

Optimization for Data Science



By

Dr. Nishant Kumar

Assistant Professor

Department of Electrical Engineering

Indian Institute of Technology (IIT) Jodhpur

Introduction

Optimization : The procedure or procedures used to make a system or design as effective or functional as possible.

Why Optimization?

- Helps improve the quality of decision-making
- Applications in Engineering, Business, Economics, Science, Military Planning etc.

Mathematical Program

Mathematical Program : A mathematical formulation of an optimization problem:

Minimize $f(x)$ subject to $x \in S$

Essential Components of a Mathematical program:

x : variables or parameters

f : objective function

S : feasible region

What is a solution of this Mathematical Program?

$x^* \in S$ such that $f(x^*) \leq f(x) \forall x \in S$

x^* : solution, $f(x^*)$: optimal objective function value

x^* may not be unique and may not even exist.

Maximize $f(x) \equiv$ Minimize $-f(x)$

Mathematical Optimization

The problem,

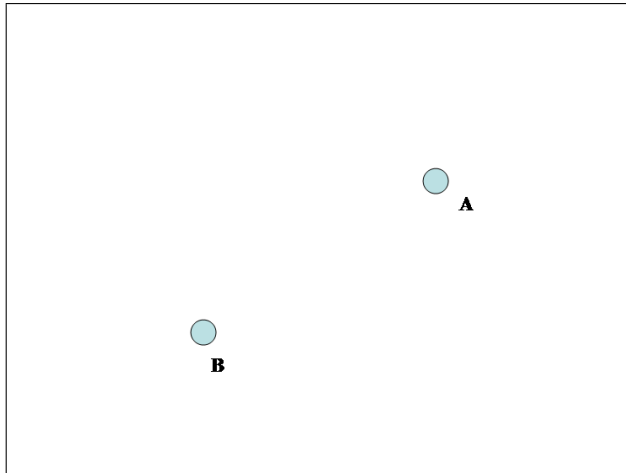
Minimize $f(x)$ subject to $x \in S$

can be written as

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & x \in S \end{array} \quad (1)$$

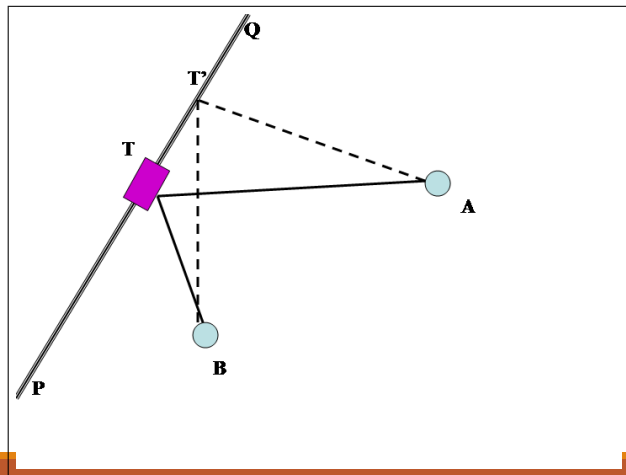
Some Optimization Problems

- Find the *shortest* path between the two points A and B in a horizontal plane



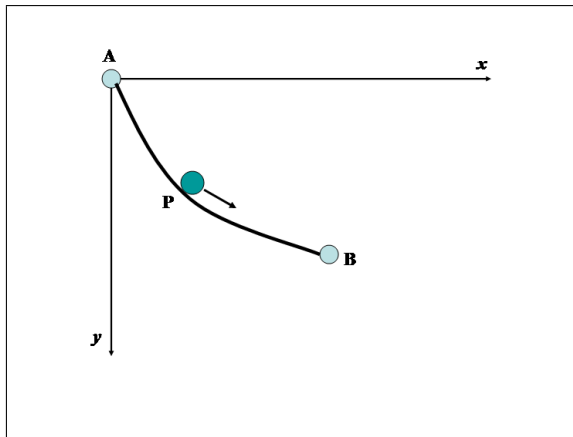
Some Optimization Problems

Bus Terminus Location Problem: Find the location of the bus terminus T on the road segment PQ such that the lengths of the roads linking T with the two cities A and B is *minimum*.



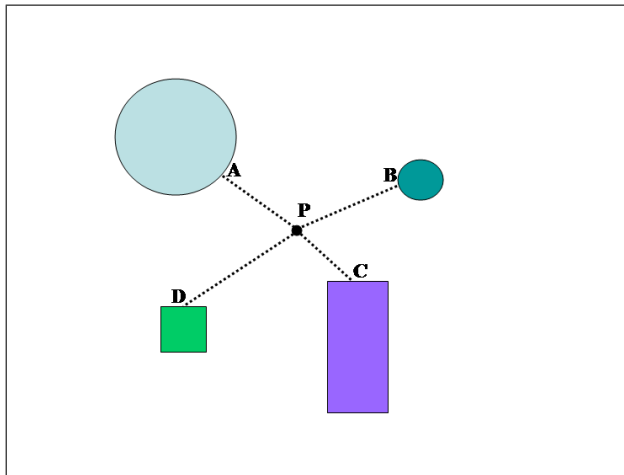
Some Optimization Problems

Given two points A and B in a vertical plane, find a path APB which an object must follow, so that starting from A, it reaches B in the *shortest* time under its own gravity.



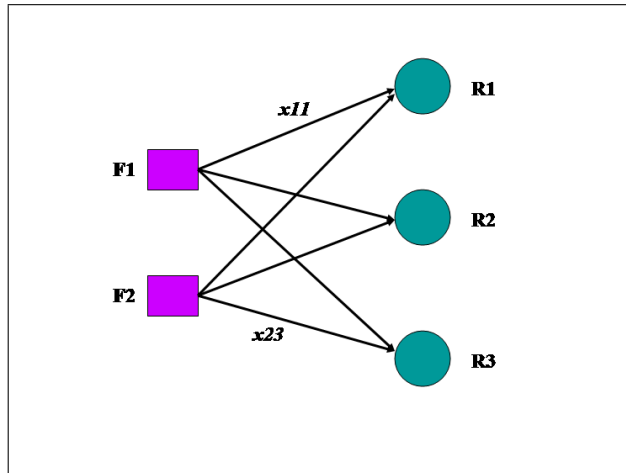
Some Optimization Problems

Facility location problem: Find a location (**within the boundary**) that *minimizes* the sum of distances to each of the locations



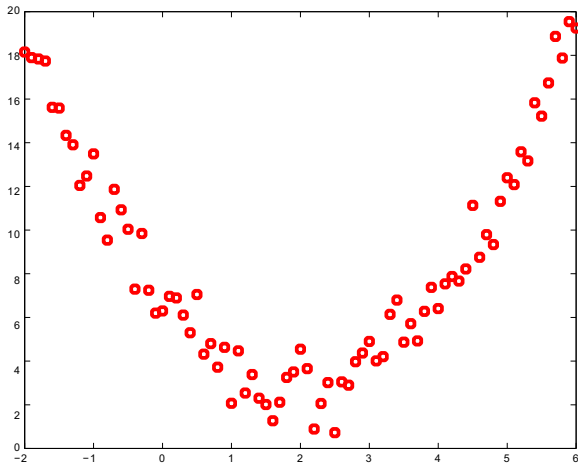
- **Transportation Problem:**

Find the “best” way to satisfy the requirement of demand points using the capacities of supply points.



Data Fitting Problem:

From a family of potential models, find a model that “best” fits the observed data.



Application Domains

- Various disciplines in Engineering
- Science
- Economics and Statistics
- Business

Some Example Problems

- Scheduling problem
- Diet Problem
- Portfolio Allocation Problem
- Engineering Design
- Manufacturing
- Robot Path Planning
- ...

Mathematical Optimization Process

- Typical steps for Solving Mathematical Optimization Problems
 - Problem formulation
 - Checking the existence of a solution
 - Solving the optimization problem, if a solution exists
 - Solution analysis
 - Algorithm analysis

Mathematical Optimization Process

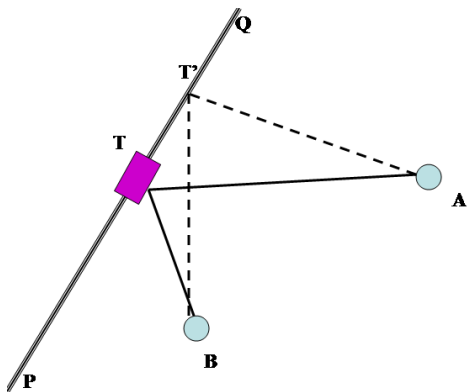
Typical steps for Solving Mathematical Optimization Problems

- Problem formulation

$$\begin{array}{ll}\min & f(x) \\ & x \\ \text{s.t.} & x \in S\end{array}$$

- Checking the existence of a solution
- Solving the optimization problem, if a solution exists
- Solution analysis
- Algorithm analysis

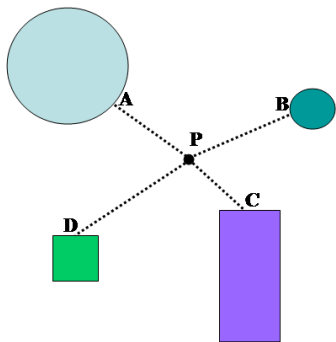
Formulation: Bus Terminus Location Problem



- Coordinates of A and B:
 $x_A = (x_{A1}, x_{A2})$ and
 $x_B = (x_{B1}, x_{B2})$
- Equation of line PQ:
 $ax_1 + bx_2 + c = 0$
- Use Euclidean distance
- $x_T = (x_{T1}, x_{T2})$ (**variables**)
- The **objective** is to *minimize*
 $d(x_A, x_T) + d(x_B, x_T)$
- T lies on PQ (**constraint**)

$$\begin{aligned} \min_{x_{T1}, x_{T2}} \quad & d(x_A, x_T) + d(x_B, x_T) \\ \text{s.t.} \quad & ax_{T1} + bx_{T2} + c = 0 \end{aligned}$$

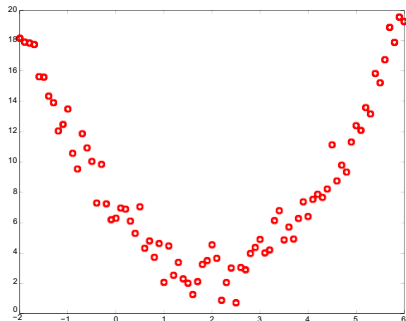
Formulation: Facility Location Problem



- x_A, x_B, x_C and x_D belong to the respective location boundaries
- Use Euclidean distance
- (x_{P1}, x_{P2}) (**variables**)
- The **objective** is to *minimize*
 $d(x_A, x_P) + d(x_B, x_P) + d(x_C, x_P) + d(x_D, x_P)$
- $x_A \in A, x_B \in B, x_C \in C$ and $x_D \in D$ (**constraints**)

$$\begin{aligned} \min_{x_{P1}, x_{P2}} \quad & d(x_A, x_P) + d(x_B, x_P) + d(x_C, x_P) + d(x_D, x_P) \\ \text{s.t.} \quad & x_A \in A, x_B \in B, x_C \in C, x_D \in D \end{aligned}$$

Formulation: Data Fitting Problem



- Given : $\{x_i, y_i\}_{i=1}^n$, n data points
- Given : Most probable model type, $f(x) = ax^2 + bx + c$
- a, b, c : **variables**
- Measure of misfit: $(y - f(x))^2$
- The **objective** is to *minimize* $\sum_i (y_i - (ax_i^2 + bx_i + c))^2$
- **No constraints**

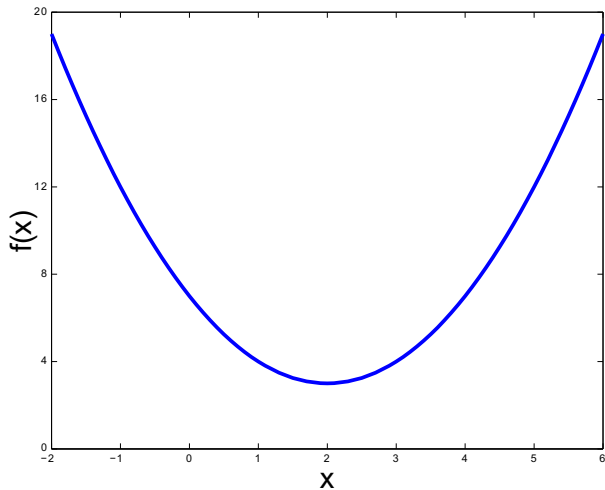
Mathematical Optimization Process

Typical steps for Solving Mathematical Optimization Problems

- Problem formulation
- Checking the existence of a solution
- Solving the optimization problem, if a solution exists
 - Graphical method
 - Analytical method
 - Numerical method
- Solution analysis
- Algorithm analysis

Functions of One Variable

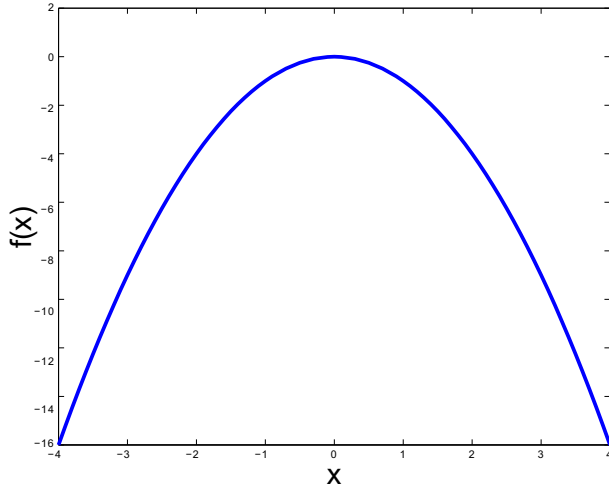
- $f(x) = (x - 2)^2 + 3$
- Minimum at $x^* = 2$, minimum function value: $f(x^*) = 3$



Functions of One Variable

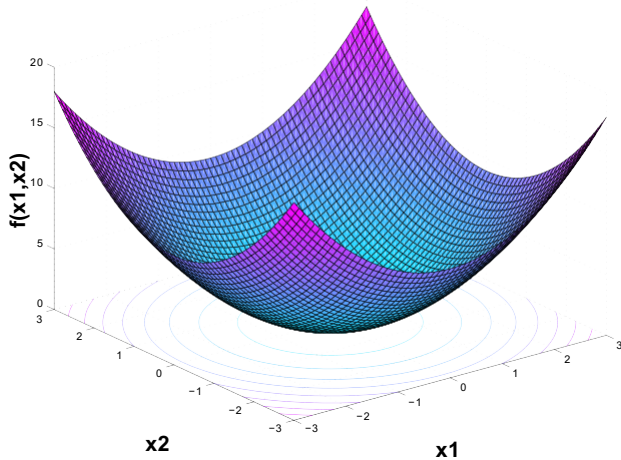
$$f(x) = -x^2$$

- Maximum at $x^* = 0$, maximum function value: $f(x^*) = 0$



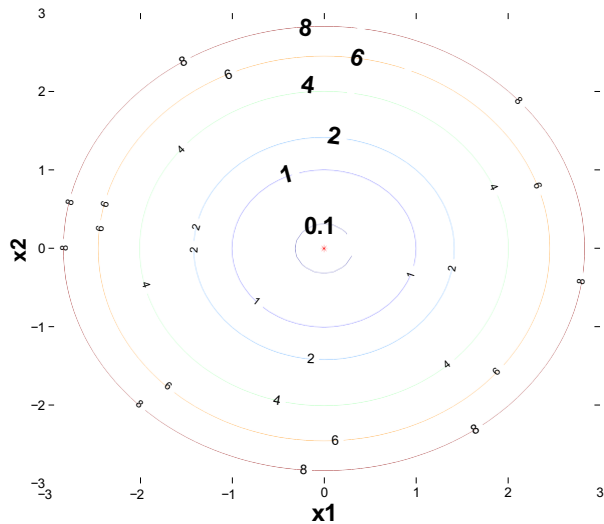
Functions of Two Variables: Surface Plots

- $f(x_1, x_2) = x_1^2 + x_2^2$
- Minimum at $x_1^* = 0, x_2^* = 0; f(x^*, x^*) = 0$



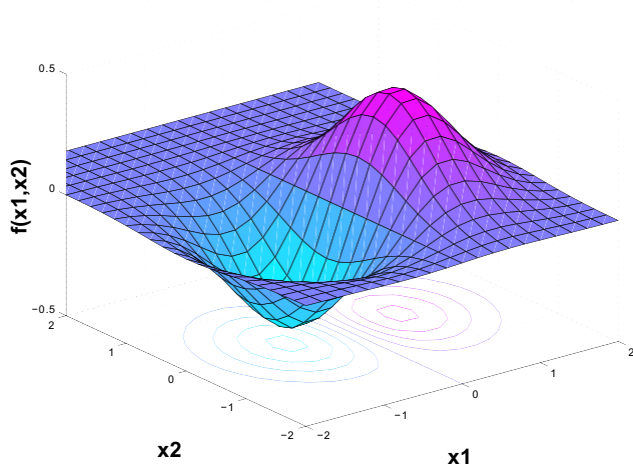
Functions of Two Variables: Contour Plots

• $f(x_1, x_2) = x_1^2 + x_2^2$



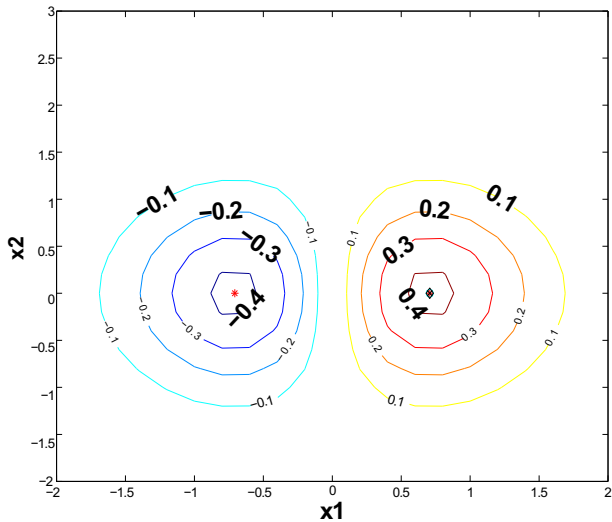
Functions of Two Variables: Surface Plots

- $f(x_1, x_2) = x_1 \exp(-x_1^2 - x_2^2)$
- Minimum at $(-1/\sqrt{2}, 0)$, maximum at $(1/\sqrt{2}, 0)$



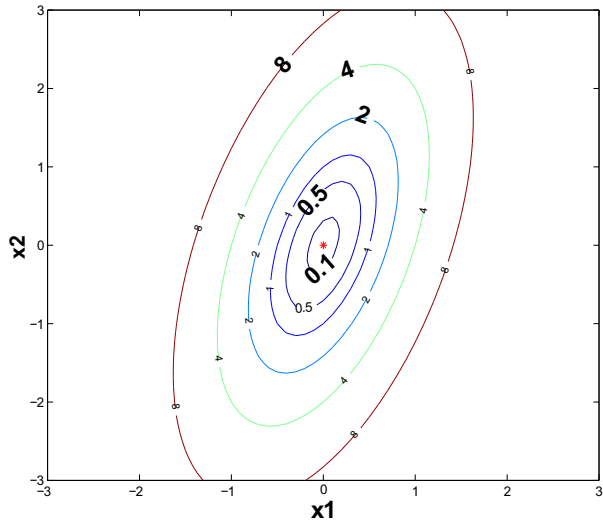
Functions of Two Variables: Contour Plots

• $f(x_1, x_2) = x_1 \exp(-x_1^2 - x_2^2)$

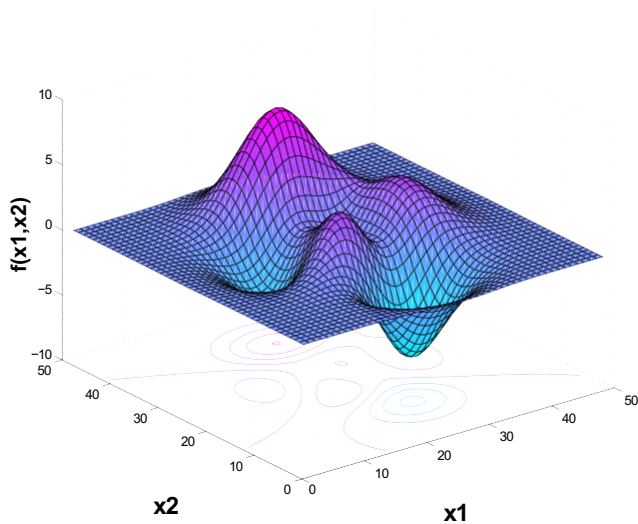


Functions of Two Variables: Contour Plots

• $f(x_1, x_2) = f(x, y) = 4x_1^2 + x_2^2 - 2x_1$



Functions of Two Variables: Surface Plots



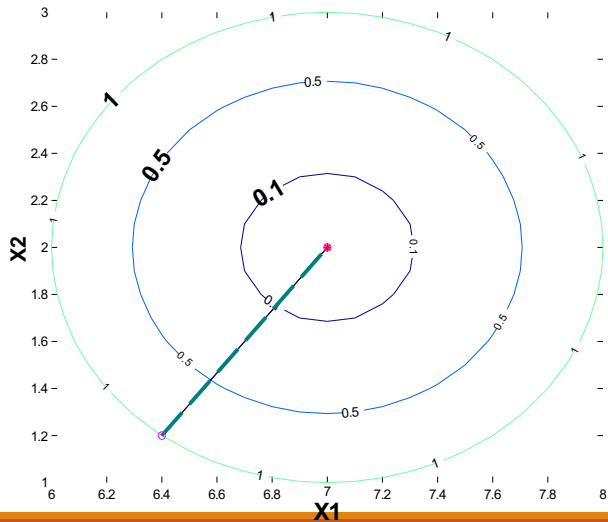
Mathematical Optimization Process

Typical steps for Solving Mathematical Optimization Problems

- Problem formulation
- Checking the existence of a solution
- Solving the optimization problem, if a solution exists
 - Graphical method
 - Analytical method
 - Numerical method
- Solution analysis
- Algorithm analysis

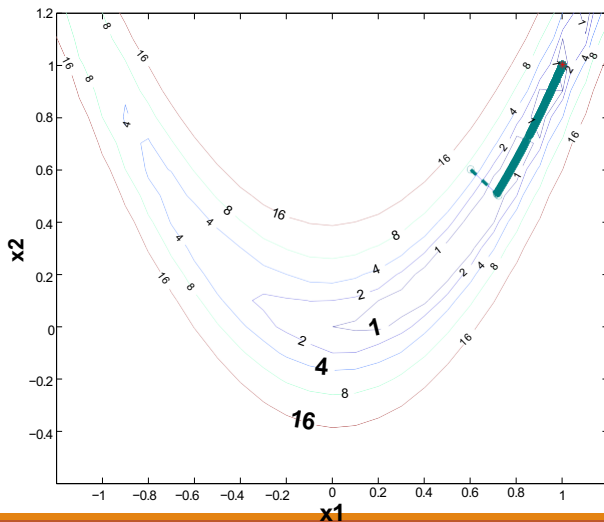
Iterates of an Optimization Algorithm

- $f(x_1, x_2) = (x_1 - 7)^2 + (x_2 - 2)^2$
- Initial Point: (6.4, 1.2), Minimum at (7, 2)



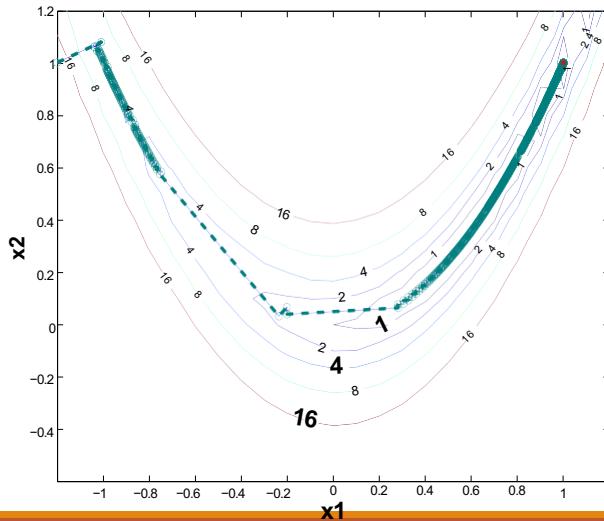
Iterates of an Optimization Algorithm

- $f(x_1, x_2) = 100(x_2 - x_1)^2 + (1 - x_1)^2$
- Initial Point: (0.6, 0.6), Minimum at (1, 1)



Iterates of an Optimization Algorithm

- $f(x_1, x_2) = 100(x_2 - x_1)^2 + (4 - x_1)^2$
- Initial Point: $(-2, 1)$, Minimum at $(1, 1)$



Types of Optimization Problems

- Constrained and unconstrained optimization
- Continuous and discrete optimization
- Stochastic and deterministic optimization

Types of Optimization Problems

- Constrained optimization problem:

$$\begin{array}{ll}\min_x & f(x) \\ \text{s.t.} & x \in S\end{array}$$

- Unconstrained optimization problem:

$$\min_x f(x)$$

Types of Optimization Problems

- Continuous optimization
 - Variables are typically real-valued

$$f(x) = x^2 + 4x + 5$$

$$x \in \mathbb{R}$$

- Discrete optimization
 - Variables are not real-valued: they take binary or integer values

$$\text{Maximize: } P(n) = 10n - 2n^2$$

$$n \in \{0, 1, 2, 3, 4, 5\}$$

Types of Optimization Problems

- Stochastic optimization
 - Some or all of the problem data are random
 - In some cases, the constraints hold with some probabilities
 - Need to define feasibility and optimality appropriately
- Deterministic optimization
 - No randomness in problem data and constraints