)) = 0$$

$$\Rightarrow \quad \begin{aligned} 2x_1 - \mu &= 0 \\ 2x_2 - \mu &= 0 \end{aligned} \quad \Rightarrow \quad x_1 = x_2 = \mu/2$$

Example

Primal Optimization:

$$\begin{aligned} & min_x \quad x_1^2 + x_2^2 \\ s.t. \quad & x_1 + x_2 - 1 \geq 0 \end{aligned}$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\left\{ \begin{array}{l} \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ \mu^*(x_1^* + x_2^* - 1) = 0 \\ \mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0 \end{array} \right.$$

$$x_1^* = x_2^* = \mu^*/2$$

$$x_1^* + x_2^* - 1 = 0$$

$$\Downarrow$$

$$x_1^* = x_2^* = 1/2, \ \mu^* = 1$$

Dual Optimization:

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad & d(\mu) \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

Lagrange Dual function:

$$\begin{aligned} d(\mu) &= \min_{x \in \mathbb{R}^n} \ L(x, \mu) \\ &= \min_{x \in \mathbb{R}^n} \ (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1) \end{aligned}$$

Optimality Condition of Unconstraint Opt:

$$\nabla_x((x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)) = 0$$

$$\begin{aligned} 2x_1 - \mu &= 0 \\ 2x_2 - \mu &= 0 \end{aligned} \quad \Rightarrow \quad x_1 = x_2 = \mu/2$$

Lagrange Dual function: $d(\mu) = \mu - \mu^2/2$

Example

Primal Optimization:

$$\begin{aligned} &min_x \quad x_1^2 + x_2^2 \\ s.t. \quad &x_1 + x_2 - 1 \geq 0 \end{aligned}$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \quad \rightarrow \quad x_1^* = x_2^* = \mu^*/2$$

$$\mu^*(x_1^* + x_2^* - 1) = 0 \quad \rightarrow \quad x_1^* + x_2^* - 1 = 0$$

$$\mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0$$

$$\Downarrow$$

$$x_1^* = x_2^* = 1/2, \; \mu^* = 1$$

Dual Optimization:

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad & d(\mu) \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

Lagrange Dual function:

$$\begin{aligned} d(\mu) &= \min_{x \in \mathbb{R}^n} \; L(x, \mu) \\ &= \min_{x \in \mathbb{R}^n} \; (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1) \end{aligned}$$

Optimality Condition of Unconstraint Opt:

$$\nabla_x ((x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)) = 0$$

$$\Rightarrow \quad \begin{aligned} 2x_1 - \mu &= 0 \\ 2x_2 - \mu &= 0 \end{aligned} \quad \Rightarrow \quad x_1 = x_2 = \mu/2$$

Lagrange Dual function: $d(\mu) = \mu - \mu^2/2$

Dual Optimization:

$$\begin{aligned} d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad & \mu - \mu^2/2 \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

Example

Primal Optimization:

$$\min_x \quad x_1^2 + x_2^2$$
$$s.t. \quad x_1 + x_2 - 1 \geq 0$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$$

$$\begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \rightarrow x_1^* = x_2^* = \mu^*/2$$

$$\mu^*(x_1^* + x_2^* - 1) = 0 \rightarrow x_1^* + x_2^* - 1 = 0$$

$$\mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0$$

$$x_1^* = x_2^* = 1/2, \; \mu^* = 1$$

Dual Optimization:

$$d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad d(\mu)$$
$$\text{subject to} \quad \mu \geq 0$$

Lagrange Dual function:

$$d(\mu) = \min_{x \in \mathbb{R}^n} \; L(x,\mu)$$
$$= \min_{x \in \mathbb{R}^n} \; (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)$$

Optimality Condition of Unconstraint Opt:

$$\nabla_x((x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)) = 0$$

$$2x_1 - \mu = 0$$
$$2x_2 - \mu = 0$$
$$\Rightarrow x_1 = x_2 = \mu/2$$

Lagrange Dual function: $d(\mu) = \mu - \mu^2/2$

Dual Optimization:

$$d^* = \underset{\mu \in \mathbb{R}^{n_g}}{\text{maximize}} \quad \mu - \mu^2/2$$
$$\text{subject to} \quad \mu \geq 0$$

Optimality Condition

$$\nabla_\mu(\mu - \mu^2/2)) = 0$$
$$\mu \geq 0$$
$$\Rightarrow \mu^* = 1 \qquad x_1^* = x_2^* = 1/2$$

Example

Primal Optimization:

$$\begin{aligned} min_x \quad & x_1^2 + x_2^2 \\ s.t. \quad & x_1 + x_2 - 1 \geq 0 \end{aligned}$$

Lagrange function:

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

KKT: $\nabla_x L(x^*, \mu^*) = 0$

$g(x^*) \geq 0 \qquad \mu^* \geq 0 \qquad \mu^* g(x^*) = 0$

$$\left\{\begin{aligned} & \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} - \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \\ & \mu^*(x_1^* + x_2^* - 1) = 0 \\ & \mu^* \geq 0 \qquad x_1^* + x_2^* - 1 \geq 0 \end{aligned}\right.$$

$$x_1^* = x_2^* = \mu^*/2$$

$$x_1^* + x_2^* - 1 = 0$$

$$x_1^* = x_2^* = 1/2, \quad \mu^* = 1$$

$$f^* = 1/2$$

Dual Optimization:

Lagrange Dual function:

$$\begin{aligned} d(\mu) &= \min_{x \in \mathbb{R}^n} \quad L(x, \mu) \\ &= \min_{x \in \mathbb{R}^n} \quad (x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1) \end{aligned}$$

Optimality Condition of Unconstraint Opt:

$$\nabla_x((x_1^2 + x_2^2) - \mu(x_1 + x_2 - 1)) = 0$$

$$\begin{aligned} 2x_1 - \mu &= 0 \\ 2x_2 - \mu &= 0 \end{aligned} \quad \Rightarrow \quad x_1 = x_2 = \mu/2$$

Lagrange Dual function: $d(\mu) = \mu - \mu^2/2$

Without solving the Dual optimization, since

$$d(\mu^* = 1) = 1/2 = f^*$$

Hence, $x_1^* = x_2^* = 1/2$ is global optimal solution.

Standard Linear Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange Function

Lagrange multipliers

$$\begin{aligned} L(x, y, s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

KKT Condition:

$$\nabla_x L(x, y, s) = c - A^T y - s = 0$$

$$Ax = b$$

$$s^T x = 0$$

$$s \geq 0 \quad x \geq 0$$

Standard Linear Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function:

$$\begin{aligned} L(x, y, s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Standard Linear Program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function:

$$\begin{aligned} L(x,y,s) &= c^T x - y^T(Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s)$$

$$\nabla_x L(x,y,s) = c - A^T y - s = 0$$

Unconstrained Optimization

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Standard Linear Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

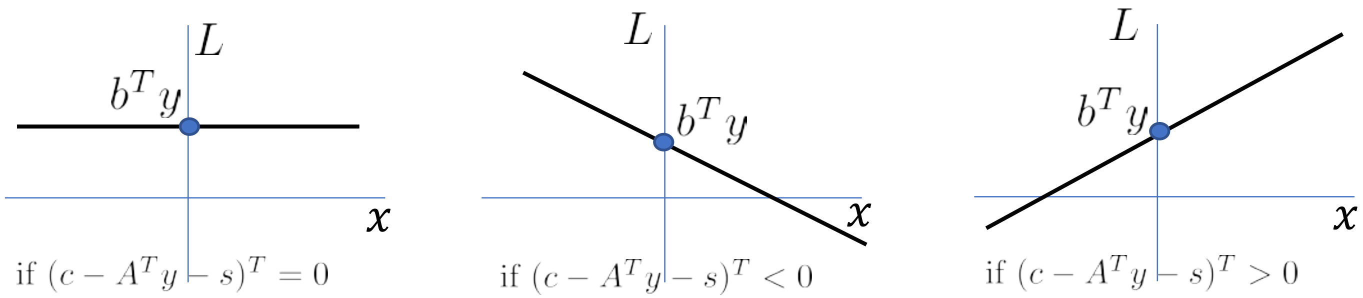
Lagrange multipliers

Lagrange function:
$$\begin{aligned} L(x, y, s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) \xrightarrow[\text{Unconstrained Optimization}]{\nabla_x L(x, y, s) = c - A^T y - s = 0} d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- We want to minimize Lagrange function with respect to x.
- $b^T y + (c - A^T y - s)^T x$: Lagrange function is a line in terms of x.



$$d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) = b^T y \qquad d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) = -\infty$$

- We are looking for the best minimum point (best lower bound of the objective function of the primal).

Standard Linear Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function:

$$\begin{aligned} L(x, y, s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}} \; L(x, y, s)$$

$$\xrightarrow[\textit{Unconstrained Optimization}]{\nabla_x L(x, y, s) = c - A^T y - s = 0}$$

$$d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}} \; L(x, y, s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual LP:

$$\underset{y, s \geq 0}{\text{maximize}} \; d(y, s)$$

$$\boxed{\begin{aligned} \underset{y, s}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y = s \\ & s \geq 0. \end{aligned} \quad \longrightarrow \quad \begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y \geq 0 \end{aligned}}$$

Standard Linear Program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function:

$$\begin{aligned} L(x,y,s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s)$$

$$\nabla_x L(x,y,s) = c - A^T y - s = 0$$

Unconstrained Optimization

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual LP:

$$\underset{y,s\geq 0}{\text{maximize}}\ d(y,s)$$

$$\boxed{\begin{aligned} \underset{y,s}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y = s \\ & s \geq 0. \end{aligned} \quad \longrightarrow \quad \begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y \geq 0 \end{aligned}}$$

Weak Duality: $b^T y \leq c^T x \longrightarrow c^T x - b^T y \geq 0 \quad c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = s^T x \geq 0$

Standard Linear Program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function:
$$\begin{aligned} L(x, y, s) &= c^T x - y^T (Ax - b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) \xrightarrow[\textit{Unconstrained Optimization}]{\nabla_x L(x,y,s) = c - A^T y - s = 0} d(y, s) = \underset{x \in \mathbb{R}^n}{\text{minimize}}\ L(x, y, s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual LP: $\underset{y, s \geq 0}{\text{maximize}}\ d(y, s)$

$$\begin{aligned} \underset{y, s}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y = s \\ & s \geq 0. \end{aligned} \longrightarrow \begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y \geq 0 \end{aligned}$$

Weak Duality: $b^T y \leq c^T x \longrightarrow c^T x - b^T y \geq 0 \quad c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = s^T x \geq 0$

Strong Duality:

$$c^T x - b^T y = 0 \longrightarrow s^T x = 0$$

Strong duality holds if the primal problem is feasible.

Standard Linear Program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Lagrange multipliers

Lagrange function: $$\begin{aligned} L(x,y,s) &= c^T x - y^T(Ax-b) - s^T x \\ &= b^T y + (c - A^T y - s)^T x \end{aligned}$$

Dual function

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s)$$

$\nabla_x L(x,y,s) = c - A^T y - s = 0$

Unconstrained Optimization

$$d(y,s) = \underset{x\in\mathbb{R}^n}{\text{minimize}}\ L(x,y,s) = \begin{cases} b^T y & c - A^T y - s = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual LP: $\underset{y,s\geq 0}{\text{maximize}}\ d(y,s)$

$$\begin{aligned} \underset{y,s}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y = s \\ & s \geq 0. \end{aligned} \quad \longrightarrow \quad \begin{aligned} \underset{y}{\text{maximize}} \quad & b^T y \\ \text{subject to} \quad & c - A^T y \geq 0 \end{aligned}$$

Weak Duality: $b^T y \leq c^T x \longrightarrow c^T x - b^T y \geq 0 \quad c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = s^T x \geq 0$

Strong Duality:

$$c^T x - b^T y = 0 \longrightarrow s^T x = 0 \xrightarrow{\textit{Optimality Cond. Based on Strong Duality}} \begin{cases} Ax = b \quad x \geq 0 & \textit{Feasibility condition of Primal} \\ c - A^T y = s \quad s \geq 0 & \textit{Feasibility condition of Dual} \\ s^T x = 0 & \textit{Strong duality} \end{cases}$$

KKT Condition

Strong duality holds if the primal problem is feasible.

Overview of Nonlinear Optimization:

- Optimality Conditions
- Newton’s Method
- Interior Point Method
- Convex Optimization
- Dual Optimization

Books:

"Nonlinear Programming" Dimitri Bertsekas, MIT.

"Nonlinear Programming Theory and Algorithms" Mokhtar Bazaraa, Hanif Sherali, Shetty.

"Convex Optimization", Stephen Boyd, Lieven Vandenberghe, Stanford University.

Lecture Notes:

"Optimization methods", Pablo A. Parrilo, MIT.

"Nonlinear Programming", Pablo A. Parrilo, MIT.

"Introduction to Convex Optimization", Pablo A. Parrilo, MIT.

"Convex and Conic Optimization", Amir Ali Ahmadi, Princeton.

"Convex Optimization", Laurent El Ghaoui, UC Berkeley.

"Convex Optimization", Ryan Tibshirani, Carnegie Mellon University.