# Mathematics for Computer Science Engineers

## Unit - 4

## Optimization in Statistics

**Dr. UMA D.**
Professor
Department of Computer Science and Engineering

Acknowledgement: Authors of the prescribed textbooks and materials sourced online.

# Optimization in Engineering

## Optimization in Engineering

Optimization is a fundamental aspect of engineering that involves finding the best solution from a set of feasible alternatives.

In engineering, optimization techniques are used to enhance the performance, efficiency, and reliability of systems and processes while minimizing costs, resources, and time.

The goal is to achieve the best possible outcome while adhering to constraints and requirements specific to the engineering domain.

# Optimization in Engineering

## Optimization in Engineering

In various fields of engineering such as **computer science**, **mechanical, electrical, civil, and chemical,** optimization problems arise frequently, ranging from design and production processes to resource allocation and operational management.

In **computer science**, optimization techniques are vital for **improving algorithms, enhancing resource allocation in networks, and maximizing** the **performance of software applications**.

By employing optimization methods, engineers and computer scientists can make informed decisions that **lead to better designs, improved efficiency, and enhanced performance.**

# What is Optimization?

## Popular Optimization Problem in Computer Science

### Network Routing

One of the most significant optimization problems is **network routing**, which is crucial for managing data transfer efficiently across computer networks.

Given a network of routers and switches, the goal is to determine the most efficient path for data packets to travel from a source node to a destination node while minimizing latency, maximizing bandwidth utilization, and ensuring network reliability.

**Company**: **Cisco Systems**, a leader in networking technology.

# What is Optimization?

## Network Routing

**Network routing**, which is crucial for managing data transfer efficiently across computer networks.

Given a network of routers and switches, the goal is to determine the most efficient path for data packets to travel from a source node to a destination node while minimizing latency, maximizing bandwidth utilization, and ensuring network reliability.

**Objective:** Minimize the total latency of data packets sent between multiple offices while also considering constraints like bandwidth limitations, network congestion, and maintenance costs of network devices.

PES UNIVERSITY

# What is Optimization?

**Network Routing can be formulated as a Constrained Optimization Problem**

$$\text{Minimize } L = \sum_{(i,j)\in E} c_{ij} \cdot x_{ij}$$

$c_{ij}$ is the latency on each link

$x_{ij}$ is a binary variable that is 1 if the link (i,j) is part of the path P from s to t, and 0 otherwise.

s.t.

$$x_{ij} \leq C_{ij} \cdot y_{ij} \quad \forall (i,j) \in E \qquad y_{ij} \in \{0,1\} \quad \forall (i,j) \in E$$

**Flow Conservation Constraints**: Ensure data only flows from the source s to the destination t, maintaining the path continuity.

For each node i

$$\sum_j x_{ji} - \sum_j x_{ij} = \begin{cases} 1 & \text{if } i = \text{source} \\ -1 & \text{if } i = \text{destination} \\ 0 & \text{otherwise} \end{cases}$$

PES UNIVERSITY

# What is Optimization?

The expression shown in the image is:

$$y_{ij} \in \{0,1\} \quad \forall(i,j) \in E$$

This typically represents a binary variable $y_{ij}$ that can take the value of either 0 or 1 for each edge $(i,j)$ in the set of edges $E$. In the context of a network routing or optimization problem:

- $y_{ij} = 1$ would indicate that the edge $(i,j)$ is selected as part of the solution, such as being included in a path.
- $y_{ij} = 0$ would indicate that the edge $(i,j)$ is not selected.

This binary variable is commonly used in network optimization to specify whether a particular link or edge is included in the optimal route or path.

# What is Optimization?

**Network Routing can be formulated as a Linear Programming Optimization Problem**

$$\text{Minimize } Z = 10x_{AB} + 20x_{AC} + 5x_{BC} + 15x_{BD} + 30x_{CD}$$

Subject to:

$$x_{AB} \leq 100$$
$$x_{AC} \leq 50$$
$$x_{BC} \leq 80$$
$$x_{BD} \leq 60$$
$$x_{CD} \leq 90$$

Capacity constraint

$$x_{AB} + x_{AC} = 1$$
$$x_{AB} - x_{BC} - x_{BD} = 0$$
$$x_{AC} + x_{BC} - x_{CD} = 0$$
$$x_{BD} + x_{CD} = 1$$

Flow Conservation Constraints

$$x_{AB}, x_{AC}, x_{BC}, x_{BD}, x_{CD} \geq 0$$

Non-negativity constraints

# What is Optimization?

## Popular Optimization Problem in Electronics

### Optimal Component Placement in Printed Circuit Board (PCB) Design

Designing a printed circuit board (PCB) is a crucial step in the development of electronic devices.

The placement of components on a PCB significantly affects performance, manufacturability, and overall cost.

Optimizing the placement of components while adhering to various constraints can be framed as an optimization problem.

# What is Optimization?

## Optimal Component Placement in Printed Circuit Board (PCB) Design

Determine the optimal placement of components on a PCB to minimize the overall length of the connections (wires) between them, while ensuring that the layout meets certain constraints, such as avoiding overlaps, respecting component sizes, and adhering to manufacturing limitations.

The goal is to **minimize the total wire length** connecting all components on the PCB. The objective function can be expressed as:

$$\text{Minimize } Z = \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} d_{ij}(x_i, y_i, x_j, y_j)$$

Where $d_{ij}$ is the Euclidean distance between components $i$ and $j$, calculated as:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

# What is Optimization?

## Optimal Component Placement in Printed Circuit Board (PCB) Design

**Component Placement Constraints**:

- Each component must be placed within the bounds of the PCB. If $L$ and $W$ are the length and width of the PCB, then:

$$0 \leq x_i \leq L - s_i \quad \text{for } i = 1, 2, \ldots, n$$

$$0 \leq y_i \leq W - s_i \quad \text{for } i = 1, 2, \ldots, n$$

**Overlap Constraints**:

- No two components should overlap. For components $i$ and $j$, this can be expressed as:

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq s_i + s_j \quad \text{for } i \neq j$$

This constraint ensures that the distance between the centers of the components is at least the sum of their sizes.

# What is Optimization?

## Optimal Component Placement in Printed Circuit Board (PCB) Design

Manufacturing Constraints:

$$|x_i - x_j| \geq d_{\min} \quad \text{for connected components}$$

Connection Constraints:

- Specific constraints based on required connections between components.

# What is Optimization?

## Popular Optimization Problem in Electrical Engineering

### Power Distribution Optimization

Optimizing the power distribution network is a significant problem.

The goal is to minimize the total cost of power distribution while satisfying the demand of various consumers connected to the grid.

The network consists of substations, transmission lines, and consumers with specific power requirements.

# What is Optimization?

## Popular Optimization Problem in Electrical Engineering

### Power Distribution Optimization

**Objective**: A utility company wants to determine the optimal distribution of power to different consumers while minimizing the costs associated with generation, transmission losses, and infrastructure.

# What is Optimization?

## Power Distribution Optimization

$$\text{Minimize } Z = C_{gen}(P_{gen}) + C_{trans}(P_{trans})$$

**Subject to Constraints:**

1. Power Balance Constraint:

$$P_{gen} = \sum_{i=1}^{n} P_i + P_{loss}$$

2. Demand Constraints:

$$P_i \geq D_i \quad \text{for all } i$$

3. Transmission Limits:

$$P_{trans,j} \leq T_j \quad \text{for all } j$$

4. Generation Capacity:

$$P_{gen} \leq G_{max}$$

# What is Optimization?

Note: Power Distribution Optimization can be modelled as a

LINEAR PROGRAMMING (LPP) PROBLEM

INTEGER LINEAR PROGRAMMING (ILP) PROBLEM

DYNAMIC PROGRAMMING (DP) PROBLEM

# What is Optimization?

## Popular Optimization Problem in Biotechnology

### Drug Development Optimization

Developing a new drug involves numerous variables, including the amount of resources allocated to different stages of the drug development process (e.g., research, clinical trials, and production).

A pharmaceutical company aims to optimize its resource allocation to maximize the effectiveness of the drug development process while minimizing costs.

# What is Optimization?

## Drug Development Optimization as a LPP

**Objective Function:** Maximize effectiveness $E$:

$$E = 0.5x_1 + 0.3x_2 + 0.2x_3$$

**Subject to Constraints:**

1. Total budget constraint:

$$x_1 + x_2 + x_3 \leq 1,000,000$$

2. Minimum allocation to clinical trials:

$$x_2 \geq 0.4 \times 1,000,000 = 400,000$$

3. Budget allocation for research and production:

$$x_1 + x_3 \leq 600,000$$

4. Non-negativity constraints:

$$x_1, x_2, x_3 \geq 0$$

# What is Optimization?

## What is Optimization?

•Optimization is the process of finding the best possible solution to a problem within given constraints.

•It involves maximizing or minimizing an objective function, such as cost, time, efficiency, or resource utilization.

# Why Optimization?

## Why Optimization?

**Efficient Resource Utilization:** Optimization helps in making the best use of available resources like time, money, or materials.

**Improved Decision-Making:** It enables selecting the most effective solution from multiple alternatives, ensuring better outcomes.

**Maximizing Performance:** By optimizing processes or systems, we can maximize performance, such as increasing speed, accuracy, or output.

**Cost Reduction:** Optimization minimizes waste, reducing costs in industrial, technological, and business processes.

**Handling Complex Problems:** It provides structured approaches to solve complex real-world problems with multiple constraints and objectives.

**Scalability:** Optimization techniques allow systems or solutions to scale efficiently as they grow in complexity or size.

# Types of Optimization

- **Constrained Optimization**
  - **Linear Optimization**
    - **Linear Programming Problems (LPP)**
    - **Integer Programming**
    - Dynamic Programming
  - **Non-Linear Optimization**
    - Non-Linear Programming
    - Non-Linear Integer Programming
    - Non-Linear Dynamic Programming
- **Unconstrained Optimization**
  - Unconstrained Linear Optimization
  - Unconstrained Non-Linear Optimization
  - Global Optimization
- **Discrete Optimization**
  - Combinatorial Optimization
    - Knapsack Problem
    - Traveling Salesman Problem
  - Integer Programming
- **Continuous Optimization**
  - Linear Optimization
  - Non-Linear Optimization
  - Calculus-Based Optimization

# Applications in Computer Science

| Category | Example | Company | Purpose of Use |
|---|---|---|---|
| **Constrained Optimization** | Resource Allocation Problem | Google | Optimizing server resources under budget and performance constraints. |
| **Unconstrained Optimization** | Neural Network Training | Facebook | Minimizing the error in predictions by adjusting weights without constraints. |
| **Discrete Optimization** | Traveling Salesman Problem (TSP) | UPS | Optimizing delivery routes to minimize travel distance and time. |
| **Continuous Optimization** | Gradient Descent for Linear Regression | Amazon | Finding the best-fit line for sales data to predict future trends. |

PES UNIVERSITY

# Applications in Electronics

| Category | Example | Company | Purpose of Use |
|---|---|---|---|
| **Constrained Optimization** | Circuit Design Optimization | Intel | Optimizing circuit parameters while adhering to size and power constraints. |
| **Unconstrained Optimization** | Signal Processing Filter Design | Sony | Designing filters to enhance audio quality without constraints on filter order. |
| **Discrete Optimization** | PCB Layout Design | Qualcomm | Optimizing the layout of printed circuit boards (PCBs) to minimize space and interference. |
| **Continuous Optimization** | Power Amplifier Design | Texas Instruments | Designing amplifiers to maximize gain while minimizing distortion in a continuous range. |

# Applications in Electrical Engineering

| Category | Example | Company | Purpose of Use |
|---|---|---|---|
| **Constrained Optimization** | Power System Load Flow Optimization | General Electric | Optimizing the distribution of electrical loads while adhering to capacity and stability constraints in power grids. |
| **Unconstrained Optimization** | Electrical Circuit Design | Siemens | Designing circuits to minimize power loss and maximize efficiency without specific constraints on component values. |
| **Discrete Optimization** | Relay Coordination in Protection Systems | ABB | Optimizing the settings of relays in electrical protection systems to ensure reliability and minimize outages. |
| **Continuous Optimization** | Transformer Sizing | Schneider Electric | Determining the optimal size of transformers to maximize efficiency while minimizing costs in continuous load scenarios. |

PES UNIVERSITY

# Applications in Biotechnology

| Category | Example | Company | Purpose of Use |
|---|---|---|---|
| **Constrained Optimization** | Protein Folding Optimization | Genentech | Optimizing protein folding to ensure stability and functionality under specific biological constraints. |
| **Unconstrained Optimization** | DNA Sequencing Design | Illumina | Designing DNA sequencing techniques to maximize speed and accuracy without specific technological constraints. |
| **Discrete Optimization** | CRISPR Gene Editing Target Selection | Editas Medicine | Optimizing the selection of gene targets in CRISPR systems to ensure high specificity and efficacy. |
| **Continuous Optimization** | Bioreactor Control Optimization | Thermo Fisher Scientific | Determining the optimal control settings for bioreactors to maximize cell growth and product yield under continuous operation. |

# Applications in Computer Science

## 1. Combinatorial Optimization

| Example Problem | Description | Company/Industry |
|---|---|---|
| Traveling Salesman Problem | Finding the shortest possible route that visits a set of cities and returns to the origin city. | **Logistics and Delivery Companies** (e.g., UPS, FedEx) |
| Knapsack Problem | Selecting a subset of items with given weights and values to maximize the total value without exceeding a weight limit. | **E-commerce Platforms** (e.g., Amazon) for optimizing inventory and shipment |
| Graph Coloring Problem | Assigning colors to the vertices of a graph such that no two adjacent vertices share the same color using the least number of colors. | **Scheduling Applications** (e.g., Google Calendar) |

# Applications in Computer Science

## 2. Network Optimization

| Example Problem | Description | Company/Industry |
|---|---|---|
| **Shortest Path Problem** | Finding the shortest path between two nodes in a graph (e.g., Dijkstra's algorithm). | **Navigation Apps** (e.g., Google Maps, Waze) |
| **Maximum Flow Problem** | Determining the maximum flow in a flow network from a source to a sink (e.g., Ford-Fulkerson algorithm). | **Telecommunications** (e.g., optimizing data flow in networks) |
| **Minimum Spanning Tree Problem** | Finding the subset of edges that connects all vertices in a graph with the minimum total edge weight (e.g., Kruskal's or Prim's algorithm). | **Network Design** (e.g., ISP companies) for optimizing cable layouts |

# Applications in Computer Science

## 3. Integer Programming

| Example Problem | Description | Company/Industry |
|---|---|---|
| Job Scheduling Problem | Assigning jobs to machines or resources in such a way that minimizes the total completion time or maximizes throughput. | **Manufacturing Firms** (e.g., General Motors) for optimizing production schedules |
| Bin Packing Problem | Packing a set of items of different sizes into a finite number of bins with a fixed capacity in a way that minimizes the number of bins used. | **Shipping Companies** (e.g., FedEx, UPS) for optimizing package sizes in containers |
| Vehicle Routing Problem | Determining the optimal routes for a fleet of vehicles to deliver goods to a set of customers while minimizing costs. | **Food Delivery Services** (e.g., DoorDash, Uber Eats) |

# Applications in Computer Science

## 4. Dynamic Programming

| Example Problem | Description | Company/Industry |
|---|---|---|
| Fibonacci Sequence Calculation | Using a recursive approach with memoization to efficiently compute Fibonacci numbers. | **Software Development** (any algorithm optimization) |
| Edit Distance Problem | Finding the minimum number of edits (insertions, deletions, substitutions) required to transform one string into another. | **Text Processing Software** (e.g., Grammarly for spell-checking and suggestions) |
| Longest Common Subsequence | Finding the longest subsequence that appears in the same relative order in both sequences (not necessarily contiguous). | **Genetic Research** (bioinformatics for DNA sequence analysis) |

# Applications in Computer Science

## 5. Linear Programming

| Example Problem | Description | Company/Industry |
|---|---|---|
| **Resource Allocation Problem** | Allocating resources among competing activities in such a way as to maximize profit or minimize cost while satisfying constraints. | **Financial Institutions** (e.g., investment companies optimizing portfolios) |
| **Diet Problem** | Determining the optimal combination of food items to meet nutritional requirements while minimizing cost. | **Nutrition Apps** (e.g., MyFitnessPal) |
| **Production Planning Problem** | Optimizing production levels of multiple products to maximize profit while adhering to resource constraints. | **Manufacturing Companies** (e.g., Toyota) |

# Applications in Computer Science

| Name of the Problem | Type of Optimization Problem |
|---|---|
| Shortest path routing in networks | Graph Optimization |
| Job scheduling in operating systems | Scheduling Optimization |
| Resource allocation in cloud computing | Integer Programming |
| Data clustering for machine learning | Heuristic Optimization |
| Load balancing in distributed systems | Network Optimization |
| Network design for telecommunications | Linear Programming |
| Vehicle routing problem (VRP) | Combinatorial Optimization |
| Text classification | Quadratic Programming |
| Image recognition | Nonlinear Programming |
| Database query optimization | Heuristic Optimization |

# Applications in Computer Science

| Name of the Problem | Type of Optimization Problem |
|---|---|
| Recommendation systems | Multi-objective Optimization |
| Search engine optimization | Constrained Optimization |
| Memory management in operating systems | Dynamic Programming |
| Bandwidth allocation in communication networks | Linear Programming |
| Graph coloring problem for register allocation | Graph Optimization |
| Path planning for autonomous vehicles | Motion Planning Optimization |
| Game theory strategy optimization | Strategic Optimization |
| Compiler optimization techniques | Heuristic Optimization |
| Network traffic management | Network Optimization |
| Feature selection in data mining | Nonlinear Programming |

# Components of Optimization Problem

## Model Design of an Optimization Problem

1. Decision Variables
2. Objective Function
3. Constraints
4. Feasible Region
5. Optimal Solution
6. Parameters
7. Type of Optimization Problem

# Components of Optimization Problem

## Components of an Optimization Problem

1. Decision Variables
2. Objective Function
3. Constraints
4. Feasible Region
5. Optimal Solution
6. Parameters
7. Type of Optimization Problem

# Components of Optimization Problems

1. **Decision Variables**

These are the variables that can be **controlled or adjusted** in order to **achieve the optimal value of the objective function.** They are the **unknown quantities** that we need to **determine.**

**Example:** In the **house price prediction**, the slope ($\beta_1$) represents how much the **price changes per unit** of **square footage**, and the intercept ($\beta_0$) represents the price when square footage is zero.

# Components of Optimization Problems

## 2. Objective Function

This is the function that needs to be **maximized or minimized.** The goal of optimization is to find the **input** values that either **maximize or minimize** the output of this function. It **quantifies** the goal of the optimization.

**Example:** If we are predicting **house prices** based on **square footage**, the **objective function** is the squared difference between the predicted price (from the regression line) and the actual price for each house.

# Components of Optimization Problems

**3. Constraints**

Constraints are the **conditions or limitations imposed on the decision variables**.

These can be **equality constraints** (where variables must satisfy an equation) or **inequality constraints** (where variables must lie within a certain range). It defines the search space for the optimal solution.

**Example:** You might impose a constraint that the slope ($\beta_1$) must be **positive**, meaning that house prices should always increase with more square footage.

# Components of Optimization Problems

**4. Feasible Region**

The feasible region represents the **set of all possible values** for the **decision variables** that **satisfy all the constraints including non-negativity.** The optimal solution must lie within this region.

**Example:** If there’s no constraint, the feasible region includes any possible slope and intercept for the regression line.

# Components of Optimization Problems

## 5. Feasible and an Optimal Solution

**Feasible Solution**: A feasible solution is any solution that satisfies all the constraints of the optimization problem.
It may or may not provide the best outcome for the objective function.

**Optimal Solution**: Among all feasible solutions, the optimal solution is the one that maximizes (or minimizes) the objective function, depending on the goal of the optimization problem.

It represents the best achievable outcome while still adhering to all the constraints.

# Mathematical Formulation of Optimization Problems

**General Form:**

Optimize $f(x)$

Subject to $g_i(x) \le 0$, $i = 1, ..., m$

$h_j(x) = 0$, $j = 1, ..., p$

$x \in X$

**Where:**

**$f(x)$** is the **objective function**

**$g_i(x)$** are **inequality constraints**

**$h_j(x)$** are **equality constraints**

**$x$** are the **decision variables**

**$X$** is the **feasible region**

**Components**

**1. Objective Function:** $f(x)$
- To be maximized or minimized

**2. Constraints:**
- Inequality: $g_i(x) \le 0$
- Equality: $h_j(x) = 0$

3. **Decision Variables:** $x$
- Often denoted as a vector $(x_1, x_2, ..., x_n)$

4. **Feasible Region:** $X$
- Set of all $x$ that satisfy all constraints

# Mathematical Formulation of Optimization Problems

**Objective**: Predict the price of a house based on its size (in square feet) and the number of bedrooms, while minimizing the prediction error.

Decision Variables:

- $\beta_0$: Intercept (the base price when size and bedrooms are zero)
- $\beta_1$: Coefficient for size (the effect on price per square foot)
- $\beta_2$: Coefficient for the number of bedrooms (the effect on price per bedroom)

# Mathematical Formulation of Optimization Problems

**Objective Function**: We want to minimize the total prediction error (or residual sum of squares) across all the training data points. The objective function can be defined as:

$$\text{Minimize } E = \sum_{i=1}^{n} (P_i - (\beta_0 + \beta_1 \times S_i + \beta_2 \times B_i))^2$$

Where:

- $E$ is the total error (sum of squared errors).
- $P_i$ is the actual price of the house $i$.

**Objective Function**: We want to minimize the total prediction error (or residual sum of squares) across all the training data points. The objective function can be defined as:

$$\text{Minimize } E = \sum_{i=1}^{n} (P_i - (\beta_0 + \beta_1 \times S_i + \beta_2 \times B_i))^2$$

# Mathematical Formulation of Optimization Problems

## Constraints:

There are no explicit constraints in this problem regarding the decision variables since we're focusing on minimizing the error based on the features available. However, we can impose some practical constraints:

- $\beta_0 \geq 0$: The base price should be non-negative.
- $\beta_1 \geq 0$: The price increase per square foot should be non-negative.
- $\beta_2 \geq 0$: The price increase per additional bedroom should be non-negative.

# Mathematical Formulation of Optimization Problems

## Mathematical Formulation:

Putting it all together, the optimization problem can be stated as:

$$\text{Minimize } E = \sum_{i=1}^{n} \left(P_i - (\beta_0 + \beta_1 S_i + \beta_2 B_i)\right)^2$$

$$\begin{aligned} \text{Subject to: } \beta_0 &\geq 0 \\ \beta_1 &\geq 0 \\ \beta_2 &\geq 0 \end{aligned}$$

# Mathematical Formulation of Optimization Problems

## 6. Type of Optimization Problem:

$$\text{Minimize } E = \sum_{i=1}^{n} \left(P_i - (\beta_0 + \beta_1 S_i + \beta_2 B_i)\right)^2$$

$$\begin{aligned} \text{Subject to: } \beta_0 &\geq 0 \\ \beta_1 &\geq 0 \\ \beta_2 &\geq 0 \end{aligned}$$

**Objective Function**: The objective function involves minimizing the sum of squared errors, which is a nonlinear function due to the squaring operation. Therefore, this problem is not linear.

# Mathematical Formulation of Optimization Problems

## 7. Method to Solve the above Optimization Problem:

1. **Gradient Descent**:
   - This iterative optimization algorithm minimizes the objective function by updating the coefficients in the direction of the steepest descent (negative gradient).
2. **Newton's Method**:
   - This method uses second-order derivative information to find the minimum of a nonlinear function efficiently.
3. **Least Squares Estimation**:
   - This is a specific method for fitting a linear regression model by minimizing the sum of squared residuals, which can be done using techniques such as Ordinary Least Squares (OLS).
4. **Optimization Libraries**:
   - Python libraries like **SciPy** can be utilized for non-linear optimization using functions like `scipy.optimize.minimize()`.

# THANK YOU

## Mathematics for Computer Science Engineers