

Optimization Techniques (OT)

Well, As If That Would Ever Happen.

Rupak R. Gupta (RRG)

2025



Veermata Jijabai Technological Institute (VJTI)

For Nana

Abstract

This book provides a concise introduction to optimization methods, spanning linear programming, unconstrained and constrained optimization, and modern metaheuristics. Topics include the simplex and dual simplex methods, gradient- and Hessian-based approaches, Lagrangian and KKT conditions, and algorithms such as simulated annealing, genetic algorithms, TLBO, PSO, GWO, and NSGA-II. Applications to multiobjective problems, machine learning, and parallel computation highlight the practical relevance. Worked examples and algorithms ensure both conceptual clarity and problem-solving ability.

Contents

1	Linear Programming	3
1.1	Introduction	3
1.2	Simplex Method	3
1.3	Dual Simplex Method	5
1.4	Non-Simplex Methods	5
1.4.1	Graphical Method	5
1.4.2	Duality	6
2	Unconstrained Optimization	8
2.1	Condition for Local Minimizer	8
2.1.1	The Hessian Matrix	8
2.2	One-Dimensional Search Methods	10
2.2.1	Bounding Phase Method	10
2.2.2	Interval Halving Method	11
2.2.3	Golden Section Method	14
2.3	Gradient Methods	14
2.4	Newton's Method	16
2.4.1	Newton-Raphson Method	17
3	Non-Linear Constrained Optimization	19
3.1	Problems with Equality Constraints	19
3.2	Lagrange Multiplier Method	19
3.3	Karush-Kuhn-Tucker (KKT) Conditions	19
3.3.1	Stationarity	20
3.3.2	Primal Feasibility	20
3.4	Constrained Optimization	21

4	Single-Objective Optimization	23
4.1	Simulated Annealing	23
4.2	Genetic Algorithms	23
4.3	Evolutionary Algorithms	23
4.3.1	Teaching–Learning-Based Optimization (TLBO)	23
4.3.2	Particle Swarm Optimization (PSO)	24
4.3.3	Grey Wolf Optimization (GWO)	24
5	Multiobjective Optimization	26
	List of Algorithms	27
	List of Figures	27
	List of Tables	28
	Glossary	30
	References	31

Linear Programming

Outline

1.1	Introduction	3
1.2	Simplex Method	3
1.3	Dual Simplex Method	5
1.4	Non-Simplex Methods	5
1.4.1	Graphical Method	5
1.4.2	Duality	6

1.1 Introduction

Linear programs have a linear objective function and linear constraints, which may include both equalities and inequalities.^[1]

Linear programs are usually stated and analyzed in the following *standard form*:

$$\max \mathbf{c}^\top \mathbf{x}, \text{ subject to } \mathbf{Ax} = \mathbf{b}, x_i \geq 0$$

where $\mathbf{c} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m, \mathbf{A} \in \mathcal{M}_{m \times n}(\mathbb{R})$. Any problem can be easily transformed into the standard form. for instance, given the problem

$$\max \mathbf{c}^\top \mathbf{x}, \text{ subject to } \mathbf{Ax} \leq \mathbf{b}, x_i \geq 0$$

we can convert the inequality constraints to equalities by introducing a vector of *slack variables* ($\mathbf{s}$) and writing

$$\max \mathbf{c}^\top \mathbf{x}, \text{ subject to } \mathbf{Ax} + \mathbf{s} = \mathbf{b}, x_i \geq 0, s_i \geq 0$$

1.2 Simplex Method

Q. Consider the given Linear Programming Problem (LPP),

$$\max Z := x_1 + 4x_2, \text{ subject to}$$

$$2x_1 + x_2 \leq 3$$

$$3x_1 + 5x_2 \leq 9$$

$$x_1 + 3x_2 \leq 5$$

$$x_1, x_2 \geq 0$$

To convert it into the *standard form*, introduce slack variables s_i for each constraint.

$$\therefore Z - x_1 - 4x_2 - 0s_1 - 0s_2 - 0s_3 = 0$$

$$\therefore 2x_1 + x_2 + s_1 + 0s_2 + 0s_3 = 3$$

$$\therefore 3x_1 + 5x_2 + 0s_1 + s_2 + 0s_3 = 9$$

$$\therefore x_1 + 3x_2 + 0s_1 + 0s_2 + s_3 = 5$$

$$\therefore x_1, x_2, s_1, s_2, s_3 \geq 0$$

The Simplex tableau is written as follows,

Iteration 1

Row	Basic Variable	x_1	x_2	s_1	s_2	s_3	RHS	Ratio
R ₁	Z	-1	-4	0	0	0	0	0
R ₂	s_1	2	1	1	0	0	3	$3/1 = 3$
R ₃	s_2	3	5	0	1	0	9	$9/5 = 1.8$
R ₄	s_3	1	3	0	0	1	5	$5/3 \approx 1.67$

Table 1.1: Simplex Tableau

1. Select the coefficient in Z that is the **most negative**. This column would be considered as the **key column**.
2. Compute the ratio of the RHS term with that of the **key column** for each row.
3. Select the row with the *least positive ratio* as the **key row**.
4. The intersection of the key column and key row is the **key element**, which must be scaled down to 1.

Upon scaling the row down, we obtain

Row	Basic Variable	x_1	x_2	s_1	s_2	s_3	RHS	Ratio
R ₁	Z	-1	-4	0	0	0	0	
R ₂	s_1	2	1	1	0	0	3	
R ₃	s_2	3	5	0	1	0	9	
R ₄	s_3	$1/3$	1	0	0	$1/3$	$5/3$	

5. Using the **key element**, reduce all other elements of the **key column** to zero., i.e.

$$R_3 \rightarrow R_3 - 5R_4$$

$$R_2 \rightarrow R_2 - R_4$$

$$R_1 \rightarrow R_1 + 4R_4$$

Now, replace the basic variable of the key row s_3 with the variable of the key column x_2 .

Iteration 2

Basic Variable	x_1	x_2	s_1	s_2	s_3	RHS	Ratio
Z	$1/3$	0	0	0	$4/3$	$20/3$	
s_1	$5/3$	0	1	0	$-1/3$	$4/3$	
s_2	$4/3$	0	0	1	$-5/3$	$2/3$	
x_2	$1/3$	1	0	0	$1/3$	$5/3$	

$\therefore$ There is no negative element in the Z row, there is no key column.

Therefore, the optimal solution is,

$$x_1^* = 0^1, \quad x_2^* = \frac{5}{3}, \quad Z_{\max} = \frac{20}{3}$$

(Ans.)

1.3 Dual Simplex Method

The above method worked for maximization types of problems. To solve minimization problems, we can either use the *duality principle* (§1.4.2), or use a variation of the Simplex method known as the Dual Simplex method.

1.4 Non-Simplex Methods

Non-Simplex methods do not use the Simplex tableau to solve the primal problem. Some of these methods are discussed below.

1.4.1 Graphical Method

Graphical method is typically used for 2-variable optimization problems, as the plots are relatively easier to draw and visualize as opposed to higher dimensions.

Q. Consider the following primal problem.

$$\max Z := 2x_1 + 3x_2, \text{ subject to}$$

$$x_1 + 2x_2 \leq 6$$

$$2x_1 + x_2 \leq 8$$

$$x_1, x_2 \geq 0$$

¹All variables not present in the first column are zero for the optimal solution.

Inequality Constraint	Equality Constraint	$x_1 := 0$	$x_2 := 0$	Origin
$x_1 + 2x_2 \leq 6$	$x_1 + 2x_2 = 6$	(0, 3)	(6, 0)	Feasible
$2x_1 + x_2 \leq 8$	$2x_1 + x_2 = 8$	(0, 8)	(4, 0)	Feasible

Table 1.4: Graphical method

Using the coordinates calculated in table 1.4, we plot the feasible region as follows:

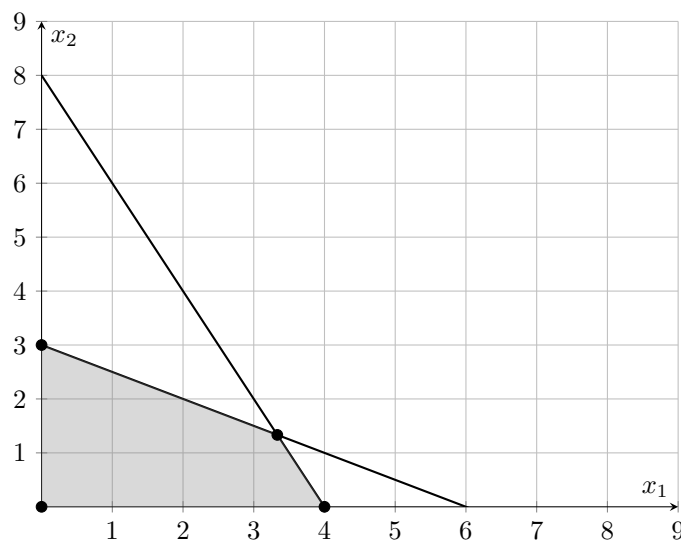


Figure 1.1: Graphical method for LPPs (The feasible region is shaded in gray.).

The intersection of the two lines is at the point $(\frac{10}{3}, \frac{4}{3})$.

Therefore, all corner points of the feasible region are candidate solutions.

Point	$Z = 2x_1 + 3x_2$
(0, 0)	0
(0, 3)	9
(4, 0)	8
$(\frac{10}{3}, \frac{4}{3})$	$\frac{32}{3} \approx 10.67$

Table 1.5: Graphical method: Evaluation of Z at candidate solutions

$\therefore$ The maximum value occurs at $(\frac{10}{3}, \frac{4}{3})$, the optimal solution is

$$x_1^* = \frac{10}{3}, \quad x_2^* = \frac{4}{3}, \quad Z_{\max} = \frac{32}{3}$$

(Ans.)

1.4.2 Duality

Duality is a principle that every LPP (the primal) has a corresponding *dual problem* such that solving the primal problem generates information about the dual problem.

Q. Consider the following primal problem.

$$\min Z := 10x_1 + 20x_2, \text{ subject to}$$

$$3x_1 + 2x_2 \geq 18$$

$$x_1 + 3x_2 \geq 8$$

$$-2x_1 + x_2 \geq -6$$

$$x_1, x_2 \geq 0$$

The **dual** of the above problem is formulated as,

$$\max W := 18y_1 + 8y_2 - 6y_3, \text{ subject to}$$

$$3y_1 + y_2 - 2y_3 \leq 10$$

$$2y_1 + 3y_2 + y_3 \leq 20$$

$$y_1, y_2, y_3 \geq 0$$

The conversion is done by transposing the coefficient matrix of the primal LPP constraints and reversing the inequalities, i.e.

$$\begin{array}{c} \begin{bmatrix} 3 & 2 \\ 1 & 3 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \geq \begin{bmatrix} 18 \\ 8 \\ -6 \end{bmatrix} \\ \min Z = 10x_1 + 20x_2 \end{array} \iff \begin{array}{c} \begin{bmatrix} 3 & 1 & -2 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \leq \begin{bmatrix} 10 \\ 20 \end{bmatrix} \\ \max W = 18y_1 + 8y_2 - 6y_3 \end{array}$$

The dual problem can be solved by Simplex method as discussed in §1.2. By the theorem of duality,

$$Z_{\min} = W_{\max} \tag{1.1}$$

Unconstrained Optimization

Outline

2.1	Condition for Local Minimizer	8
2.1.1	The Hessian Matrix	8
2.2	One-Dimensional Search Methods	10
2.2.1	Bounding Phase Method	10
2.2.2	Interval Halving Method	11
2.2.3	Golden Section Method	14
2.3	Gradient Methods	14
2.4	Newton’s Method	16
2.4.1	Newton–Raphson Method	17

2.1 Condition for Local Minimizer

Unconstrained optimization focuses on finding the minimum (or maximum) of a function without constraints. The necessary and sufficient conditions for a local minimizer rely on derivatives and **Hessian matrices**.

2.1.1 The Hessian Matrix

For a scalar multivariable function $f(\mathbf{x})$, where $\mathbf{x} := \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top \in \mathbb{R}^n$, the gradient of f is denoted by $\nabla f(\mathbf{x})$ and is defined as follows:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^\top \tag{2.1}$$

Computing the gradient is the First–Order Necessary Condition (FONC) for a local minimizer.

The *Hessian* matrix $\mathbf{H}_f$ of f is an $n \times n$ matrix comprising second-order derivatives. For $n = 2$, $f(\mathbf{x}) = f(x_1, x_2)$ and the Hessian is defined as,

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \tag{2.2}$$

For order n , the Hessian is defined as,

$$\mathbf{H}_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \quad (2.3)$$

A critical point on a function is said to be *stationary* if the gradient at that point is zero.

Depending on the nature of the Hessian matrix, the following inferences can be drawn for a stationary point:

- **Positive definite:** If $\forall \mathbf{x}, \mathbf{x}^\top \mathbf{H}_f \mathbf{x} > 0$, then f has a **local minimum**.
- **Negative definite:** If $\forall \mathbf{x}, \mathbf{x}^\top \mathbf{H}_f \mathbf{x} < 0$, then f has a **local maximum**.
- **Indefinite:** If $\mathbf{H}_f$ is neither positive nor negative definite, then f has a **saddle point**.

Computing the gradient is the Second-Order Sufficient Condition (SOSC) for a local minimizer.

Q. Consider the function:

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

Determine the nature of all critical points.

For a critical point, the gradient at that point must be zero (FONC).

$$\begin{aligned} \therefore \nabla f(x, y) &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} 2x \\ -2y \end{bmatrix} := \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

Equating the two components,

$$\begin{aligned} \implies 2x &= 0 \implies x = 0 \\ \implies -2y &= 0 \implies y = 0 \end{aligned}$$

$$\therefore \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ is a stationary point.}$$

Computing the Hessian at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ (SOSC),

$$\therefore \mathbf{H}_f = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

The eigenvalues of $\mathbf{H}_f$ are $\lambda_1 = 2$ and $\lambda_2 = -2$. Since $\lambda_1 > 0$ and $\lambda_2 < 0$, the Hessian is **indefinite** and the point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a **saddle point**. (Ans.)

Q. Consider the function:

$$f(x_1, x_2) = x_1^2 + 3x_2^2 - 4x_1x_2 + 5x_1 - 6x_2$$

Determine the nature of all critical points.

Computing the gradient (FONC),

$$\therefore \nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 4x_2 + 5 \\ -4x_1 + 6x_2 - 6 \end{bmatrix}$$

$$\begin{aligned} \therefore \nabla f(x, y) &= \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 - 4x_2 + 5 \\ -4x_1 + 6x_2 - 6 \end{bmatrix} := \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{aligned}$$

Equating the two components,

$$\implies 2x_1 - 4x_2 + 5 = 0$$

$$\implies -4x_1 + 6x_2 - 6 = 0$$

Solving the two equations, we obtain $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 2 \end{bmatrix}$ as a stationary point.

Computing the Hessian at $\begin{bmatrix} 3/2 \\ 2 \end{bmatrix}$ (SOSC),

$$\therefore \mathbf{H}_f = \begin{bmatrix} 2 & -4 \\ -4 & 6 \end{bmatrix}$$

Using the characteristic equation¹ $\lambda^2 - 8\lambda - 4 = 0$, we obtain the eigenvalues $\lambda_1 = 4 + 2\sqrt{5} > 0$ and $\lambda_2 = 4 - 2\sqrt{5} < 0$.

Therefore, the Hessian is **indefinite** and the point $\begin{bmatrix} 3/2 \\ 2 \end{bmatrix}$ is a **saddle point**. (Ans.)

2.2 One-Dimensional Search Methods

One-dimensional search methods are used to find the minimum (or maximum) of a function $f(x) \in \mathbb{R}$ along a single variable $x \in \mathbb{R}$. These methods are used to iteratively search for the optimal x that optimizes $f(x)$.

This can be done using the following bracketing methods.

2.2.1 Bounding Phase Method

Q. Consider the function $f(x) = x^2 + \frac{54}{x}$. The objective is to find $\min f(x)$.

Consider an initial guess $x_0 := 0.5$ and increment $\Delta := 0.5$.

For each iteration i , evaluate the function at $x_i - \Delta$, x_i and $x_i + \Delta$.

¹For a matrix $\mathbf{A} \in \mathcal{M}_{2 \times 2}(\mathbb{R})$, the characteristic equation is $\lambda^2 - (\text{tr } \mathbf{A})\lambda + \det \mathbf{A} = 0$.

Update the value of x_i as follows:

$$x_{i+1} := \begin{cases} x_i - 2^i \Delta; & f(x_i - \Delta) \leq f(x_i) \leq f(x_i + \Delta) \\ x_i + 2^i \Delta; & f(x_i - \Delta) \geq f(x_i) \geq f(x_i + \Delta) \end{cases} \quad (2.4)$$

The algorithm terminates when $f(x_{i+1}) \not\leq f(x_i)$. The optimal solution x^* is said to lie in the interval (x_{i-1}, x_{i+1}) .

Iteration 0

$$\begin{aligned} x_0 - \Delta = 0.5 - 0.5 = 0.0 &\implies f(0.0) = 0.0^2 + \frac{54}{0.0} \rightarrow \infty^2 \\ x_0 = 0.5 &\implies f(0.5) = 0.5^2 + \frac{54}{0.5} = 108.25 \\ x_0 + \Delta = 0.5 + 0.5 = 1.0 &\implies f(1.0) = 1.0^2 + \frac{54}{1.0} = 55.0 \end{aligned}$$

$\therefore \infty > 108.25 > 55.0$, the minimum is towards the right.

$$\begin{aligned} x_1 &:= x_0 + 2^0 \Delta \\ &= 0.5 + 1 \cdot 0.5 \\ \therefore x_1 &= 1.0 \end{aligned}$$

Tabulating remaining entries,

i	x_i	x_{i+1}	$f(x_i)$	$f(x_{i+1})$	$f(x_{i+1}) < f(x_i)$
0	0.5	1.0	108.25	55.0	✓
1	1.0	2.0	55.0	31.0	✓
2	2.0	4.0	31.0	29.5	✓
3	4.0	8.0	29.5	70.75	×

$\therefore$ Optimal solution $x^* \in (x_2, x_4) \implies x^* \in (2.0, 8.0)$.

Therefore, the approximate optimal solution $x^* \approx \frac{2.0 + 8.0}{2} = 5.0$ and $f(x^*) \approx 35.8$. (Ans.)

2.2.2 Interval Halving Method

Interval halving method is a bracketing method used to find the optimum solution within an interval $[a, b]$.

The algorithm is as follows:

²More rigorously, if $a := x_0 - \Delta$, then $f(a) := \lim_{x \rightarrow a^+} f(x)$, which in this case tends to infinity.

Algorithm 1: Interval Halving Method**Input:** Initial interval $[a, b]$, tolerance $\epsilon > 0$ **Output:** Approximate minimizer x^*

```

1 while  $b - a > \epsilon$  do
2   Compute midpoint:  $x_m \leftarrow \frac{a + b}{2}$ ;
3   Compute test points:  $x_1 \leftarrow \frac{a + x_m}{2}$ ,  $x_2 \leftarrow \frac{x_m + b}{2}$ ;
4   Evaluate:  $f(x_1), f(x_m), f(x_2)$ ;
5   if  $f(x_1) < f(x_m)$  then
6      $b \leftarrow x_m$ ;
7   else
8     if  $f(x_2) < f(x_m)$  then
9        $a \leftarrow x_m$ ;
10    else
11       $a \leftarrow x_1$ ;
12       $b \leftarrow x_2$ ;
13 Set  $x^* \leftarrow \frac{a + b}{2}$ ;
14 return  $x^*$ ;

```

Q. Consider the function $f(x) = x^2 + \frac{54}{x}$. The objective is to find $\min f(x)$.

Consider an initial interval $a := 0, b := 5$ and tolerance $\epsilon := 10^{-3}$.

Iteration 1 $x_m = \frac{0+5}{2} = 2.5$. Now, $x_1 = \frac{0+2.5}{2} = 1.25$ and $x_2 = \frac{2.5+5}{2} = 3.75$.

Evaluating x_1, x_m and x_2 ,

$$f(1.25) = 44.7625$$

$$f(2.5) = 27.85$$

$$f(3.75) = 28.4625$$

$$\therefore f(x_1) > f(x_m), f(x_2) > f(x_m).$$

$$\implies a := 1.25$$

$$\implies b := 3.75$$

$$b - a = 3.75 - 1.25 = 2.5 > \epsilon$$

Iteration 2 $x_m = \frac{1.25+3.75}{2} = 2.5$. Now, $x_1 = \frac{1.25+2.5}{2} = 1.875$ and $x_2 = \frac{2.5+3.75}{2} = 3.125$.

Evaluating x_1, x_m and x_2 ,

$$f(1.875) = 32.315625$$

$$f(2.5) = 27.85$$

$$f(3.125) = 27.045625$$

$$\therefore f(x_1) > f(x_m), f(x_2) < f(x_m).$$

$$\implies a := 2.5$$

$$b - a = 3.75 - 2.5 = 1.25 > \epsilon$$

and so on, until $b - a < \epsilon$.

The final bracket obtained would be,

$$[a, b] \approx [2.9996, 3.0002]$$

And optimal solution $x^* = \frac{a+b}{2} \approx 2.9999$

Therefore, the approximate optimal solution $x^* \approx 2.9999$ and $f(x^*) \approx 27.0000$. **(Ans.)**

2.2.3 Golden Section Method

The golden section method is very similar to the method discussed in §2.2.2. Instead of dividing the two segments of the interval equally, they are divided in the *golden ratio* (φ).

Where,

$$\varphi := \frac{\sqrt{5} + 1}{2} \approx 1.618$$

We define a constant φ' as follows:

$$\varphi' := \frac{1}{\varphi} = \frac{\sqrt{5} - 1}{2} \approx 0.618$$

The algorithm is as follows:

Algorithm 2: Golden Section Search

Input: Initial interval $[a, b]$, tolerance $\epsilon > 0$

Output: Approximate minimizer x^*

```

1 Set  $\varphi' \leftarrow \frac{\sqrt{5} - 1}{2}$ ;
2 Compute  $x_1 \leftarrow a + (1 - \varphi')(b - a)$ ;
3 Compute  $x_2 \leftarrow a + \varphi'(b - a)$ ;
4 Evaluate  $f(x_1), f(x_2)$ ;
5 while  $b - a > \epsilon$  do
6   if  $f(x_1) < f(x_2)$  then
7      $b \leftarrow x_2$ ;
8      $x_2 \leftarrow x_1$ ;
9      $f(x_2) \leftarrow f(x_1)$ ;
10     $x_1 \leftarrow a + (1 - \varphi')(b - a)$ ;
11    Evaluate  $f(x_1)$ ;
12  else
13     $a \leftarrow x_1$ ;
14     $x_1 \leftarrow x_2$ ;
15     $f(x_1) \leftarrow f(x_2)$ ;
16     $x_2 \leftarrow a + \varphi'(b - a)$ ;
17    Evaluate  $f(x_2)$ ;
18 Set  $x^* \leftarrow \frac{a + b}{2}$ ;
19 return  $x^*$ ;
```

Solving the earlier problem using this method is left as an exercise to the reader.

2.3 Gradient Methods

Gradient-based methods are iterative techniques used to find the minimum (or maximum) of a function when no constraints are present. These methods rely on the first derivative (gradient) to guide the search toward an optimal solution.

The rule for minimization is,

$$\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)})$$

where we define α to be the *step size* (or *learning rate*).

Q. Minimize the function $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1 + 2x_2 + 3$ using gradient descent.

The gradient vector is,

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 - 2 \\ 2x_2 + 2 \\ 2x_3 \end{bmatrix}$$

Consider an initial guess of $\mathbf{x}^{(0)} := \begin{bmatrix} 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$ and $\alpha := 0.5$.

Iteration 1

$$\begin{aligned} \mathbf{x}^{(1)} &:= \mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)}) \\ &= \begin{bmatrix} 0.00 \\ 0.00 \\ 0.00 \end{bmatrix} - 0.5 \begin{bmatrix} -2.00 \\ 2.00 \\ 0.00 \end{bmatrix} \\ \therefore \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} &= \begin{bmatrix} 1.00 \\ -1.00 \\ 0.00 \end{bmatrix} \end{aligned}$$

Iteration 2

$$\begin{aligned} \mathbf{x}^{(2)} &:= \mathbf{x}^{(1)} - \alpha \nabla f(\mathbf{x}^{(1)}) \\ &= \begin{bmatrix} 1.00 \\ -1.00 \\ 0.00 \end{bmatrix} - 0.5 \begin{bmatrix} 0.00 \\ 0.00 \\ 0.00 \end{bmatrix} \\ \therefore \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} &= \begin{bmatrix} 1.00 \\ -1.00 \\ 0.00 \end{bmatrix} \end{aligned}$$

Since there is no further change, the solution has converged.

Therefore, the optimal solution is $\mathbf{x}^* = \begin{bmatrix} 1.00 \\ -1.00 \\ 0.00 \end{bmatrix}$ and $f(x_1^*, x_2^*, x_3^*) = f(1.00, -1.00, 0.00) = 1.00$. **(Ans.)**

Q. Minimize the function $f(x_1, x_2) = x_1^2 + 3x_2^2 - 4x_1x_2 + 5x_1 - 6x_2$ using gradient descent.

The gradient vector is,

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 4x_2 + 5 \\ -4x_1 + 6x_2 - 6 \end{bmatrix}$$

Consider an initial guess of $\mathbf{x}^{(0)} := \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$ and $\alpha := 0.01$.

Iteration 1

$$\begin{aligned}\mathbf{x}^{(1)} &:= \mathbf{x}^{(0)} - \alpha \nabla f(\mathbf{x}^{(0)}) \\ \therefore \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} &= \begin{bmatrix} -0.05 \\ 0.06 \end{bmatrix}\end{aligned}$$

Iteration 2

$$\begin{aligned}\mathbf{x}^{(2)} &:= \mathbf{x}^{(1)} - \alpha \nabla f(\mathbf{x}^{(1)}) \\ \therefore \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} &= \begin{bmatrix} -0.0966 \\ 0.1144 \end{bmatrix}\end{aligned}$$

Iteration 3

$$\begin{aligned}\mathbf{x}^{(3)} &:= \mathbf{x}^{(2)} - \alpha \nabla f(\mathbf{x}^{(2)}) \\ \therefore \begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \end{bmatrix} &\approx \begin{bmatrix} -0.1401 \\ 0.1637 \end{bmatrix}\end{aligned}$$

It is observed that there is no convergence as the function is unbounded below. Therefore, the algorithm **diverges**. (Ans.)

2.4 Newton's Method

Newton's method is a second-order optimization technique that converges faster than gradient descent by incorporating second derivatives (Hessian matrix).

The rule for minimization is,

$$\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} - \mathbf{H}_f^{-1} \nabla f(\mathbf{x}^{(k)}) \quad (2.5)$$

Newton's method converges quadratically if $\mathbf{H}_f$ is *positive definite*.

Q. Minimize the function $f(x_1, x_2) = x_1^2 + 3x_2^2 - 4x_1x_2 + 5x_1 - 6x_2$ using Newton's method.

The gradient vector is,

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 - 4x_2 + 5 \\ -4x_1 + 6x_2 - 6 \end{bmatrix}$$

The Hessian matrix is,

$$\mathbf{H}_f = \begin{bmatrix} 2 & -4 \\ -4 & 6 \end{bmatrix}$$

Inverting the Hessian,

$$\therefore \mathbf{H}_f^{-1} = \begin{bmatrix} 3/2 & 1 \\ 1 & 1/2 \end{bmatrix}$$

Consider an initial guess of $\mathbf{x}^{(0)} := \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$.

Iteration 1

$$\begin{aligned}\mathbf{x}^{(1)} &:= \mathbf{x}^{(0)} - \mathbf{H}_f^{-1} \nabla f(\mathbf{x}^{(0)}) \\ \therefore \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} &= \begin{bmatrix} 1.50 \\ 2.00 \end{bmatrix}\end{aligned}$$

Iteration 2

$$\mathbf{x}^{(2)} := \mathbf{x}^{(1)} - \mathbf{H}_f^{-1} \nabla f(\mathbf{x}^{(1)})$$

$$\therefore \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} = \begin{bmatrix} 1.50 \\ 2.00 \end{bmatrix}$$

Since there is no further change, the solution has converged.

$$\therefore \text{The optimal solution is } \mathbf{x}^* = \begin{bmatrix} 1.50 \\ 2.00 \end{bmatrix} \text{ and } f(\mathbf{x}_1^*, \mathbf{x}_2^*) = f(1.50, 2.00) = -2.25. \quad (\text{Ans.})$$

2.4.1 Newton–Raphson Method

For one-dimensional problems ($f : \mathbb{R} \rightarrow \mathbb{R}$), the objective typically is to **find a root** of the real-valued function f (i.e. x s.t. $f(x) = 0$). This root can be approximated iteratively by using the following rule,

$$x^{(k+1)} := x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} \quad (2.6)$$

Important Distinction

If the objective is to **optimize** the function f , we are looking for the stationary points (i.e. x s.t. $f'(x) = 0$). In this case, the problem is identical to the one presented in §2.4, where $f'(x)$ is equivalent to the gradient and $f''(x)$ is equivalent to the Hessian for one dimension. The corresponding rule of minimization is therefore,

$$x^{(k+1)} := x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})} \quad (2.7)$$

which is the same as eq. (2.5).

Q. Approximate a root of the polynomial $x^3 - 4x + 1$ using Newton–Raphson method.

Define $f(x) := x^3 - 4x + 1$.

Differentiating f with respect to x ,

$$\implies f'(x) = 3x^2 - 4$$

Consider an initial guess of $x^{(0)} := 0.00$.

Applying Newton–Raphson method,

Iteration 1

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}$$

$$= 0.00 - \frac{1}{-4}$$

$$\therefore x^{(1)} = -0.25$$

Iteration 2

$$\begin{aligned}x^{(2)} &= x^{(1)} - \frac{f(x^{(1)})}{f'(x^{(1)})} \\&\approx -0.25 - \frac{1.9844}{-3.8125} \\ \therefore x^{(2)} &\approx 0.2705\end{aligned}$$

Iteration 3

$$\begin{aligned}x^{(3)} &= x^{(2)} - \frac{f(x^{(2)})}{f'(x^{(2)})} \\&\approx 0.2705 - \frac{-0.0622}{-3.7805} \\ \therefore x^{(3)} &\approx 0.2540\end{aligned}$$

Iteration 4

$$\begin{aligned}x^{(4)} &= x^{(3)} - \frac{f(x^{(3)})}{f'(x^{(3)})} \\&\approx 0.2540 - \frac{0.0002}{-3.8063} \\ \therefore x^{(4)} &\approx 0.2541\end{aligned}$$

Since there is no further change, the solution has converged.

$\therefore$ A root of the polynomial $x^3 - 4x + 1$ is at $\boxed{x \approx 0.2541}$.

(Ans.)

Non-Linear Constrained Optimization

Outline

3.1 Problems with Equality Constraints	19
3.2 Lagrange Multiplier Method	19
3.3 Karush–Kuhn–Tucker (KKT) Conditions	19
3.3.1 Stationarity	20
3.3.2 Primal Feasibility	20
3.4 Constrained Optimization	21

3.1 Problems with Equality Constraints

A general non-linear optimization problem with equality constraints is given by,

$$\min f(\mathbf{x})$$

Subject to equality constraints $h_i(\mathbf{x}) = 0$, for $i \in \{1, 2, \dots, m\}$, where

$f(\mathbf{x})$ is the objective function to minimize.

$h_i(\mathbf{x})$ are equality constraints.

3.2 Lagrange Multiplier Method

The Lagrangian function incorporating equality constraints is,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i h_i(\mathbf{x}) \tag{3.1}$$

where λ_i are the *Lagrange multipliers*.

3.3 Karush–Kuhn–Tucker (KKT) Conditions

To find the optimal solution, we set the Karush–Kuhn–Tucker (KKT) conditions:

3.3.1 Stationarity

This condition is obeyed when the critical point is stationary, i.e. there is no gradient.

$$\therefore \nabla \mathcal{L} = \nabla f(\mathbf{x}) + \sum_{i=1}^m \lambda_i \nabla h_i(\mathbf{x}) = 0 \quad (3.2)$$

3.3.2 Primal Feasibility

This condition is obeyed when all equality constraints are satisfied.

$$\implies h_i(\mathbf{x}) = 0 \quad \forall i \in \{1, 2, \dots, m\}.$$

Q. Consider the non-linear optimization problem

$$\min f(x_1, x_2) = x_1^2 + x_2^2$$

Subject to the equality constraint:

$$x_1 + x_2 - 1 = 0$$

Define the Lagrangian function as follows:

$$\mathcal{L}(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1)$$

Applying first derivative conditions,

$$\begin{aligned} \therefore \frac{\partial \mathcal{L}}{\partial x_1} &= 2x_1 + \lambda := 0 \\ \therefore \frac{\partial \mathcal{L}}{\partial x_2} &= 2x_2 + \lambda := 0 \\ \therefore \frac{\partial \mathcal{L}}{\partial \lambda} &= x_1 + x_2 - 1 := 0 \end{aligned}$$

Upon solving the two equations, we obtain $2x_1 = -\lambda = 2x_2 \implies x_1 = x_2$.

Substituting $x_1 = x_2$ in the constraint equation,

$$x_1 + x_2 = 1 \implies 2x_1 = 1 \implies \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

Therefore, the function value at the optimal point,

$$f(x_1^*, x_2^*) = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$$

Thus, the optimal solution is,

$x_1^* = \frac{1}{2}, \quad x_2^* = \frac{1}{2}, \quad \lambda = -1, \quad f(\mathbf{x}^*) = 0.5$

(Ans.)

Q. Given the optimization problem $\min f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 3)^2$ subject to $x_1^2 + x_2^2 - 4 = 0$, find the values of x_1 and x_2 that minimize $f(\mathbf{x})$ while satisfying the constraint.

Define the Lagrangian function as follows:

$$\mathcal{L}(x_1, x_2, \lambda) = (x_1 - 2)^2 + (x_2 - 3)^2 + \lambda(x_1^2 + x_2^2 - 4)$$

Applying first derivative conditions,

$$\begin{aligned}\therefore \frac{\partial \mathcal{L}}{\partial x_1} &= 2(x_1 - 2) + 2\lambda x_1 := 0 \\ \therefore \frac{\partial \mathcal{L}}{\partial x_2} &= 2(x_2 - 3) + 2\lambda x_2 := 0 \\ \therefore \frac{\partial \mathcal{L}}{\partial \lambda} &= x_1^2 + x_2^2 - 4 := 0\end{aligned}$$

Upon solving the two equations, we obtain $x_1(1 + \lambda) = 2$ and $x_2(1 + \lambda) = 3$.

Substituting x_1 and x_2 in the constraint equation,

$$\begin{aligned}\therefore \frac{4}{(1 + \lambda)^2} + \frac{9}{(1 + \lambda)^2} &= 4 \\ \therefore (1 + \lambda)^2 &= \frac{13}{4} \\ \therefore 1 + \lambda &= \pm \frac{\sqrt{13}}{2} \\ \therefore \lambda &= \frac{-2 \pm \sqrt{13}}{2}\end{aligned}$$

Therefore, we obtain $x_1 = \frac{2}{1 + \lambda}$ and $x_2 = \frac{3}{1 + \lambda} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \left\{ \underbrace{\begin{bmatrix} \frac{4}{\sqrt{13}} \\ \frac{6}{\sqrt{13}} \end{bmatrix}}_{x_+}, \underbrace{\begin{bmatrix} -\frac{4}{\sqrt{13}} \\ -\frac{6}{\sqrt{13}} \end{bmatrix}}_{x_-} \right\}$

Evaluating, $f(x_+) \approx 4.3675$ and $f(x_-) \approx 26.555$.

Thus, the optimal solution is,

$$\boxed{x_1^* = \frac{4}{\sqrt{13}}, \quad x_2^* = \frac{6}{\sqrt{13}}, \quad \lambda = \frac{-2 + \sqrt{13}}{2}, \quad f(x^*) \approx 4.3675}$$

(Ans.)

3.4 Constrained Optimization

Constrained Optimization aims to optimize an objective function subject to equality as well as inequality constraints.

The general form for a constrained optimization problem,

$$\min f(x) \text{ subject to } g_j(x) \leq 0, h_i(x) = 0$$

For a candidate solution, the following KKT conditions are necessary for optimality:

1. **Stationarity:** There must be no gradient at the optimal point.

$$\nabla f(\mathbf{x}^*) + \sum \lambda_i \nabla g_i(\mathbf{x}^*) + \sum \mu_j \nabla h_j(\mathbf{x}^*) = 0 \quad (3.3)$$

2. **Primal feasibility:** The candidate solution must satisfy all constraints, i.e. the solution must be *feasible*.

$$g_i(\mathbf{x}^*) \leq 0, h_j(\mathbf{x}^*) = 0 \quad (3.4)$$

3. **Dual feasibility:** Inequality multipliers cannot be negative, as they must preserve the nature of the inequality.

$$\lambda_i \geq 0 \quad (3.5)$$

4. **Complementary slackness:** For each inequality constraint, one of the following hold:

- The constraint is *active* ($g_i(\mathbf{x}^*) = 0$ and $\lambda_i > 0$), which influences the solution.
- The constraint is *inactive* ($g_i(\mathbf{x}^*) < 0$ and $\lambda_i = 0$), which means it has no effect.

$$\lambda_i g_i(\mathbf{x}^*) = 0 \quad (3.6)$$

Single-Objective Optimization

Outline

4.1	Simulated Annealing	23
4.2	Genetic Algorithms	23
4.3	Evolutionary Algorithms	23
4.3.1	Teaching–Learning-Based Optimization (TLBO)	23
4.3.2	Particle Swarm Optimization (PSO)	24
4.3.3	Grey Wolf Optimization (GWO)	24

4.1 Simulated Annealing

$$P = \exp\left(-\frac{\Delta E}{T}\right) \tag{4.1}$$

4.2 Genetic Algorithms

4.3 Evolutionary Algorithms

Define clamp(x ; a , b) as follows:

$$\text{clamp}(x; a, b) := \begin{cases} a; & x \leq a \\ x; & a < x < b \\ b; & b \leq x \end{cases} \tag{4.2}$$

4.3.1 Teaching–Learning-Based Optimization (TLBO)

Teaching–Learning-Based Optimization (TLBO)

Boundary condition: $\forall i, X_i \in [a, b]$

Teaching Phase

$$X_{\text{new}} = X + \text{rand} \cdot (X_{\text{best}} - T_F X_{\text{mean}}) \tag{4.3}$$

where

$\mathbf{X}_{\text{new}}$ = new solution

$\mathbf{X}$ = current solution

rand = random number $\in (0, 1)$

$\mathbf{X}_{\text{best}}$ = teacher

T_F = teaching factor (1 or 2)

$\mathbf{X}_{\text{mean}}$ = mean of the solution

$$\mathbf{X}_{\text{new}} := \text{clamp}(\mathbf{X}_{\text{new}}; \mathbf{a}, \mathbf{b})$$

Learning Phase

For a given solution $\mathbf{X}$, choose a partner solution $\mathbf{X}_p$.

For a maximization problem,

$$\mathbf{X}_{\text{new}} = \begin{cases} \mathbf{X} + \text{rand} \cdot (\mathbf{X} - \mathbf{X}_p); & f(\mathbf{X}) > f(\mathbf{X}_p) \\ \mathbf{X} - \text{rand} \cdot (\mathbf{X} - \mathbf{X}_p); & f(\mathbf{X}) < f(\mathbf{X}_p) \end{cases} \quad (4.4)$$

For a minimization problem,

$$\mathbf{X}_{\text{new}} = \begin{cases} \mathbf{X} + \text{rand} \cdot (\mathbf{X} - \mathbf{X}_p); & f(\mathbf{X}) < f(\mathbf{X}_p) \\ \mathbf{X} - \text{rand} \cdot (\mathbf{X} - \mathbf{X}_p); & f(\mathbf{X}) > f(\mathbf{X}_p) \end{cases} \quad (4.5)$$

4.3.2 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO)

The movement of the particles is given by,

$$\mathbf{V}_i^{(t+1)} = w\mathbf{V}_i^{(t)} + c_1\text{rand}_1 \cdot (\mathbf{P}_{\text{best},i} - \mathbf{X}_i^{(t)}) + c_2\text{rand}_2 \cdot (\mathbf{G}_{\text{best}} - \mathbf{X}_i^{(t)}) \quad (4.6)$$

$$\mathbf{X}_i^{(t+1)} = \mathbf{X}_i^{(t)} + \mathbf{V}_i^{(t+1)} \quad (4.7)$$

4.3.3 Grey Wolf Optimization (GWO)

Grey Wolf Optimization (GWO)

$$\mathbf{X}^{(t+1)} = \frac{\mathbf{X}_1^{(t)} + \mathbf{X}_2^{(t)} + \mathbf{X}_3^{(t)}}{3} \quad (4.8)$$

$$\begin{array}{l|l|l} A_1 = 2a \cdot \text{rand} - a & A_2 = 2a \cdot \text{rand} - a & A_3 = 2a \cdot \text{rand} - a \\ C_1 = 2 \cdot \text{rand} & C_2 = 2 \cdot \text{rand} & C_3 = 2 \cdot \text{rand} \\ \mathbf{D}_\alpha = |C_1\mathbf{X}_\alpha - \mathbf{X}| & \mathbf{D}_\beta = |C_2\mathbf{X}_\beta - \mathbf{X}| & \mathbf{D}_\delta = |C_3\mathbf{X}_\delta - \mathbf{X}| \\ \mathbf{X}_1 = \mathbf{X}_\alpha - A_1\mathbf{D}_\alpha & \mathbf{X}_2 = \mathbf{X}_\beta - A_2\mathbf{D}_\beta & \mathbf{X}_3 = \mathbf{X}_\delta - A_3\mathbf{D}_\delta \end{array}$$

Chapter 5

Multiobjective Optimization

List of Algorithms

1	Interval Halving Method	12
2	Golden Section Search	14

List of Figures

1.1	Graphical method for LPPs (The feasible region is shaded in gray).	6
-----	--	---

List of Tables

1.1	Simplex Tableau	4
1.4	Graphical method	6
1.5	Graphical method: Evaluation of Z at candidate solutions	6

Glossary

FONC First-Order Necessary Condition 8–10

GWO Grey Wolf Optimization 1, 24

KKT Karush–Kuhn–Tucker 1, 19, 21

LPP Linear Programming Problem 3, 6, 7, 28

OT Optimization Techniques 1

PSO Particle Swarm Optimization 1, 24

RHS Right-Hand Side 4, 5

RRG Rupak R. Gupta 1

SOSC Second-Order Sufficient Condition 9, 10

TLBO Teaching–Learning-Based Optimization 1, 23

VJTI Veermata Jijabai Technological Institute 1

References

- [1] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2nd edition, 2006.

