23MAT204

Mathematics for Intelligent Systems - 3

OPTIMIZATION

Dictionary Meaning

- Optimization : an act, process, methodology or procedure of making something (as a design, system, or decision) as fully perfect, functional, or effective as possible (Minimization or Maximization)

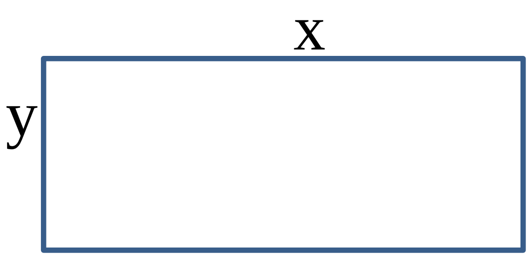
Why Optimization?

- To understand and get solutions to many questions like:
  - How can a car manufacturer get the most parts out of a piece of sheet metal?
  - How can a household moving company fit the most furniture into a truck of a certain size?
  - How can the phone company route calls to get the best use of its lines and connections?
  - How can a university schedule its classes to make the best use of classrooms without conflicts?

Similarly, engineers have to take many technological and managerial decisions at several stages. The main and important goal of all such decisions is either to minimize the effort (time / cost / power consumption) required or to maximize the desired benefit (profit / efficiency of the engine / processer speed) .

Mathematical model of an Optimization Problem

- Use a string of 100 m to form a rectangle of maximum area



Variable definitions

x = length in meters

y = breadth in meters

Objective function

Maximize xy

subject to $2x + 2y \leq 100$, $x > 0$, $y > 0$

constraints

- Form a rectangle of area $1000 m^2$, whose perimeter is as small as possible.

  Variable: x = length, y = breadth

  Minimize $2x + 2y$

  subject to $xy = 1000$, $x > 0$, $y > 0$.

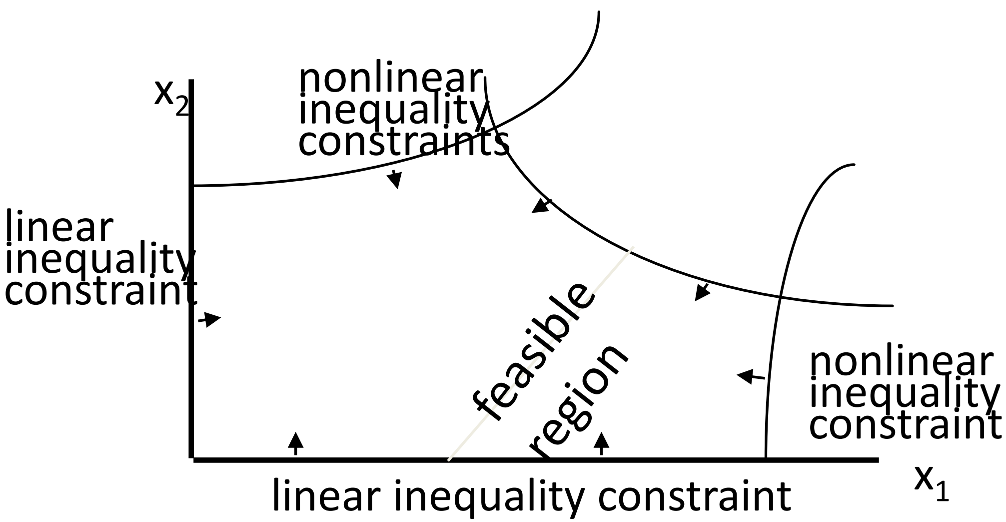
- Essential Components in an Optimization problem

  Minimize f(**x**) subject to **x** ϵ S → feasible region) set of all pts. satisfying all constraints

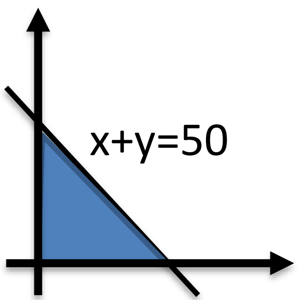
  x: set of variables

  f : objective function

  S: feasible region (set of all points that satisfy all the constraints) S = Unbdd F.R



Feasible Region for the rectangular problem



- Solution of an optimization problem: Minimize f(**x**) subject to **x** ϵ S

  x* ϵ S such that f (**x***) ≤ f (**x**) for every **x** ϵ S

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

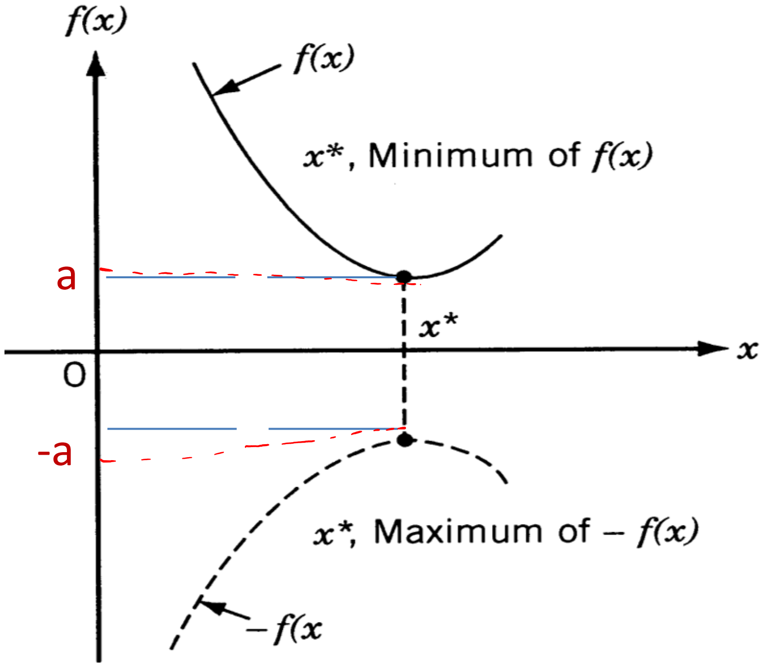
  x* may not be unique and may not even exist.

- A mathematical formulation of a standard optimization problem:

$$\textbf{Minimize } f(\mathbf{x}) \text{ subject to } \mathbf{x} \in S$$

- If x^* has a minimum of f(x), then x^* will also have the maximum of -f(x) (and viceversa) and

$$\text{Minimum value of } f(\mathbf{x}) = -\text{ Maximum value of } -f(\mathbf{x})$$



Minimum of $f(x)$ is same as maximum of $-f(x)$.

Applications

- Design of Aircrafts, racing cars
- Maximize the efficiency of a power plant
- Maximize the efficiency of a IC engine
- Optimized Scheduling of a machine
- Minimization of the manufacturing cost
- Optimum product mixing in a fractionating problem (petroleum refinery)
- Minimizing transmission loss
- Faster Communication between nodes in a communication network
- Maximizing the processor speed
- Robot path planning
- Fitting of a data
- …

Types of Optimization Problems

- Single Variable and Multi variable Optimization Problems
- Constrained and Unconstrained Optimization Problems
- Linear and Nonlinear Optimization Problems
- Continuous (real variables) and Discrete optimization (binary or integer variables)
- **Single-objective** and Multi-objective optimization
- Stochastic and deterministic optimization

Stochastic

  - Some or all of the problem data are random
  - In some cases, the constraints hold with some probabilities

Deterministic

  - No randomness in problem data and constraints

Some Optimization Problems

- Maximum area rectangle problem

Variables : x = length of rectangle, y = breadth of the rectangle

Maximize xy

subject to $2x + 2y \leq 100$, $x > 0$, $y > 0$.

Multivariable, Constrained, Non-linear, Continuous, Deterministic

- Minimum perimeter rectangle problem

Variables : x = length of rectangle, y = breadth of the rectangle

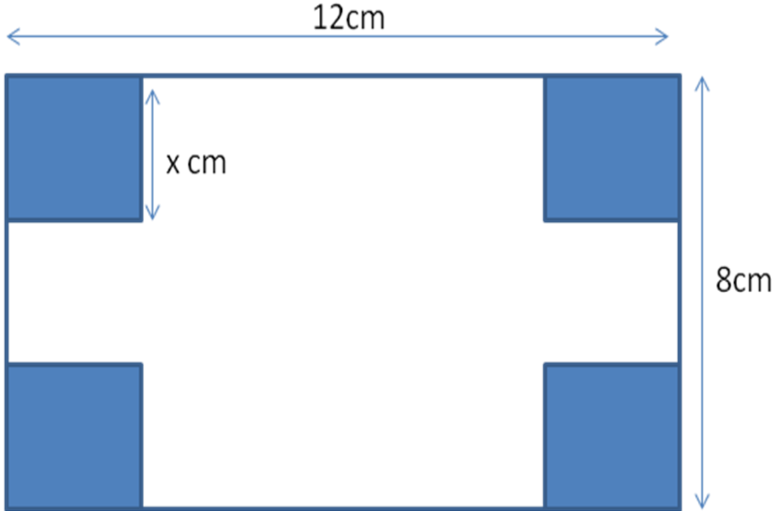
Minimize $2x + 2y$

subject to $xy = 1000$, $x > 0$, $y > 0$.

Multivariable, Constrained, Non-linear, Continuous, Deterministic

Some Optimization Problems

- An open rectangular box need to be made using a sheet of length 12cm and breadth 8cm. What would be the height of the box if the volume of it should be maximum?



Variable: Let 'x' be the height

Objective:

Maximize $V = (12-2x)(8-2x)x = 4(6-x)(4-x)x = 4x^3 - 40x^2 + 96x$

subject to x > 0

Single variable, Unconstrained, Non-linear, Continuous, Deterministic

Some Optimization Problems

Multivariable, Constrained, Linear, Discrete, Deterministic

- Two Products - Bulbs and Batteries
- Each requires two raw materials - P and Q
- One unit of bulb needs
  - 2 units of P, 4 units of Q
- One unit of battery needs
  - 3 units of P, 2 units of Q
- Raw material availability is limited
  - **24 units of P, 32 units of Q**
- Each product yields different profit/unit
  - bulb: $4 per unit
  - battery: $5 per unit

Variables:
X units of bulbs
Y units of batteries
Objective: Maximize 4x+5y

	P	Q
Bulb	2	4
Battery	3	2
Avail	24	32

Constraints:
Usage of P : $2x+3y \leq 24$
Usage of Q : $4x+2y \leq 32$

$x \geq 0, y \geq 0$, x,y positive integers

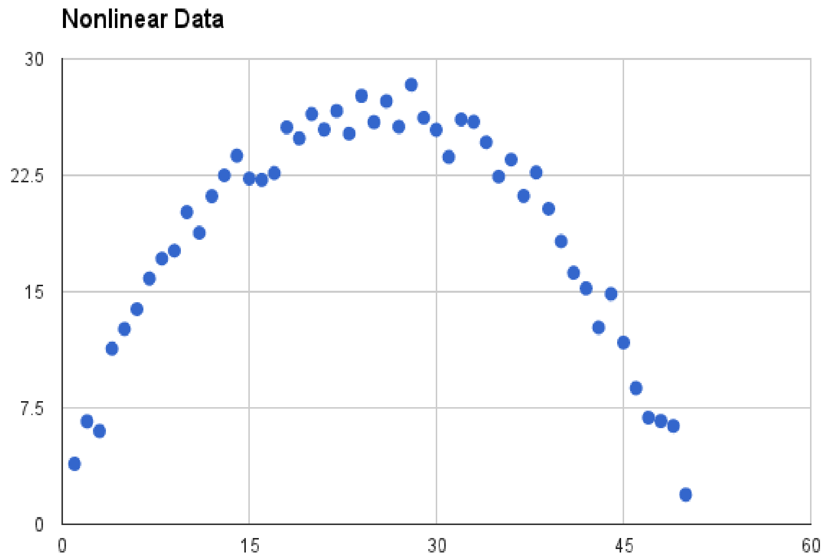
Problem: How much of each product should be produced so that total Profit is maximized?

Some Optimization Problems Contd…

- **Data Fitting Problem**: Find a model that "best" fits the observed data.

Given : (1) $\{x_i , y_i\}$ for i = 1 to n (data points)

(2) Fit Model $f(x) = ax^2 + bx + c$



Variables : a, b, c

Objective : Minimize $\sum_{i=1}^{n}(y_i - (a\,x_i{}^2 + bx_i + c))^2$ $a \neq 0$ $\boxed{\text{ie } a>0 \text{ or } a<0}$

Multivariable, Unconstrained, Non-linear, Continuous, Deterministic

Data Fitting Model

Formulate an optimization problem to fit a quadratic curve to the given data points:

X_i	Y_i
-3	8.2
-2	3.9
-1	1.1
0	0.003
1	0.99
2	4.2
3	9.8

Let the fit be $f(x) = ax^2 + bx + c$

Variables: *a, b, c*

$f(X_i) = aX_i^2 + bX_i + c$	$[Y_i-f(X_i)]^2$
9a-3b+c	$[8.2-9a+3b-c]^2$
4a-2b+c	$[3.9-4a+2b-c]^2$
a-b+c	$[1.1-a+b-c]^2$
c	$[0.003-c]^2$
a+b+c	$[0.99-a-b-c]^2$
4a+2b+c	$[4.2-4a-2b-c]^2$
9a+3b+c	$[9.8-9a-3b-c]^2$

Objective: Minimize $[8.2-9a+3b-c]^2 + [3.9-4a+2b-c]^2 + [1.1-a+b-c]^2 + [0.003-c]^2 + [0.99-a-b-c]^2 + [4.2-4a-2b-c]^2 + [9.8-9a-3b-c]^2$

Multivariable, Unconstrained, Non-linear, Continuous, Deterministic

Data Fitting Model

Least Squares

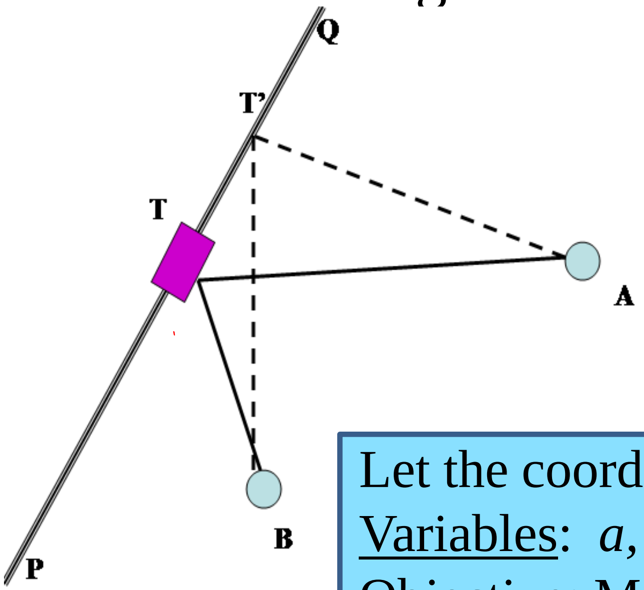
This is a least square approximation problem, the solution of which is the solution of $A^TAx=A^Tb$

$$\text{Where } A=\begin{bmatrix} 9 & -3 & 1 \\ 4 & -2 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix},\ x=\begin{bmatrix} a \\ b \\ c \end{bmatrix},\ b=\begin{bmatrix} 8.2 \\ 3.9 \\ 1.1 \\ 0.003 \\ 0.99 \\ 4.2 \\ 9.8 \end{bmatrix}$$

$$x = \overbrace{(A^TA)^{-1}A^T}^{\text{pinv}(A)}\, b$$
$$= \text{pinv}(A)\, b$$

Some Optimization Problems Contd…

- **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.



Find the coordinates of the point T on the line PQ, x + y = 5, such that the lengths of the line segments AT and BT is minimum, Given, coordinate of A is (1,2) and B is (3,1).

Let the coordinates of the point T be (a, b)

Variables: a, b

Objective: Minimize the distance AT + BT

i.e. Minimize $\sqrt{(a-1)^2+(b-2)^2} + \sqrt{(a-3)^2+(b-1)^2}$

subject to $a + b = 5$

- Multivariable, Constrained, Non-linear, Continuous, Deterministic

Some Optimization Problems Contd…

- **Diet Problem:** Propose a diet containing at least 2,000 (Kcal), at least 55 grams of protein and 800 (mg) of calcium with reference to the given table and additionally at minimum cost.

Food	Portion Size	Energy (Kcal)	Proteins (grams)	Calcium (mg)	Price ($/portion)	Limit (portions/day)
Oats	28 g	110	4	2	30	4
Chicken	100 g	205	32	12	240	3
Eggs	2 big ones	160	13	54	130	2
Milk	237 cc	160	8	285	90	8
Kuchen	170 g	420	4	22	200	2
Beans	260g	260	14	80	60	2

- **Variables:** Let X_i be the portion of food i to eat during a day

  (i=1 for oats, i=2 for chicken, i=3 for egg, i=4 for milk, i=5 for kuchen, i=6 for beans)

Diet Problem continued…

Food	Portion Size	Energy (Kcal)	Proteins (grams)	Calcium (mg)	Price ($/portion)	Limit (portions/day)
Oats	28 g	110	4	2	30	4
Chicken	100 g	205	32	12	240	3
Eggs	2 big ones	160	13	54	130	2
Milk	237 cc	160	8	285	90	8
Kuchen	170 g	420	4	22	200	2
Beans	260g	260	14	80	60	2

Minimize $30X_1 + 240\,X_2 + 130\,X_3 + 90\,X_4 + 200\,X_5 + 60\,X_6$

s.t. $110\,X_1 + 205\,X_2 + 160\,X_3 + 160\,X_4 + 420\,X_5 + 260\,X_6 \geq 2000$

$4\,X_1 + 32\,X_2 + 13\,X_3 + 8\,X_4 + 4\,X_5 + 14\,X_6 \geq 55$

$2\,X_1 + 12\,X_2 + 54\,X_3 + 285\,X_4 + 22\,X_5 + 80\,X_6 \geq 800$

$X_1 \leq 4$;

$X_2 \leq 3$;

$X_3 \leq 2$;

$X_4 \leq 8$;

$X_5 \leq 2$;

$X_6 \leq 2$;

$X_i \geq 0$ for i = 1, 2, 3, 4, 5, 6

Multivariate, Constrained, Linear, Discrete, Deterministic

Some Optimization Problems Contd…

A gear manufacturing company received an order for three specific types of gears for regular supply. The management is considering to devote the available excess capacity to one or more of the types, say A, B and C. The available capacity on the machines which might limit the output and the number of machines hours required for each unit of the respective gear is also given below:

M/C type	Available M/C hours/week	Productivity in M/C hours/unit		
		Gear A	Gear B	Gear C
Gear Hobbling M/C	250	8	2	3
Gear Shaping M/C	150	4	3	0
Gear Grinding M/C	50	2	4	1

The unit profit would be Rs. 20, Rs. 6 and Rs. 8 respectively for the gears A, B and C. Formulate the model to find how much of the gear the company should produce in order to maximize the profit?

Some Optimization Problems Contd…

- **Portfolio Optimization**
  - **Variables:** amounts to be invested in different assets
  - **Objective:** Minimize the overall risk or return variance
  - **Constraints:** budget, max/min investment per asset, minimum return

- **Development of device in electronic circuit**
  - **Variables:** device width and length
  - **Objective:** Minimize power consumption
  - **Constraints:** manufacturing limits, timing requirements, maximum area

Formulation Exercise:

1. Write a mathematical model to find the dimensions of a cylindrical tin (with top and bottom) made up of sheet metal to maximize its volume such that the total surface area is equal to 50 sqm. Classify the model based on all five classifications.
2. A firm manufactures two products A and B on which the profit earned per unit are Rs.3 and Rs.4 respectively. Each product is processed on two machines M_1 and M_2. Product A requires one minute of processing time on M_1 and two minutes on M_2 , while B requires one minute on M_1 and one minute on M_2. Machine M_1 is available for not more than 7 hours and 30minutes, while machine M_2 is available for 10 hours. Formulate the problem as a mathematical model, if the objective is to maximize the profit.

Formulation Exercise:

3. A firm manufactures headache pills in two sizes A and B. Size A contains 2 grains of Aspirin, 5 grains of bicarbonate and 1 grain of codein. Size B contains 1 grain of aspirin, 8 grains of bicarbonate and 6 grains of codein. It is found by users that it requires at least 12 grains of aspirin, 74 grains of bicarbonate and 24 grains of codein for providing immediate effect. It is required to determine the least number of pills a patient should take to get immediate relief. Formulate the problem.
4. The final product of a firm has a requirement that it must weigh exactly 150 kg. The two raw materials used are product A with cost of Rs.2/unit and B with a cost of Rs.8/unit. At least 14 units of B and not more than 20 units of A must be used. Each unit of A weighs 5kg and that of B weighs 10kg. Formulate the problem so as to know how much of each type of raw materials should be used for each unit if the cost is to be minimized.

Formulation Exercise:

5. A balanced diet for a five year old boy should contain at least 1400 Kcal of energy, 75 grams of proteins and 800 mg of calcium per day. With reference to the given table, formulate a mathematical model to decide the diet of a five year old boy, with cost minimization as the objective.

Food	Portion size	Energy (Kcal)	Proteins (gram)	Calcium (mg)	Price (Rs./portion)	Limit portion/ day
Rice	150 gram	380	14	19	25	3
Wheat	150 gram	295	21	33	42	2
Milk	250 ml	160	11	285	20	4
Dal	250 gram	180	23	52	165	2

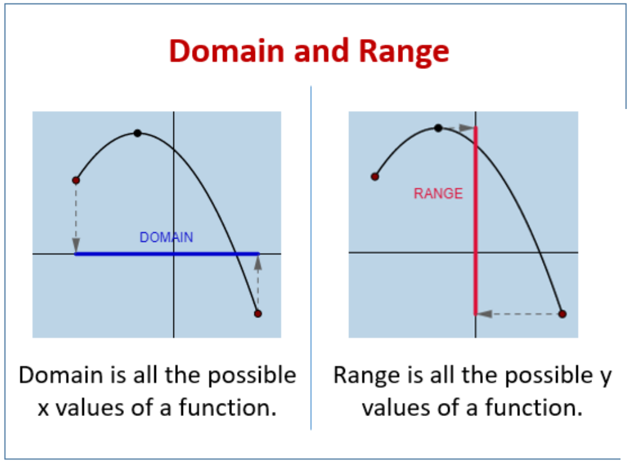
Methods of Solving Optimization Problems

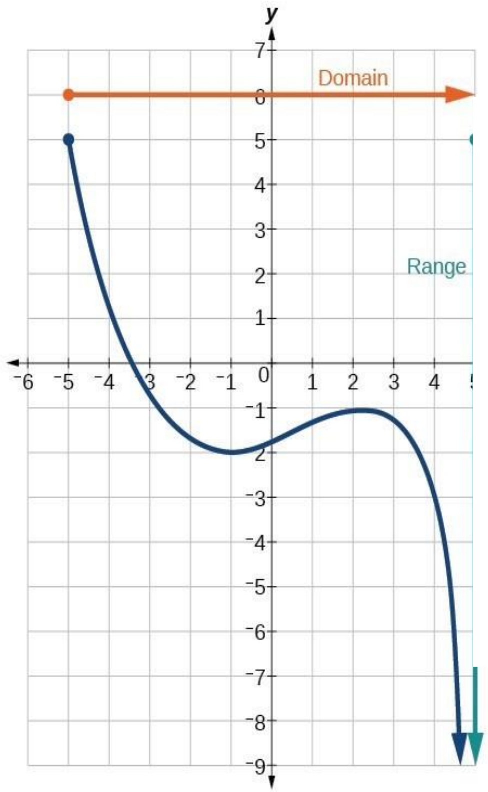
– Graphical method
– Analytical methods (or classical methods)
– Numerical Methods

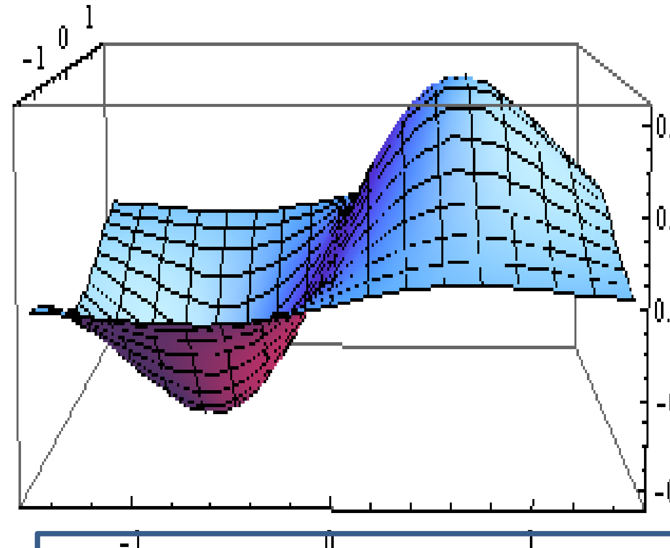
Graphical method

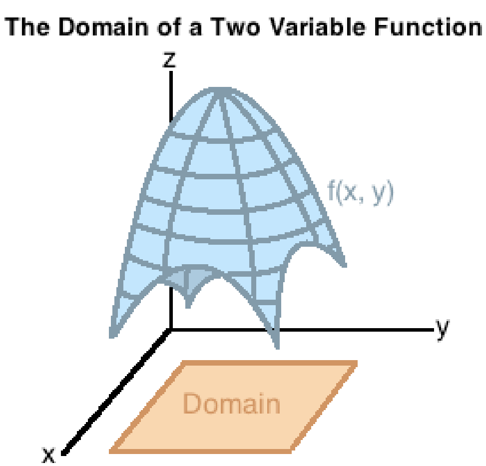
- Unconstrained or Constrained
- Optimization problems with one and two variables only

Graphical method – Domain and Range









Domain: $-1 \leq x, y \leq 1$

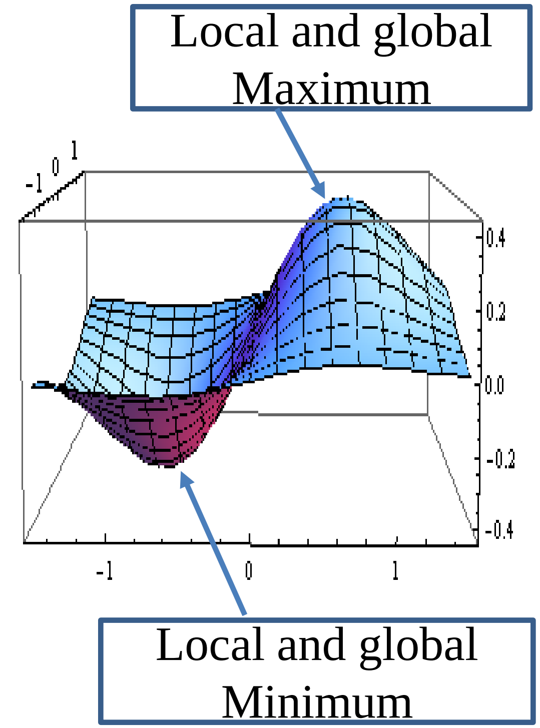
Range: $-0.4 \leq f(x, y) \leq 0.4$

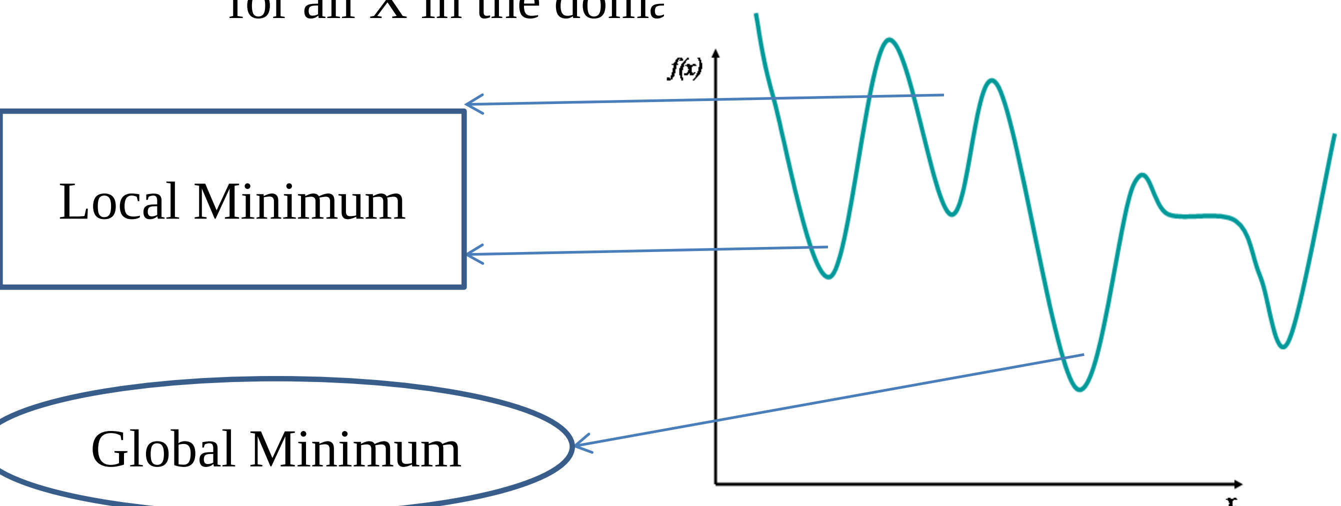
Solutions for an Optimization problem

- **Local Optimum Solution**
  - A function f(**X**) is said to have
    - A local minimum at $\mathbf{X}^*$ if $f(\mathbf{X}^*) \leq f(\mathbf{X})$
    - A local maximum at $\mathbf{X}^*$ if $f(\mathbf{X}^*) \geq f(\mathbf{X})$

    for all X in an open neighbourhood of **X***.
- **Global Optimum Solution**
  - A function f(**X**) is said to have
    - A global minimum at $\mathbf{X}^*$ if $f(\mathbf{X}^*) \leq f(\mathbf{X})$
    - A global maximum at $\mathbf{X}^*$ if $f(\mathbf{X}^*) \geq f(\mathbf{X})$

    for all X in the domain of f(**X**)





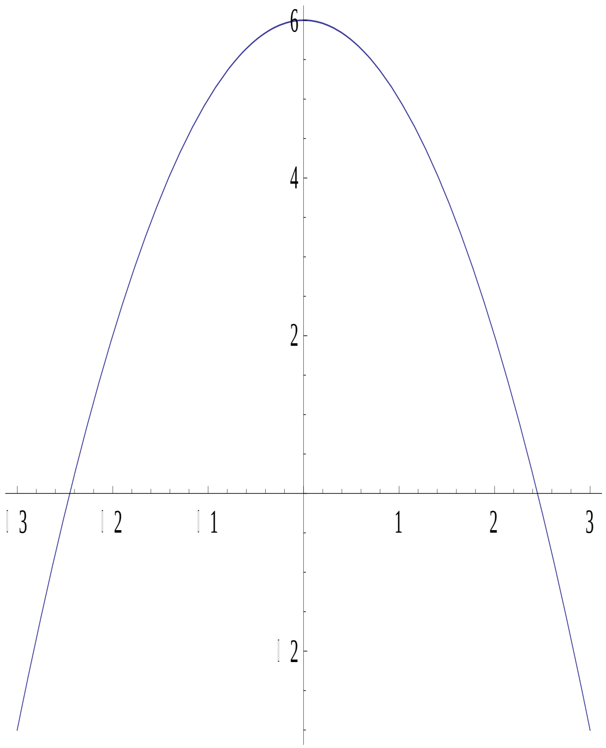
Graphical method
-Unconstrained, Single Variable Optimization Problems

$f(x) = 6 - x^2$

Local maximum and Global maximum

$x^* = 0$

$f(x^*) = 6$



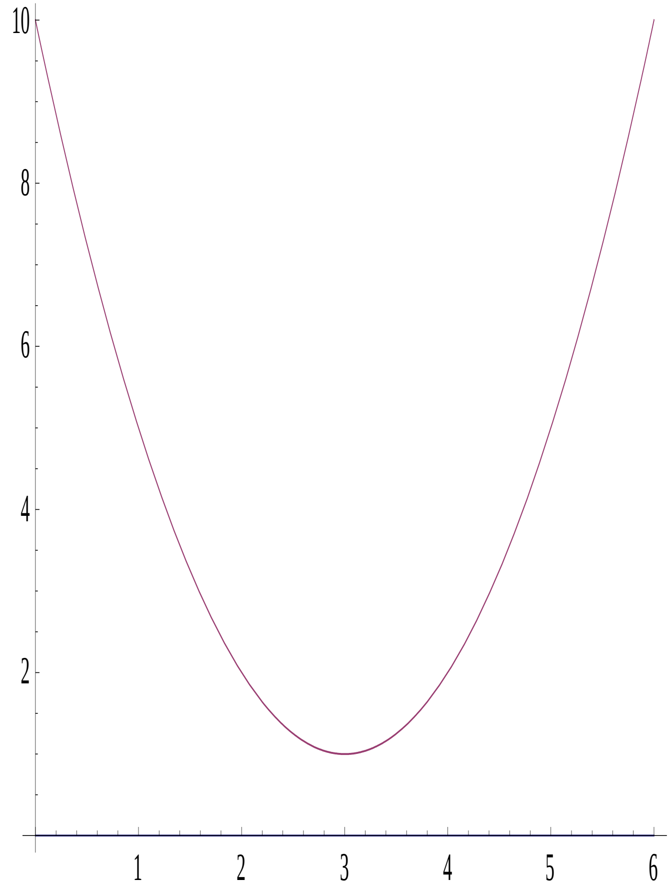
Graphical method
-Unconstrained, Single Variable Optimization Problems

$f(x) = (x-3)^2 + 1$

Local minimum and Global minimum

$x^* = 3$

$f(x^*) = 1$



Graphical method

- Single Variable Optimization Problems

Find the points of minimum and maximum (local and global) for the following functions

(a) $f(x) = x^2$ in $[-1,1]$

(b) $f(x) = x^2$ in $[0,1]$

(c) $f(x) = x^2$ in $(0,1]$

(d) $f(x) = x^2$ in $[-1,0)$

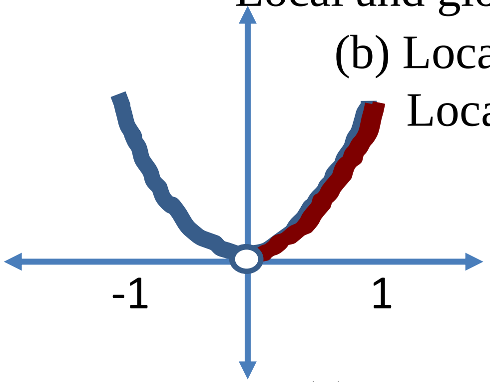
(e) $f(x) = x^2$ in $(-1,0)$

(f) $f(x) = 5x + 3$

(g) $f(x) = 5x + 3$ in $[0,5]$

(a) Local and global minimum at x=0
Local and global maximum at x= -1 and x=1

(b) Local and global minimum at x=0
Local and global maximum at x=1



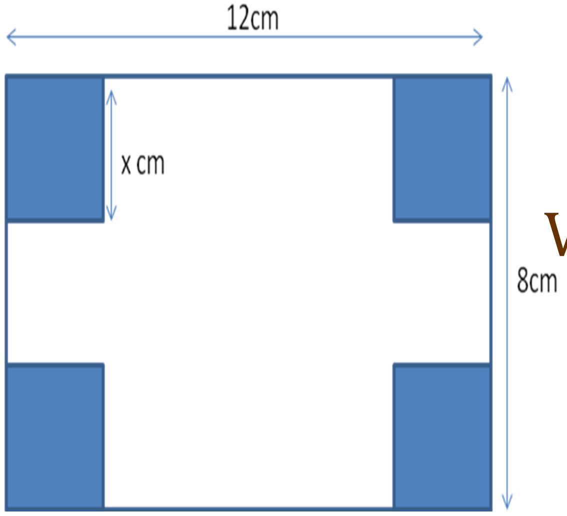
(c) No Local and global minimum
Local and global maximum at x=1

(e) and (f) No Local/global minimum/maximum

The minimum / maximum depends on not only the function but also on the region in which the optimum has to be found.

Graphical Method – Single Variable O.P.

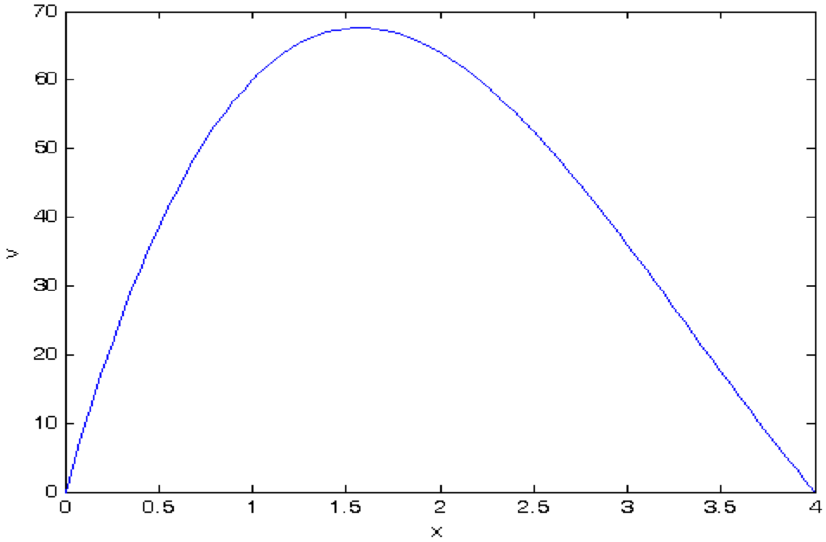
- An open rectangular box need to be made using a sheet of length 12cm and breadth 8cm. What would be the height of the box if the volume of it should be maximum?



Problem: Maximize V

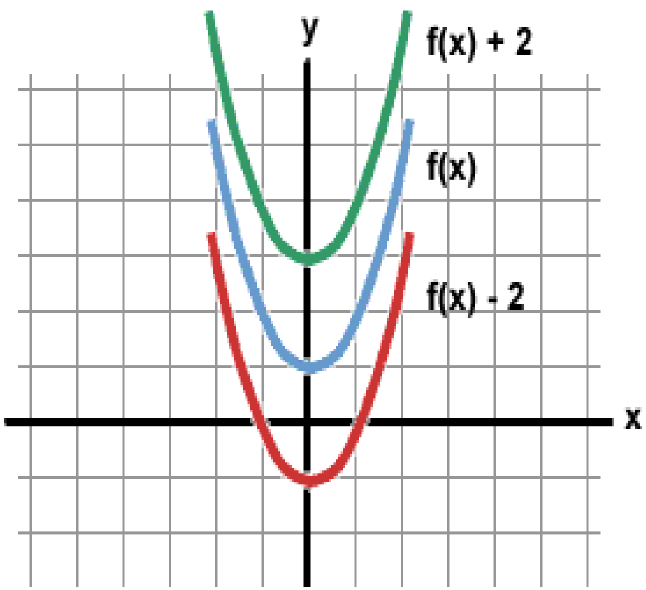
$$V = (12-2x)(8-2x)x = 4x^3 - 40x^2 + 96x$$

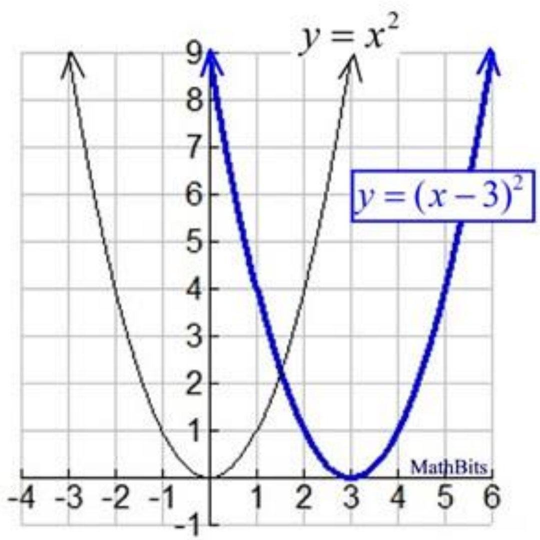
Graph of V

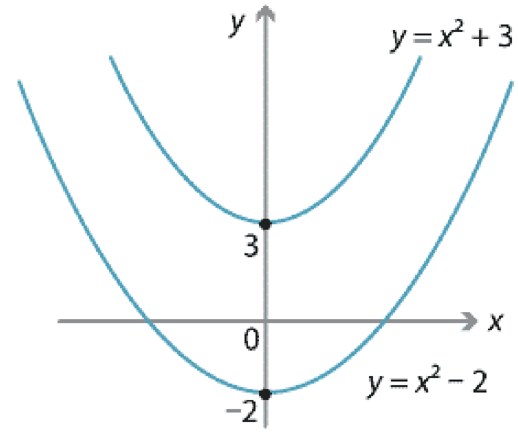


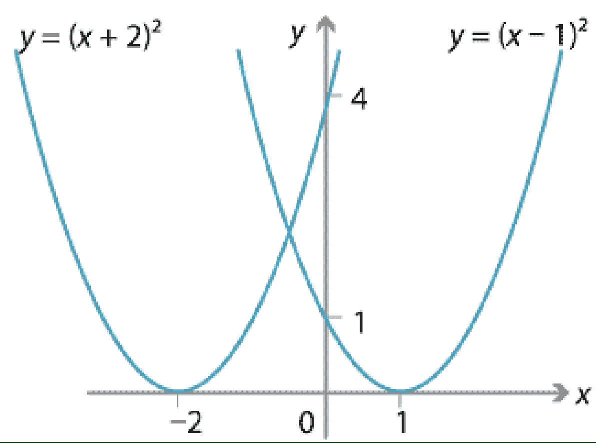
Graphical Method – S.V.O.P. – Using Shifting

- If x^* is the minimum of f(x), then x^* will also have the minimum of f(x)+a and Minimum value of [f(**x**)+a] = Minimum value of f(**x**) + a.
- If x^* is the minimum of f(x), then minimum of f(x+a) will be at x^*- a. The minimum values of f(x) and f(x+a) is the same.









Single Variable Optimization Problems

Analytical solution / Solution by Calculus Methods

Calculus - Revision

- **Increasing / Decreasing functions**
  - A function f(x) is said to be **increasing** in an interval (a,b) if the derivative, $f'(x) > 0$ in (a,b).
  - A function f(x) is said to be **decreasing** in an interval (a,b) if the derivative, $f'(x) < 0$ in (a,b).

- **Local Minimum / Local maximum**
  - The point where the function changes its nature from decreasing to increasing is a **local minimum point**($f'' > 0$).
  - The point where the function changes its nature from increasing to decreasing is a **local maximum point**($f'' < 0$).
  - Local extremum/optimum points will be points where either $\boldsymbol{f' = 0}$ or where $\boldsymbol{f'}$ **is undefined** or one of the **end-points** of the domain.

- **Absolute(Global) Minimum / Absolute(Global) maximum**
  - If absolute minimum or absolute maximum exists for a function, it should be at one of the points which has a local minimum or local maximum.
  - Can be found by finding function values at all local extremum points.

Calculus - Revision

Find the points of minimum and maximum for the following functions

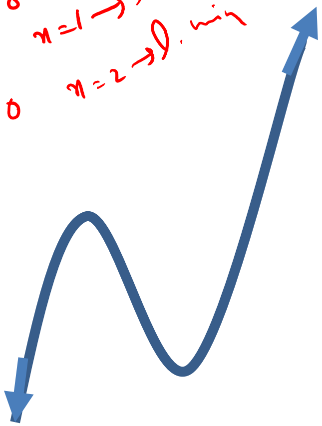
1. $2x^3 - 9x^2 + 12x + 6$

$f'(x) = 6x^2 - 18x + 12 = 6(x-1)(x-2)$

$f'(x)$ =0 at x = 1 and at x = 2 (critical points)

$f'' = 12x - 18$
$f''(1) < 0$ $x = 1 \rightarrow$ lo. max
$f''(2) > 0$ $x = 2 \rightarrow$ l. min

Interval:	(-∞, 1)	(1,2)	(2, ∞)
Sign of f':	+	-	+
Nature of fn.:	Increasing	Decreasing	Increasing



Local maximum at x=1 and local minimum at x=2.

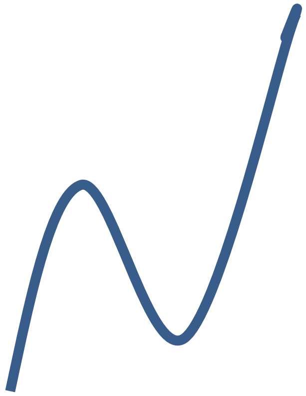
No absolute/global minimum and no absoluteglobal maximum.

2. $2x^3 - 9x^2 + 12x + 6$ in [-1, 4]

$f'(x) = 6x^2 - 18x + 12 = 6(x-1)(x-2)$

$f'(x)$ =0 at x = 1 and at x = 2 (critical points)

Interval:	(-1, 1)	(1,2)	(2, 4)
Sign of f':	+	-	+
Nature of fn.:	Increasing	Decreasing	Increasing



Local minimum at x= -1 and x=2, local maximum at x=1 and x=4.

Absolute minimum at x= -1(as f(-1)<f(2)) and

Absolute maximum at x=4 (as f(4)>f(1))

Single Variable Optimization Problem

Theorem 1:**Necessary Condition-** If a function f(x) is defined in the interval $a \leq x \leq b$ and has a local minimum/maximum at $x=x^*$, where $a<x^*<b$, and if first derivative exists at x^*, then $f'(x^*)=0$.

Theorem 2: **Sufficient Condition-** Let $f'(x^*)= f''(x^*)=\ldots=f^{n-1}(x^*)=0$, but $f^{n}(x^*)\neq 0$.

Then $x=x^*$ is a minimum point if $f^{n}(x^*)>0$ and *n is even.*

$x=x^*$ is a maximum point if $f^{n}(x^*)<0$ and *n is even.*

$x=x^*$ is neither minimum nor maximum *if n is odd.*

Single Variable Optimization Problem

Minimize f(x)

A **necessary condition** for a function with one variable to have a minimum at an interior point x=a is that, the derivative of the function should vanish at x=a or the derivative of the function does not exist at x=a.

A **sufficient condition** for a function with one variable to have a minimum at x=a is that the first non-zero derivative is an even derivative and the value of this even derivative at x=a is positive.

Single Variable Optimization Problems

Find the points of local minimum/maximum for the following functions

1. $f(x) = 2x^3 - 9x^2 + 12x + 6$

$f'(x) = 6x^2 - 18x + 12 = 6(x-1)(x-2)$

$f'(x) = 0$ at $x = 1$ and at $x = 2$ (critical points)

$f''(x) = 12x - 18;$

$f''(1) < 0,$ so x=1 is a local maximum.

$f''(2) > 0,$ so x=2 is a local minimum.

2. $f(x) = x^6 - 6x^5 + 9x^4$

$f'(x) = 6x^5 - 30x^4 + 36x^3 = 6x^3(x-2)(x-3)$

$f'(x) = 0$ at $x = 0$, $x = 2$ and $x = 3$ (critical points)

$f''(x) = 30x^4 - 120x^3 + 108x^2$

$f''(2) = -48 < 0,$ so **x=2 is a local maximum point.**

$f''(3) = 162 > 0,$ so **x=3 is a local minimum point.**

$f''(0) = 0;$

$f'''(x) = 120x^3 - 360x^2 + 216x;\ f'''(0) = 0$

$f^{iv}(x) = 360x^2 - 720x + 216;$

$f^{iv}(0) = 216 > 0,$ First non-zero derivative at zero is the fourth derivative (even)

So **x=0 is a local minimum point.**

AMRITA VISHWA VIDYAPEETHAM UNIVERSITY Established u/s 3 of UGC Act 1956

Single Variable Optimization Problems

Find the points of local minimum / local maximum for the following functions

3. $f(x) = 12x^5 - 45x^4 + 40x^3 + 5$

Answer: Critical points are 1,2,0.
Local Maximum at x=1, Local Minimum at x=2
At x=0, there is neither a local minimum nor a local maximum.

4. $f(x) = (x-1)^6$

Answer: Critical point x=1, Local Minimum at x=1

5. $f(x) = (2x-3)^7$

Answer: Critical point x=1.5 is neither a local minimum nor a local maximum.

Multivariable Unconstrained Optimization Problems

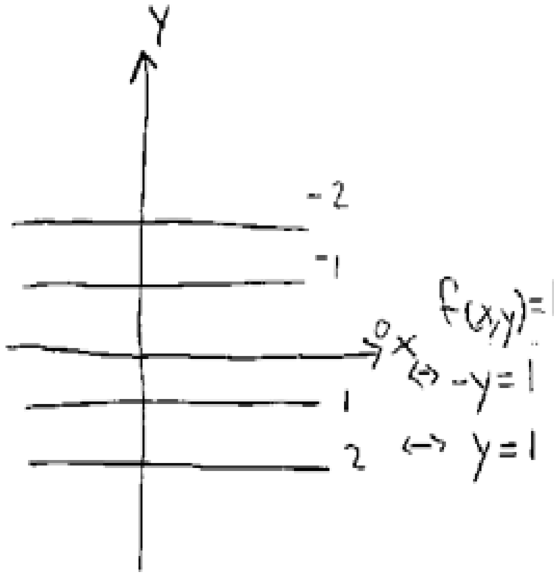
Minimize $f(x_1, x_2, \ldots, x_3)$

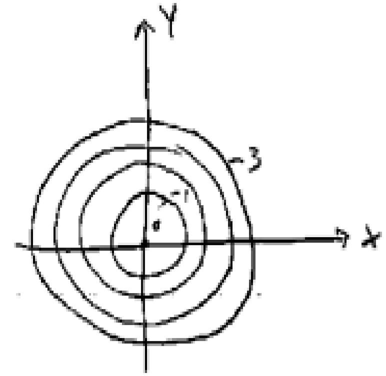
– Graphical method

– Analytical methods (or classical methods)

Contour Plots/Level curves for two variable functions

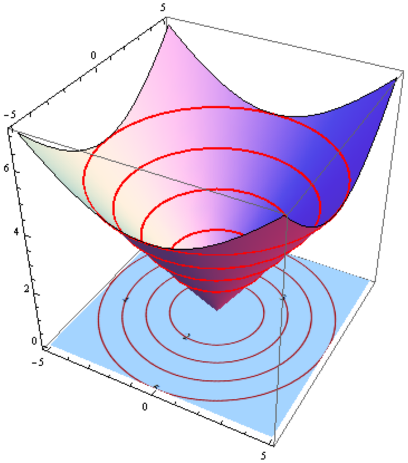
Example $f(x,y) = -y$ $\quad$ $f(x,y) = 1 - (x^2 + y^2)$

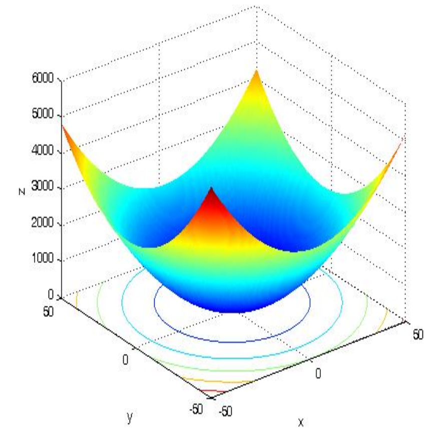


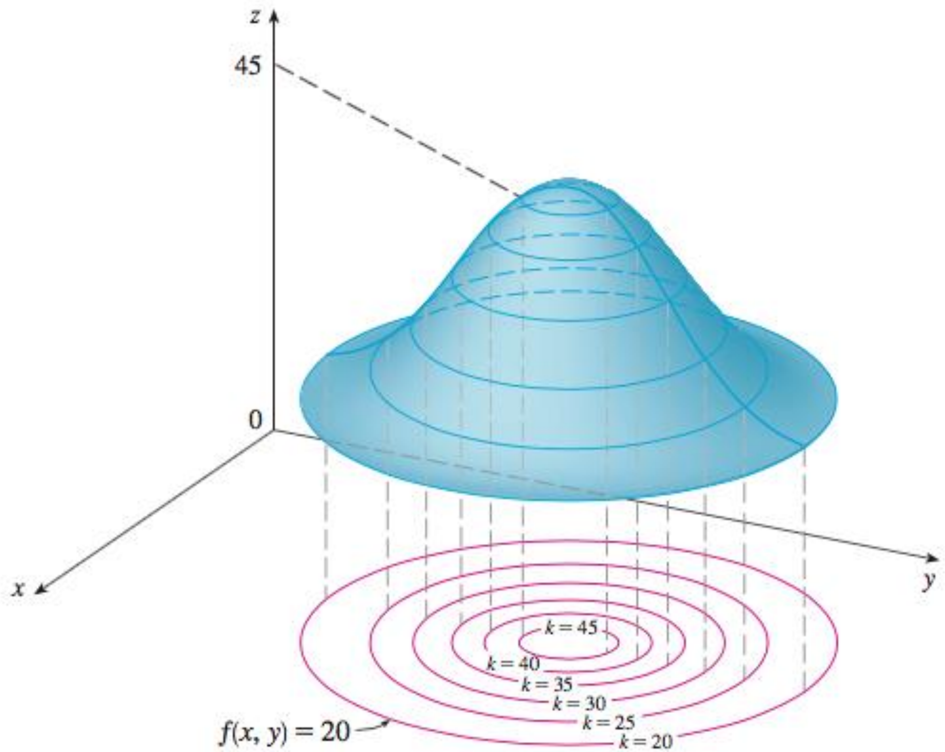


$$f = 0 \quad \Leftrightarrow \quad x^2 + y^2 = 1$$
$$f = 1 \quad \Leftrightarrow \quad x^2 + y^2 = 0$$
$$f = -1 \quad \Leftrightarrow \quad x^2 + y^2 = 2$$

Relations between Contour Plots/Level curves and surfaces





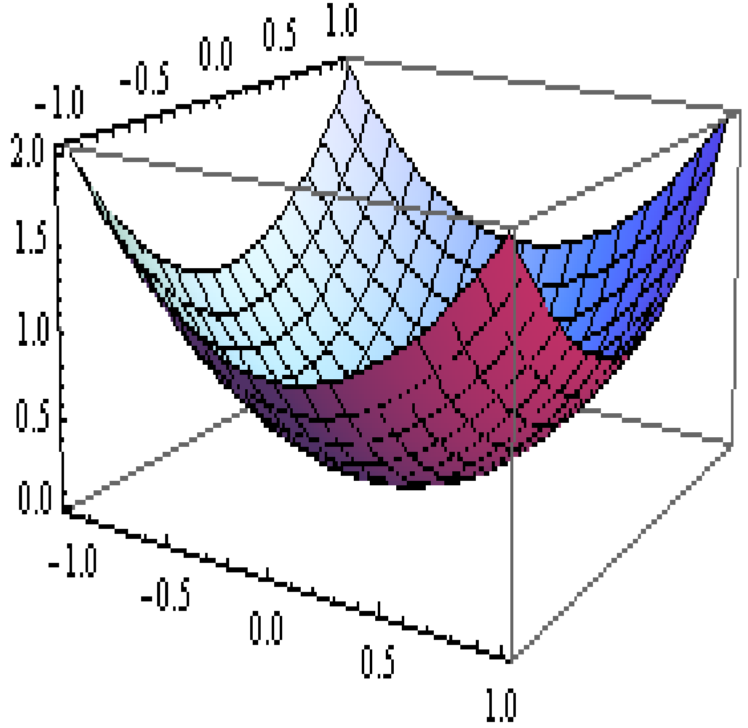


Surface Plots of some two variable functions

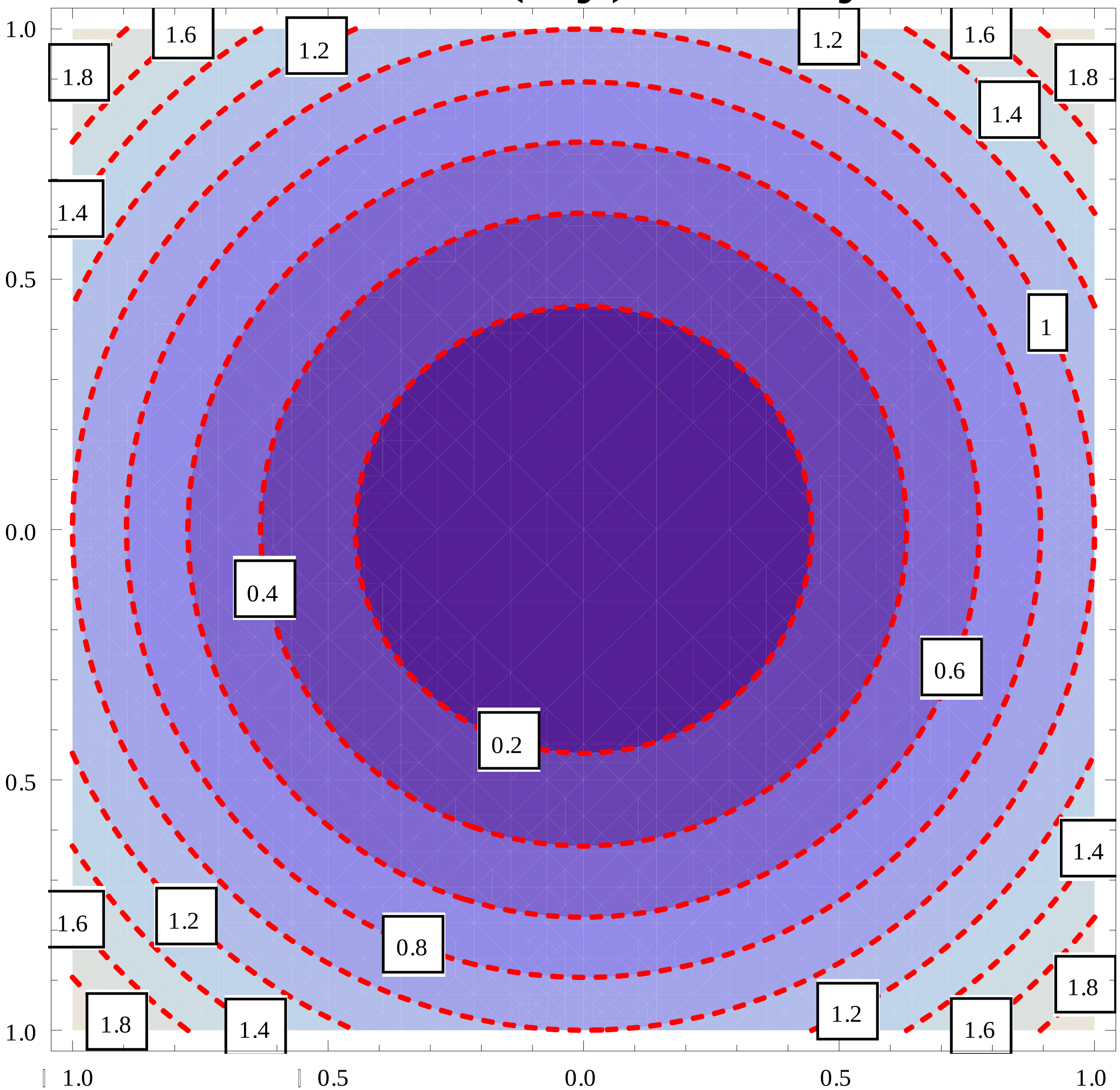
$f(x,y) = x^2 + y^2$

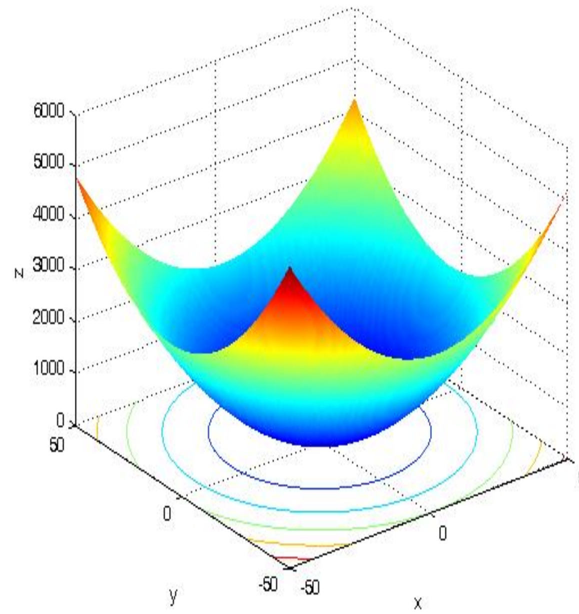
Local minimum and global minimum at $x^*=0$, $y^*=0$

Minimum value = $f(x^*, y^*)=0$



Contour Plot of $f(x,y) = x^2 + y^2$

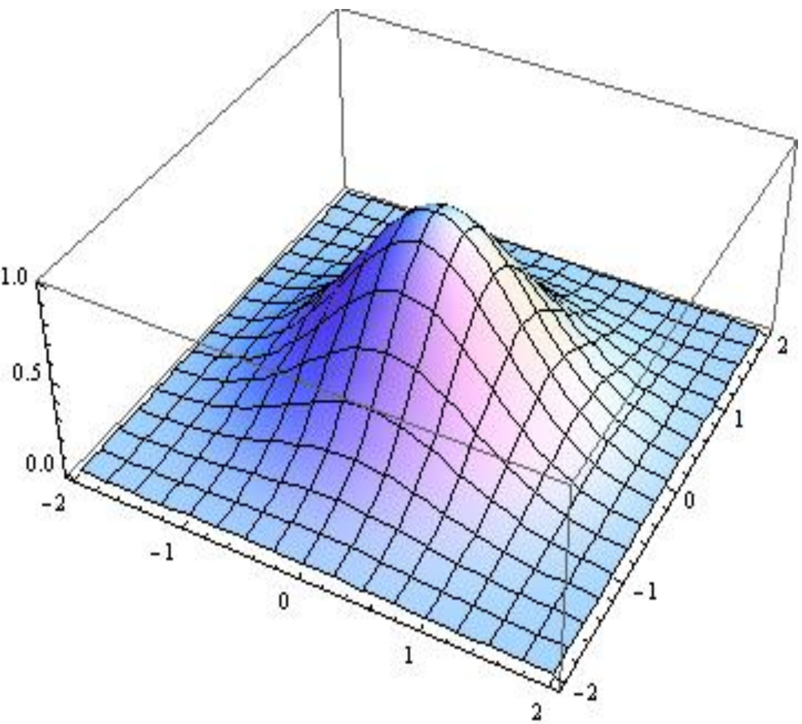


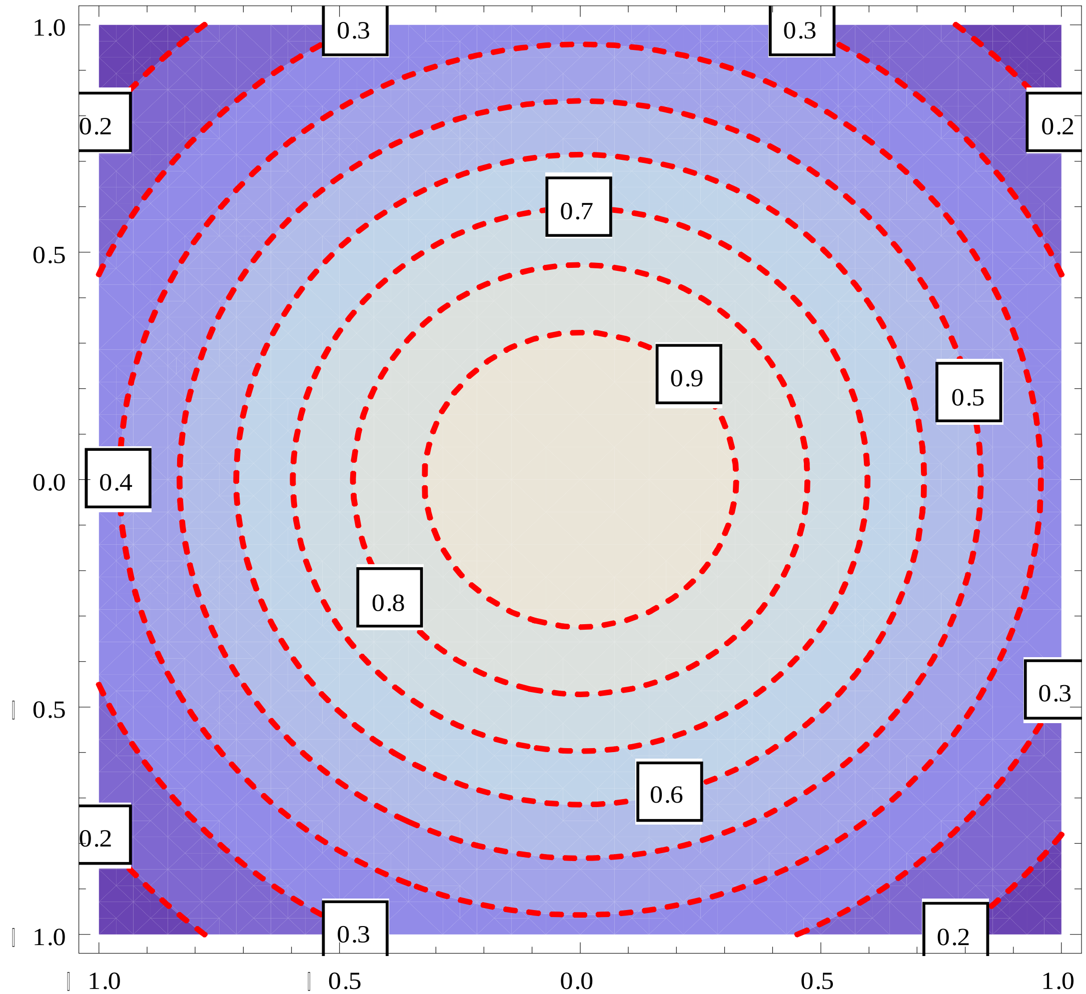


Surface Plots and Contour Plots of f(x,y) =Exp[-($x^2 + y^2$)]

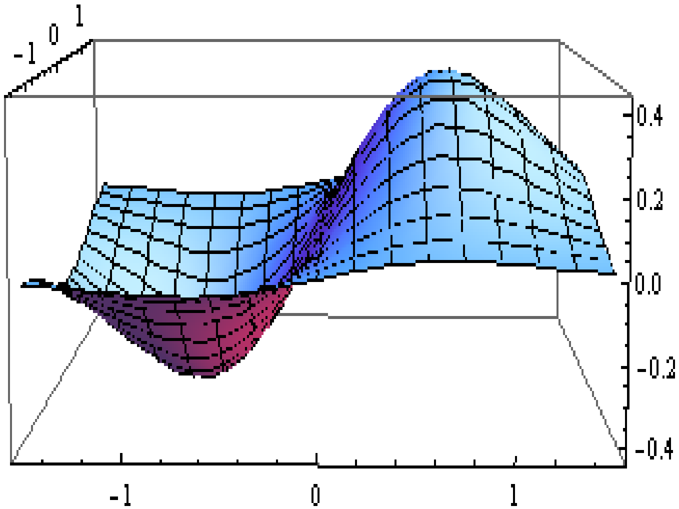
Local maximum and global maximum at $x^*=0$, $y^*=0$

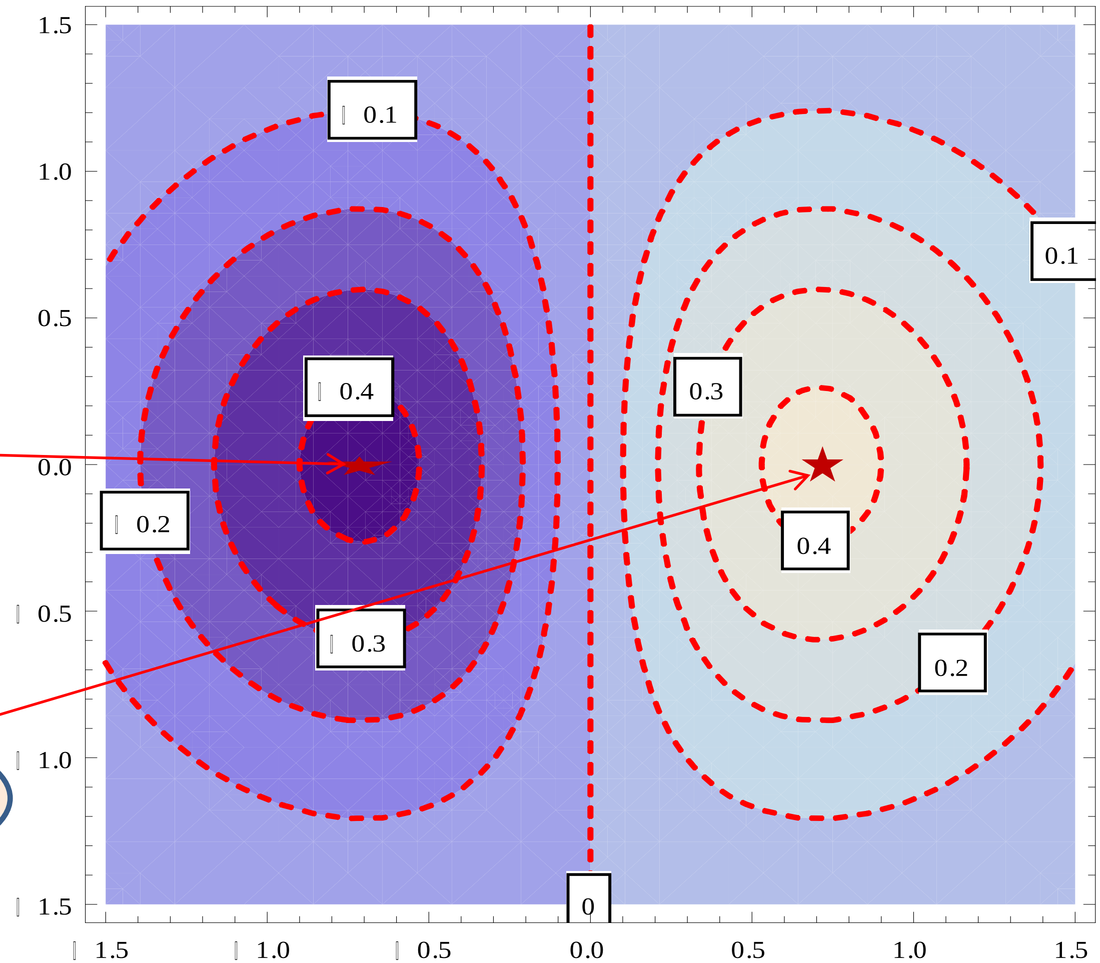
Minimum value = $f(x^*, y^*)=1$





Surface Plot and Contour plot $f(x,y) = x \, \mathrm{Exp}[-(x^2 + y^2)]$





Minimum point

Maximum point

Multivariable Unconstrained Optimization Problems

Minimize $f(x_1, x_2, \ldots, x_n)$

- Graphical method
- Analytical methods

Positive/Negative Definite Matrices

- A symmetric matrix A is defined to be a **positive definite matrix** if for any point x in the search space the quantity $\boldsymbol{x^T A x \geq 0}$.
- A symmetric matrix A is called a **negative definite matrix** if for any point $\boldsymbol{x}$ the quantity $\boldsymbol{x^T A x} \leq 0$.

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^T A\mathbf{x};\ \ A \text{ is symmetric}$$

$$= \begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} ax_1 + bx_2 \\ bx_1 + cx_2 \end{pmatrix}$$

$$= x_1\left(ax_1 + bx_2\right) + x_2\left(bx_1 + cx_2\right)$$

$$= ax_1^2 + 2bx_1x_2 + cx_2^2$$

Positive/Negative Definite Matrices

- How to investigate whether a symmetric matrix is positive definite or negative definite?
  - A symmetric matrix is **positive definite** if all its eigenvalues are positive.
  - A symmetric matrix is **positive definite** if all its principal minors are positive.
  - A matrix A is **negative definite** if –A is positive definite.

Examples of positive definite matrices:

- $A = \begin{bmatrix} 5 & 0 \\ 0 & 6 \end{bmatrix}$ eigenvalues are 5,6 (positive)
- $B = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix}$

First Principal minor is 2

Second Principal minor is $\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3$

Third Principal minor is |B| = 4

Examples of negative definite matrices:

- $C = \begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$
- $D = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & -1 \\ 1 & -1 & -2 \end{bmatrix}$

Eigenvalues: negative

Principal minors : -,+,-,+,-,+,…

- A matrix is **indefinite** if it is neither positive definite or negative definite.
- A matrix is **positive semi-definite** if one of the eigenvalues or one of the principal minors is zero and all others are positive.
- A matrix A is **negative semi-definite** if –A is positive semi-definite.

Two Variable Unconstrained Optimization Problems

A **necessary condition** for a function, f(x,y) with two variables to have a minimum at a point is that, all the first partial derivatives of the function should vanish at that point $\left(i.e.\ \frac{\partial f}{\partial x} = 0; \frac{\partial f}{\partial y} = 0\right)$.

A **sufficient condition** for a function with two variables to have a minimum at a point is that the Hessian matrix of the function at the point is positive definite.

$$\text{Hessian matrix, } H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}.$$

$f'' > 0$

$\rightarrow f'' < 0$

$f'' < 0.$

- If Hessian is negative definite at a stationary point the point would be a local maximum point.
- If the Hessian is indefinite at a stationary point then the point will neither have a minimum nor a maximum at that point. It will be called as a saddle point.
- If the Hessian is positive semi-definite or negative semi-definite at a stationary point then the point may or may not be a minimum/maximum point.

Find the local optimum points for the given functions:

1. $f(x, y) = (x - 3)^2 + (y - 4)^2$

Answer: $\frac{\partial f}{\partial x} = 0, \frac{\partial f}{\partial y} = 0 \rightarrow 2(x - 3) = 0,\ 2(y - 4) = 0$

$\rightarrow$ (x,y)=(3,4) is the stationary point.

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$ is a positive definite matrix.

Hence **(3,4) is a local minimum point**.

2. $f(x_1, x_2) = {x_1}^2 + 2{x_2}^2 - 4x_1 - 2x_1x_2$

$$\frac{\partial f}{\partial x_1} = 0, \frac{\partial f}{\partial x_2} = 0 \rightarrow 2x_1 - 4 - 2x_2 = 0, \qquad 4x_2 - 2x_1 = 0$$

$\rightarrow (x_1, x_2) = (4, 2)$ is the stationary point.

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial {x_2}^2} \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$ is a positive definite matrix.

Hence **(4,2) is a local minimum point**.

Find the local optimum points for the given function:

3. $f(x,y) = x^3 + y^3 + 2x^2 + 4y^2 + 6$

Answer: $\frac{\partial f}{\partial x} = 0 \rightarrow \ 3x^2 + 4x = 0 \ \rightarrow \ x = 0, x = -\frac{4}{3}$

$$\frac{\partial f}{\partial y} = 0 \rightarrow 3y^2 + 8y = 0 \ \rightarrow \ y = 0, y = -\frac{8}{3}$$

The stationary points are $(0,0), \left(0, -\frac{8}{3}\right), \left(-\frac{4}{3}, 0\right), \left(-\frac{4}{3}, -\frac{8}{3}\right)$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 6x+4 & 0 \\ 0 & 6y+8 \end{bmatrix}$$

$H_{(0,0)} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$ is positive definite. So **(0,0) is a local minimum**.

$H_{\left(0,-\frac{8}{3}\right)} = \begin{bmatrix} 4 & 0 \\ 0 & -8 \end{bmatrix}$ is indefinite. So $\left(0, -\frac{8}{3}\right)$ is neither a local minimum nor a local maximum.

$H_{\left(-\frac{4}{3},0\right)} = \begin{bmatrix} -4 & 0 \\ 0 & 8 \end{bmatrix}$ is indefinite. So $\left(-\frac{4}{3}, 0\right)$ is neither a local minimum nor a local maximum.

$\boldsymbol{\left(0, -\frac{8}{3}\right)}$ **and** $\boldsymbol{\left(-\frac{4}{3}, 0\right)}$ **are saddle points.**

$H_{\left(-\frac{4}{3},-\frac{8}{3}\right)} = \begin{bmatrix} -4 & 0 \\ 0 & -8 \end{bmatrix}$ is negative definite. So $\boldsymbol{\left(-\frac{4}{3}, -\frac{8}{3}\right)}$ **is a local maximum**.

n - Variable Unconstrained Optimization Problems

A **necessary condition** for a function with n variables, $f(x_1,x_2,\ldots,x_n)$ to have a minimum at a point is that, all the first partial derivatives of the function should vanish at that point.

A **sufficient condition** for a function with two variables to have a minimum at a point is that the Hessian matrix of the function at the point is positive definite.

$$\text{Hessian matrix, } H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Exercise

1. Find the optimum points and the optimum values for the given functions:

   (a) $f(x_1, x_2) = 0.5{x_1}^2 + 0.5(x_2 - x_1)^2 + 0.5{x_2}^2 - 2x_2$

   (b) $f(x, y) = \frac{1}{3}(x^3 + y^3) - \frac{3}{2}(x^2 + y^2) + 2x$

   (c) $\mathrm{f(x, y) = 9x^2 + 4y^2 - 6x - 8y + 8xy + 5.}$

   (d) $f(x, y, z) = x^2 + 16y^2 + 4z^2 - 2x - 16y - 12z + 9$

   (e) $f(x, y, z) = x^2 + y^2 + z^2 - 4x - 8y - 12z + 59$

2. Find the values of *a* and *b* in the given function *f(x,y)* if the function has a minimum at (2,3) and a maximum at (1,0).

$$f(x, y) = a(x^3 + y^3) + b(x^2 + y^2) + 2x$$

3. Find all values of *k* for which the matrix, $A = \begin{bmatrix} -\frac{1}{\sqrt{2}} & k \\ k & -\frac{1}{\sqrt{2}} \end{bmatrix}$ is negative definite.

4. Find the optimum values for the function $f(x, y) = x^2 + 2y^2 - 4x - 2xy$.

5. Find the value of 'a' for which the matrix $A = \begin{bmatrix} a & 1 & 1 \\ 1 & a & 1 \\ 1 & 1 & a \end{bmatrix}$ is positive definite.

Newton's Method to find a minimum of a function:

Minimize f(x)

In this case Newton's method is applied to find the root of the equation $f'(x)=0$.

Iterative Formula: $x_{n+1} = x_n - f'(x_n)/f''(x_n)$

Find the minimum of $f(x) = x^2 + \frac{54}{x}$ using Newton's Method

(i) Choose $x^{(0)}=2$ and perform 3 iterations manually.

(ii) Choose $x^{(0)}=1$ and perform 10 iterations using excel.

(iii) Choose $x^{(0)}=1$ and perform 100 iterations using MATLAB.

$f'(x) = 2x - \frac{54}{x^2}$; $f''(x) = 2 + \frac{108}{x^3}$.

$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$

$x_{(1)} = x_0 - \frac{f'(x_0)}{f''(x_0)} = 2 - \frac{(-9.5)}{(2+13.5)} =$

AMRITA VISHWA VIDYAPEETHAM UNIVERSITY Established u/s 3 of UGC Act 1956

Find the minimum of $f(x) = x^2 + \frac{54}{x}$ using Newton's Method

(i) Choose $x^{(0)}$=2 and perform 3 iterations manually.

(ii) Choose $x^{(0)}$=1 and perform 10 iterations using excel.

(iii) Choose $x^{(0)}$=1 and perform 100 iterations using MATLAB.

xn	f'(xn)=2*xn-54/xn^2	f''(xn)=2+108/xn^3	xn+1=xn-f'(xn)/f''(xn)
2	-9.5	15.5	2.612903226
2.612903226	-2.683658569	8.054158411	2.946104838
2.946104838	-0.329322696	6.223564818	2.999020282
2.999020282	-0.005880227	6.003921432	2.99999968
2.99999968	-1.92011E-06	6.00000128	3
3	0	6	3
3	0	6	3

Find the minimum of $f(x) = x^2 + \frac{54}{x}$ using Newton's Method

(i) Choose $x^{(0)}=2$ and perform 3 iterations manually.

(ii) Choose $x^{(0)}=1$ and perform 10 iterations using excel.

(iii) Choose $x^{(0)}=50$ and perform 100 iterations using MATLAB.

```
%Find minimum of f(x)=x^2+54/x,
%Iterative formula is x_n+1=x_n-f'(x_n)/f''(x_n)
%f'(x)=2x-54/x^2; f''(x)=2+108/x^3
x0 = 50  % initial point
N=100  % number of iterations
x = x0
xstart=x; % initial starting positions
for k=1:N
    x=x-(2*x-54/x^2)/(2+108/x^3);
end
xend=x;
disp('starting and ending positions')
disp([xstart' xend'])
```

Newton's method can be practically done for all functions by evaluating the first and second derivatives numerically and then apply the formula:

$x_{n+1} = x_n - f(x_n)/f'(x_n)$ to find root of $f(x)=0$ /

$x_{n+1} = x_n - f'(x_n)/f''(x_n)$ to find minimum of $f(x)$ (root of $f'(x)=0$).

- Evaluation of first and second derivatives numerically:

$$f'(a) = \frac{f(a+\Delta a) - f(a-\Delta a)}{2\Delta a};$$

$$f''(a) = \frac{f\big((a+\Delta a)\big) - 2f(a) + f((a-\Delta a))}{(\Delta a)^2};$$

$$\text{where, } \Delta a = \begin{Bmatrix} 0.01|a|, & if \quad a < -0.01 \\ 0.01|a|, & if \quad a > 0.01 \\ 0.0001, & if \quad -0.01 < a \quad < 0.01 \end{Bmatrix}$$

Newton Raphson Method to find a minimum of a function

Algorithm:

Step 1: Choose an initial guess $x^{(0)}$ and a small number ε. Set k=0. Compute $f'(x^{(k)})$.

Step 2: Compute $f''(x^{(k)})$.

Step 3: Calculate $x^{(k+1)} = x^{(k)} - \frac{f'(x^{(k)})}{f''(x^{(k)})}$. Compute $f'(x^{(k+1)})$.

Step 4: If $\left|f'(x^{(k+1)})\right| < \varepsilon$ Terminate;

Else set k=k+1 and go to step 2.

Problem

1. Find the minimum of $f(x) = x^2 + \frac{54}{x}$ using Newton Raphson Method. [Choose $x^{(0)}$=2 and termination parameter as 0.5].

$$f'(x^0) = f'(2) = \frac{f(2+0.02) - f(2-0.02)}{2(0.02)} = -9.50135$$

$$f''(x^0) = f''(2) = \frac{f(2+0.02) - 2f(2) + f(2-0.02)}{(0.02)^2} = 15.50135$$

$$x^{(1)} = x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = 2 - \frac{(-9.50125)}{15.475} = 2.612937$$

$$f'(x^{(1)}) = \frac{27.42387 - 27.56376}{2(0.0261397)} = -2.68418$$

$$\left| f'(x^{(1)}) \right| > \varepsilon$$

Iteration 2:-

$$f''(x^{(1)}) = f''(2.61397)$$
$$= \frac{f(2.612937 + 0.02612937) - 2f(2.612937) + f(2.612937 - 0.02612937)}{(0.02612937)^2}$$
$$= 8.05453$$

$$x^{(2)} = x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = 2.946188$$

$$f'(x^{(2)}) = -0.32943$$

$$\left| f'(x^{(2)}) \right| < \varepsilon$$

Minimum of *f(x)* is at $x^{(2)} = 2.946188$

Problem

1. Find the minimum of $f(x) = x^2 + \frac{54}{x}$ using Newton Raphson Method. [Choose $x^{(0)}$=2 and termination parameter as 0.00001] using excel.

xk	f(xk)	del(xk)	xk+del(xk)	f(xk+del(xk))	xk-del(xk)	f(xk-del(xk))	f'(xk)	f''(xk)	x(k+1)=xk-f'(xk)/f''(xk
2	31	0.02	2.02	30.813073	1.98	31.193127	-9.50135	15.5014	2.61294
2.61294	27.4938	0.02613	2.63907	27.426453	2.58681	27.566725	-2.684178	8.05453	2.94619
2.94619	27.0088	0.02946	2.97565	27.001789	2.91673	27.0212	-0.329429	6.22363	2.99912
2.99912	27	0.02999	3.02911	27.002526	2.96913	27.002879	-0.005884	6.00392	3.0001
3.0001	27	0.03	3.0301	27.0027	2.9701	27.0027	-1.88E-06	6	3.0001

f'(3.0001)<0.00001 So minimum at 3.0001

Problem

2. Find a minimum of the function: f(x) $= 0.5x^3 - 1.75x^2 + 2x - 9$ using Newton Raphson method starting with the initial point x=2 and with the termination parameter 0.3 in excel.

Solution:

f'(2)=1.0002

f''(2)=2.5

$x^{(1)}$=1.59992

f'(1.59992)=0.240087 < ε

So minimum at 1.59992

Numerical Algorithms for Multivariable Unconstrained Problems

Minimize $f(x_1, x_2, \ldots x_n)$ $\longrightarrow$ analytically, $\nabla f = 0$, $H \succ 0$
$\longrightarrow$ Newton's method

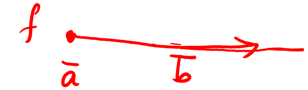
- Unidirectional Search / Line Search
- Gradient Descent Algorithm / Steepest Descent Algorithm
- Newton's Method ✓

uses gradient → 1st order algorithms

gradient, hessian
2nd order algorithms

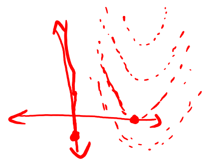
Unidirectional Search / Line search

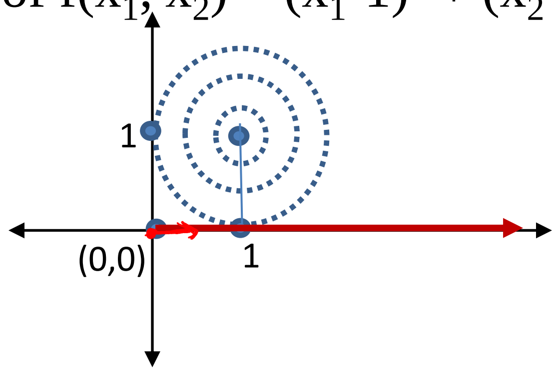
- Unidirectional Search is the search for a minimum from a point with position vector $\boldsymbol{\bar{a}}$ along a specified direction $\boldsymbol{\bar{b}}$.



  – For Example:

  Find the minimum of $f(x_1, x_2) = (x_1-1)^2 + (x_2-1)^2$ along the direction of X axis from the origin.

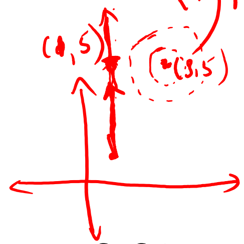




$\bar{a} = (0,0)$ $\bar{b} = (1,0)$

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

Qn) Min. $(x-3)^2 + (y-5)^2$ starting $(1,1)$ +ve Y dirn.



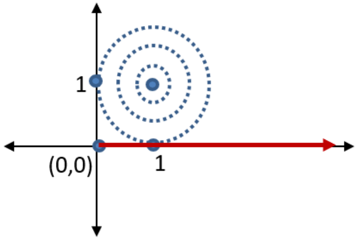
Min. on the given line is at $(1,5)$.

Geometrically: Find the minimum of $f(x_1, x_2)$ on the X axis.

$\therefore$Min is at (1,0)

Unidirectional Search / Line search

Find the minimum of $f(x_1, x_2) = (x_1-1)^2 + (x_2-1)^2$ along the direction of X axis from the origin.



How to perform unidirectional search?

Step 1: Write the parametric representation of the line of search

$$\bar{\boldsymbol{s}}\,(t) = \bar{\boldsymbol{a}} + t\bar{\boldsymbol{b}}$$

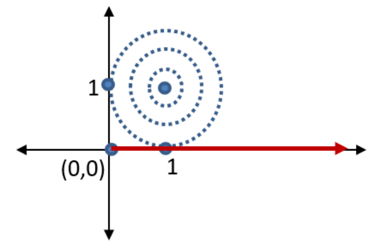
Step 2: Write the function in terms of t, $f(\bar{\boldsymbol{s}}\,(t))$. This is a single variable function.

Step 3: Find the minimum of $f(\bar{\boldsymbol{s}}\,(t))$, t^*.

Step 4: Obtain the solution for unidirectional search by substituting t^* in $\bar{\boldsymbol{s}}\,(t)$.

Unidirectional Search / Line search

Find the minimum of $f(x_1, x_2) = (x_1-1)^2 + (x_2-1)^2$ along the direction of X axis from the origin.



How to perform unidirectional search?

Step 1: Write the parametric representation of the line of search

$$\bar{\boldsymbol{s}}\,(t) = \bar{\boldsymbol{a}} + t\bar{\boldsymbol{b}}$$

Step 2: Write the function in terms of t, $f(\bar{\boldsymbol{s}}\,(t))$. This is a single variable function.

Step 3: Find the minimum of $f(\bar{\boldsymbol{s}}\,(t))$, t^*.

Step 4: Obtain the solution for unidirectional search by substituting t^* in $\bar{\boldsymbol{s}}\,(t)$.

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

Step 1: $\bar{\boldsymbol{s}}\,(t) = \bar{\boldsymbol{a}} + t\bar{\boldsymbol{b}} = (\overset{\bar{a}}{0,0}) + t(\overset{\bar{b}}{1,0}) = (t, 0)$ ✓ $x_1 = t$, $x_2 = 0$

Step 2: $f\big(\bar{\boldsymbol{s}}\,(t)\big) = (t-1)^2 + (0-1)^2 = t^2 - 2t + 2$ ✓

Step 3: $f'\big(\bar{\boldsymbol{s}}\,(t)\big) = 0 \rightarrow 2t - 2 = 0 \rightarrow t = 1$ is the critical point.

Also $f''\big(\bar{\boldsymbol{s}}\,(t)\big) > 0$, So $t^* = 1$ is the minimum of $f\big(\bar{\boldsymbol{s}}\,(t)\big)$.

Step 4: $\bar{\boldsymbol{s}}\,(t^*) = (1,0)$ is the minimum of $f(x_1, x_2) = (x_1-1)^2 + (x_2-1)^2$ along the direction of X axis from the origin.

Unidirectional Search / Line search

Find the minimum of f(x_1, x_2) = along the direction of X axis from the origin.

How to perform unidirectional search?

Step 1: Write the parametric representation of the line of search

$$\bar{\boldsymbol{s}}\,(t) = \bar{\boldsymbol{a}} + t\bar{\boldsymbol{b}}$$

Step 2: Write the function in terms of t, $f\,(\bar{\boldsymbol{s}}\,(t))$. This is a single variable function.

Step 3: Find the minimum of $f\,(\bar{\boldsymbol{s}}\,(t))$, t^*.

Step 4: Obtain the solution for unidirectional search by substituting t^* in $\bar{\boldsymbol{s}}\,(t)$.

Q) $f(x,y) = (x-2)^2 - y$, $\bar{a} = (-1, 0)$ $\bar{b} = (1,1)$

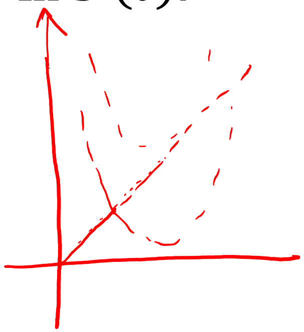
Step 1:- $\bar{s}(t) = \bar{a} + t\bar{b} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} + t\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} t-1 \\ t \end{pmatrix}$

Step 2:- $f(\bar{s}(t)) = (t-1-2)^2 - t = t^2 - 6t + 9 - t = t^2 - 7t + 9$

Step 3:- $f'(s(t)) = 0 \rightarrow 2t - 7 = 0 \Rightarrow t^* = 7/2$.

$f''(t) = 2 > 0 \Rightarrow t^* = 7/2$ has min. of $f(t)$.

Step 4:- $\bar{s}(t^*) = \begin{pmatrix} t^*-1 \\ t^* \end{pmatrix} = \begin{pmatrix} 5/2 \\ 7/2 \end{pmatrix} = \begin{pmatrix} 2.5 \\ 3.5 \end{pmatrix}$ is min. of f.



Unidirectional Search / Line search

1. Find the minimum of the given function $f(x,y,z,w)$ from $(1,1,1,1)$ along the direction $(-1,0,1,1)$ using unidirectional search.

$$f(x,y,z,w) = x^2 + y^2 + z^2 + w^2 - 2x - 4y + 6z + 8w + 15$$

$$\bar{a} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \bar{b} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad \bar{s} = \bar{a} + t\bar{b} = \begin{pmatrix} 1-t \\ 1 \\ 1+t \\ 1+t \end{pmatrix}$$

$$f(\bar{s}(t)) = (1-t)^2 + 1 + (1+t)^2 + (1+t)^2 - 2(1-t) - 4 + 6(1+t) + 8(1+t)$$

$$f'(t) = 0 \Rightarrow -2(1-t) + 2(1+t) + 2(1+t) + 2 + 6 + 8 = 0 \Rightarrow 6t = -18 \Rightarrow t = -3$$

$f''(t) = 6 > 0 \Rightarrow t^* = -3$ is the minimum of $f(s(t))$.

$\bar{s}(t^*) = \begin{pmatrix} 1-t^* \\ 1 \\ 1+t^* \\ 1+t^* \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ -2 \\ -2 \end{pmatrix}$ is the minimum of f along $(-1,0,1,1)$ from $(1,1,1,1)$

Unidirectional Search / Line search

2. Find the minimum of the function $f(x, y, z) = x^3 + y^2 - 6x - 8y + 10z + 2$ from the point (0,0,0) along the direction (-1,1,1) using unidirectional search method.

Line of Search :- $\bar{s}(t) = \bar{a} + t\bar{b} = \begin{pmatrix} -t \\ t \\ t \end{pmatrix}$

function $f(t)$:- $f(t) = -t^3 + t^2 + 6t - 8t + 10t + 2$

Min. of $f(t)$:- $f'(t) = 0 \Rightarrow -3t^2 + 2t + 6 - 8 + 10 = 0$

$\Rightarrow -3t^2 + 2t + 8 = 0$

$\Rightarrow t = 2, -4/3.$ (Critical pts.)

$f''(t) = -6t + 2$

$f''(2) = -10 < 0 \Rightarrow t = 2$ is a max. pt.

$f''(-\frac{4}{3}) = 10 > 0 \Rightarrow t^* = \frac{-4}{3}$ is a min. pt.

Min. of $f(x, y, z)$:- $\bar{s}(t^*) = \begin{pmatrix} 4/3 \\ -4/3 \\ -4/3 \end{pmatrix}$ //

& Max. of $f(x, y, z)$ is at $\begin{pmatrix} -2 \\ 2 \\ 2 \end{pmatrix}$

Descent direction

- A direction given by the vector **d** from $\boldsymbol{x}^{(k)}$ is a descent direction only if the function value decreases along that direction from the point, $\boldsymbol{x}^{(k)}$ i.e., when

$$\nabla f\left(\boldsymbol{x}^{(k)}\right) \cdot \boldsymbol{d} < \mathbf{0}$$

- A direction given by the vector **d** from $\boldsymbol{x}^{(k)}$ is an ascent direction only if the function value increases along that direction from the point, $\boldsymbol{x}^{(k)}$ i.e., when

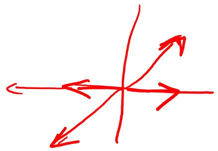
$$\nabla f\left(\boldsymbol{x}^{(k)}\right) \cdot \boldsymbol{d} > \mathbf{0}$$

Eg: f(x,y) = x+y has a descent direction along the direction (-1,0), (-1,-1), but along the direction (1,0), (1,1) the function value increases.

$\nabla f = (1,1)$;

$\nabla f \cdot (-1,0) < 0$, $\nabla f \cdot (-1,-1) < 0$,

But $\nabla f \cdot (1,0) > 0$ and $\nabla f \cdot (1,1) > 0$



Exercise

1. Check whether the given function have a descent direction from **x=** (2,-1) along the given directions $\mathbf{d_1}$ and $\mathbf{d_2}$

$$f = 2{x_1}^2 + {x_2}^2 - 2x_1x_2 + 2{x_1}^3 + {x_2}^4$$

$\mathbf{d_1}$= (-2, 3), $\mathbf{d_2}$= (1, 1)

Answer:

$$\nabla f = (4x_1 - 2x_2 + 6{x_1}^2, 2x_2 - 2x_1 + 4{x_2}^3)$$
$$\nabla f(2, -1) = (34, -10)$$

$\nabla f \cdot d_1 = (34, -10) \cdot (-2,3) = -98 < 0$. So $\mathbf{d_1}$ is a descent direction.

$\nabla f \cdot d_2 = (34, -10) \cdot (1,1) = 24 > 0$. So $\mathbf{d_2}$ is not a descent direction.

2. Check whether the given function have a descent direction from **x=** (4, 2,-1) along the given directions $\mathbf{d_1}$ and $\mathbf{d_2}$

$$f = (x_1 - 1)^4 + (x_2 - 3)^2 + 4(x_3 + 5)^4$$

$\mathbf{d_1}$= (-1,10,-1), $\mathbf{d_2}$= (-1, 2, 1)

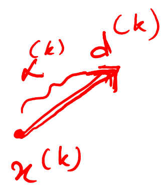
3. Consider the $f(x, y) = sin\, x + 4x^2 + y^3 - 3y + 2$

Which of the following directions are descent directions for the function from (0,0)?

(i) (1,1) (ii) (2,-1)

Direction of Descent/Ascent in numerical algorithms

- In any numerical algorithm to find the optimum of an unconstrained problem, the iterative formula used is:



$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)}$$

step length.

where $\alpha^{(k)}$ is the step length in step k and
$d^{(k)}$ is the direction of descent in step k, if it's a minimization problem or a direction of ascent if it's a maximization problem.

Newton's:- $\bar{x}_{k+1} = \bar{x}_k - (H(x_k))^{-1} \nabla f(x_k) \rightarrow d^{(k)} = -H(x_k)^{-1} \nabla f(x_k)$

$\alpha^{(k)} = 1$

Modified Newton's Alg:- $\bar{x}_{k+1} = \bar{x}_k + \alpha_k \bar{d}_k$, $d_k = -H(x_k)^{-1} \nabla f(x_k)$

α = using ~~da~~ line search.

Gradient Descent | Steepest Desc. Alg:- $x_{k+1} = x_k + \alpha_k d_k$, $d_k = -\nabla f(x_k)$

α = using line search

Steepest Descent Algorithm to minimize a multivariable function

Steepest Descent Algorithm

- The iterative formula for method of steepest descent is
$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} d^{(k)}$$
- The descent direction $d^{(k)}$ is along the steepest descent direction, $d^{(k)} = -\nabla f\left(x^{(k)}\right)$
- The step length $\alpha^{(k)}$ is obtained by performing a unidirectional search from $x^{(k)}$ along the direction $d^{(k)}$ i.e., by minimizing,
$$\varphi\left(\alpha^{(k)}\right) = f\left(x^{(k)} + \alpha^{(k)} d^{(k)}\right)$$
$$\left(\varphi'\left(\alpha^{(k)}\right) = 0, \;\; \varphi''\left(\alpha^{(k)}\right) > 0 \right)$$

<u>Algorithm:</u>

Step 1: Chose an initial starting point $x^{(0)}$ and a termination parameter ε.

Step 2: Compute steepest descent, $d^{(0)} = -\nabla f\left(x^{(0)}\right)$

Step 3: Compute the step length, $\alpha^{(0)}$[using unidirectional search from $x^{(0)}$ along $d^{(0)}$].

Step 4: Evaluate $x^{(1)} = x^{(0)} + \alpha^{(0)} d^{(0)}$

Step 5: Compute $\left\|\nabla f\left(x^{(1)}\right)\right\|$. If $\left\|\nabla f\left(x^{(1)}\right)\right\| < \varepsilon$, stop and mention $x^{(1)}$ is minimum,

Else go to step 2 with $x^{(0)} = x^{(1)}$.

[While manually solving these problems gradients in step 2 can be directly found. But while coding the numerical formula to evaluate gradients need to be used.]

Exercise

1. Find a minimum for the function, $f(x,y) = (x-1)^2 + (y-2)^2$ starting from the point (10,-1), using steepest descent method. Choose the termination parameter $\varepsilon = 0.1$

$\bar{x}^{(0)} = (10, -1)$; $\nabla f = \begin{pmatrix} 2x-2 \\ 2y-4 \end{pmatrix}$; $\nabla f(10,-1) = \begin{pmatrix} 18 \\ -6 \end{pmatrix}$

Dirn. of descent in 1st iteration $\Big\}$ $d^{(0)} = -\nabla f(x^{(0)}) = \begin{pmatrix} -18 \\ 6 \end{pmatrix}$

Line search from $x^{(0)}$ along $d^{(0)}$:- $\bar{s}(\alpha) = x^{(0)} + \alpha d^{(0)} = \begin{pmatrix} 10 \\ -1 \end{pmatrix} + \alpha \begin{pmatrix} -18 \\ 6 \end{pmatrix} = \begin{pmatrix} 10-18\alpha \\ -1+6\alpha \end{pmatrix}$

function :- $f(\bar{s}(\alpha)) = (10-18\alpha-1)^2 + (-1+6\alpha-2)^2$
$= (9-18\alpha)^2 + (6\alpha-3)^2$

Min. $f(\alpha)$:- $f'(s(\alpha)) = 0 \implies 2(9-18\alpha)(-18) + 2(6\alpha-3)(6) = 0$
$\implies -36(9-18\alpha) + 12(6\alpha-3) = 0$
$\implies (18\times 3 + 6)\alpha = 27 + 3$
$60\alpha = 30 \implies \boxed{\alpha = 1/2}$

$\therefore$ Step length $f''(s(\alpha)) = 60 > 0 \implies \alpha^* = 1/2$ is a minimum.

$\therefore \bar{x}_1 = \bar{x}_0 + \alpha \bar{d}_0$ or $\bar{s}(\alpha^*) = \begin{pmatrix} 10-18\times 1/2 \\ -1+6\times 1/2 \end{pmatrix}$
$= \begin{pmatrix} 10 \\ -1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -18 \\ 6 \end{pmatrix}$
$\bar{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\nabla f(x^{(1)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\|\nabla f(x^{(0)})\| = 0 < \varepsilon$
$\implies (1,2)$ is the minimum.

Exercise

2. Apply three iterations of steepest descent method to find a minimum for the function, $f(x,y) = 2x^2 - 2xy + y^2$ starting from the point (1,2).

Itn. 1 $\bar{x}^{(0)} = (1,2)$; $\nabla f = \begin{pmatrix} 4x-2y \\ -2x+2y \end{pmatrix}$; $\nabla f(1,2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$; $d^{(0)} = -\nabla f(1,2) = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$;

Unidirectional search from $x^{(0)}$ along $d^{(0)}$:- $s(\alpha) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2-2\alpha \end{pmatrix}$;

$f(s(\alpha)) = 2 - 2(2-2\alpha) + \underbrace{(2-2\alpha)^2}_{(4+4\alpha^2-8\alpha)}$; $f'(\alpha) = 0 \Rightarrow 4 + 8\alpha - 8 = 0 \Rightarrow \boxed{\alpha = 1/2}$

$f''(\alpha) = 8 > 0 \Rightarrow \alpha^* = 1/2$ is minimum step length.

Itn. 2 $\therefore \bar{x}^{(1)} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ //

$\nabla f(1,1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$; $d^{(1)} = -\nabla f(1,1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$

Unidirectional search from $x^{(1)}$ along $d^{(1)}$:- $s(\alpha) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1-2\alpha \\ 1 \end{pmatrix}$

$f(s(\alpha)) = \underbrace{2(1-2\alpha)^2}_{2(1-4\alpha+4\alpha^2)} - 2(1-2\alpha) + 1$; $f'(\alpha) = 0 \Rightarrow -8 + 16\alpha + 4 = 0$

$\Rightarrow \boxed{\alpha = 1/4}$

$f''(\alpha) = 16 > 0 \Rightarrow \alpha = 1/4$ is the required step length

Itn. 3 $\therefore \bar{x}^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$ //

$\nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$; $d^{(2)} = -\nabla f(0.5, 1) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

Unidirectional search from $x^{(2)}$ along $d^{(2)}$:- $s(\alpha) = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1-\alpha \end{pmatrix}$

$f(s(\alpha)) = 0.5 - 2(0.5)(1-\alpha) + (1-\alpha)^2$; $f'(\alpha) = 0 \Rightarrow 1 - 2 + 2\alpha = 0 \Rightarrow \boxed{\alpha = 1/2}$

$f''(\alpha) = 2 > 0 \Rightarrow$ Step length

$\therefore x^{(3)} = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ //

Exercise

2. Apply three iterations of steepest descent method to find a minimum for the function, $f(x,y) = 2x^2 - 2xy + y^2$ starting from the point (1,2).

Solving analytically:- $\nabla f = 0$; $\left.\begin{array}{l} 4x - 2y = 0 \\ -2x + 2y = 0 \end{array}\right\}$ $(x,y) = \underbrace{(0,0)}_{\text{st. pt.}}$

$$H = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \quad \text{+ve def.}$$

$\therefore$ Min. is $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

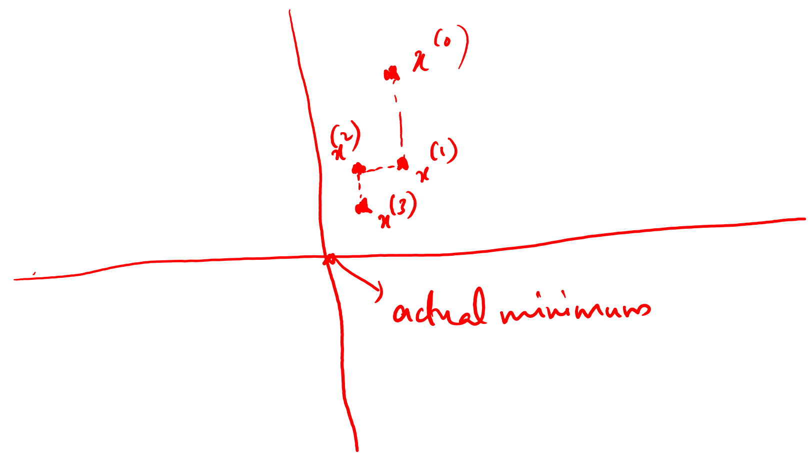
By S.D.A :-

$x^{(0)} = (1,2)$

$x^{(1)} = (1,1)$

$x^{(2)} = (0.5, 1)$

$x^{(3)} = (0.5, 0.5)$



Exercise

3. Apply 4 iterations of steepest descent method to find the minimum of the function, $f(x_1, x_2) = 3x_1^2 - 4\, x_1\, x_2 + 2x_2^2 + 4\, x_1 + 6$ starting from the origin. [Evaluate the gradients in each iteration analytically].

Ans:

k	$X^{(k)}$	$\nabla f(X^{(k)})$	$d^{(k)}$	α_k
1	(0,0)	(4,0)	(-4,0)	1/6
2	(-2/3,0)	(0,8/3)	(0,-8/3)	1/4
3	(-2/3,-2/3)	(8/3,0)	(-8/3,0)	1/6
4	(-10/9,-2/3)	(0,16/9)	(0,-16/9)	1/4

Steepest Descent method

- This method is also called as the gradient descent method as it uses the negative gradient as the search direction in each iteration.
- **The method produces successive directions that are perpendicular to each other.**
- When the point is away from the optimum, the method makes good progress towards the optimum.
- Near the optimum due to zigzagging, the convergence becomes very slow.

Calculation of gradient and Hessian numerically

In all the numerical algorithms, the calculation of gradient and Hessian is involved at different points and in many cases it may not be easy to calculate these analytically. Hence, numerically these are evaluated using finite difference formulae:

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i}\Big|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_i{}^{(\boldsymbol{t})} + \Delta x_i{}^{(\boldsymbol{t})}\right) - f\left(x_i{}^{(\boldsymbol{t})} - \Delta x_i{}^{(\boldsymbol{t})}\right)}{2\Delta x_i{}^{(t)}}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i{}^2}\Big|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_i{}^{(\boldsymbol{t})} + \Delta x_i{}^{(\boldsymbol{t})}\right) - 2f\left(x_i{}^{(\boldsymbol{t})}\right) + f\left(x_i{}^{(\boldsymbol{t})} - \Delta x_i{}^{(\boldsymbol{t})}\right)}{[\Delta x_i{}^{(t)}]^2}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i x_j}\Big|_{\boldsymbol{x}^{(t)}} = [f\left(x_i{}^{(\boldsymbol{t})} + \Delta x_i{}^{(\boldsymbol{t})}, x_j{}^{(\boldsymbol{t})} + \Delta x_j{}^{(\boldsymbol{t})}\right) - f\left(x_i{}^{(\boldsymbol{t})} + \Delta x_i{}^{(\boldsymbol{t})}, x_j{}^{(\boldsymbol{t})} - \Delta x_j{}^{(\boldsymbol{t})}\right) - f\left(x_i{}^{(\boldsymbol{t})} - \Delta x_i{}^{(\boldsymbol{t})}, x_j{}^{(\boldsymbol{t})} + \Delta x_j{}^{(\boldsymbol{t})}\right) + f\left(x_i{}^{(\boldsymbol{t})} - \Delta x_i{}^{(\boldsymbol{t})}, x_j{}^{(\boldsymbol{t})} - \Delta x_j{}^{(\boldsymbol{t})}\right)]/\ 4\Delta x_i{}^{(\boldsymbol{t})}\Delta x_j{}^{(\boldsymbol{t})}$$

Calculation of gradient and Hessian numerically

$$f'(a) = \frac{f(a+\Delta a) - f(a-\Delta a)}{2\Delta a}$$

$$f''(a) = \frac{f(a+\Delta a) - 2f(a) + f(a-\Delta a)}{(\Delta a)^2}$$

$$\Delta a = \begin{cases} 0.01\,|a|, & a > 0.01 \text{ or } a < -0.01 \\ 0.0001, & -0.01 \le a \le 0.01 \end{cases}$$

$$\frac{\partial f}{\partial x}(a,b) = \frac{f(a+\Delta a, b) - f(a-\Delta a, b)}{2\Delta a}; \quad \frac{\partial f}{\partial y}(a,b) = \frac{f(a, b+\Delta b) - f(a, b-\Delta b)}{2\Delta b}$$

$$\frac{\partial^2 f}{\partial x^2}(a,b) = \frac{f(a+\Delta a, b) - 2f(a,b) + f(a-\Delta a, b)}{(\Delta a)^2}; \quad \frac{\partial^2 f}{\partial y^2}(a,b) = f(a, b+\Delta b)$$

$$\frac{\partial^2 f}{\partial x \partial y}(a,b) = \frac{f(a+\Delta a, b+\Delta b) - f(a+\Delta a, b-\Delta b) - f(a-\Delta a, b+\Delta b) + f(a-\Delta a, b-\Delta b)}{4\Delta a\,\Delta b}$$

Calculation of gradient and Hessian numerically

If f is a function with two variables x_1 and x_2, then

$$\frac{\partial f(\boldsymbol{x})}{\partial x_1}|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_1{}^{(\boldsymbol{t})} + \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})}\right) - f(x_1{}^{(\boldsymbol{t})} - \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})})}{2\Delta x_1{}^{(t)}}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_2}|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} + \Delta x_2{}^{(\boldsymbol{t})}\right) - f(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} - \Delta x_2{}^{(\boldsymbol{t})})}{2\Delta x_2{}^{(t)}}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_1{}^2}|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_1{}^{(\boldsymbol{t})} + \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})}\right) - 2f\left(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})}\right) + f\left(x_1{}^{(\boldsymbol{t})} - \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})}\right)}{[\Delta x_1{}^{(t)}]^2}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_2{}^2}|_{\boldsymbol{x}^{(t)}} = \frac{f\left(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} + \Delta x_2{}^{(\boldsymbol{t})}\right) - 2f\left(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})}\right) + f\left(x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} - \Delta x_1{}^{(\boldsymbol{t})}\right)}{[\Delta x_2{}^{(t)}]^2}$$

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 x_2}|_{\boldsymbol{x}^{(t)}} =$$
$$\left[f\left(x_1{}^{(\boldsymbol{t})} + \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} + \Delta x_2{}^{(\boldsymbol{t})}\right) - f\left(x_1{}^{(\boldsymbol{t})} + \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} - \Delta x_2{}^{(\boldsymbol{t})}\right)\right.$$
$$\left. - f\left(x_1{}^{(\boldsymbol{t})} - \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} + \Delta x_2{}^{(\boldsymbol{t})}\right) + f\left(x_1{}^{(\boldsymbol{t})} - \Delta x_1{}^{(\boldsymbol{t})}, x_2{}^{(\boldsymbol{t})} - \Delta x_2{}^{(\boldsymbol{t})}\right)\right] / \mathbf{4}\Delta x_1{}^{(t)} \Delta x_2{}^{(t)}$$

Exercise

1. Evaluate the gradient and Hessian matrix for the function, $f(x,y) = x^2 - 6x + y^2 - 16y + 10$ computationally at (-2,1).

$$\frac{\partial f}{\partial x}\Big|_{(-\mathbf{2},\mathbf{1})} = \frac{f(-2+0.02,1)-f(-2-0.02,1)}{2(0.02)} = \frac{10.8004-11.2004}{0.04} = -10$$

$$\frac{\partial f}{\partial y}\Big|_{(-\mathbf{2},\mathbf{1})} = \frac{f(-2,1+0.01)-f(-2,1-0.01)}{2(0.01)} = \frac{10.8601-11.1401}{0.02} = -14$$

$$\text{Gradient} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (-10, -14)$$

$$\frac{\partial^2 f}{\partial x^2}\Big|_{(-\mathbf{2},\mathbf{1})} = \frac{f(-2+0.02,1)-2f(-2,1)+f(-2-0.02,1)}{[0.02]^2} = \frac{10.8004-2(11)+11.2004}{0.0004} = 2$$

$$\frac{\partial^2 f}{\partial y^2}\Big|_{(-\mathbf{2},\mathbf{1})} = \frac{f(-2,1+0.01)-2f(-2,1)+f(-2,1-0.01)}{[0.01]^2} = \frac{10.8601-2(11)+11.1401}{0.0001} = 2$$

$$\frac{\partial^2 f}{\partial x \partial y}\Big|_{(-\mathbf{2},\mathbf{1})} = \frac{f(-2+0.02,1+0.01)-f(-2+0.02,1-0.01)-f(-2-0.02,1+0.01)+f(-2-0.02,1-0.01)}{4(0.02)(0.01)}$$

$$= \frac{10.6605-10.9405-11.0605+11.3405}{0.0008} = 0$$

$$\text{Hessian} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$f(\bar{x})$, $f(\bar{x}) \pm a$

Use of Taylor series in finding the turning points/ optimum points

Let x^* be an opt. pt.

$$f(x) \approx f(x^*) + (x - x^*)^T \underbrace{\nabla f(x^*)}_{0} + \boxed{\frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*)}$$

$f(x_1, \ldots x_n)$
$\nabla f = 0$
$\longrightarrow$ St. pts, a, b
$H_f(a) \succ 0$, a is min
$H_f(b) \prec 0$, b is max.

At the turning point/ optimum, if we approximate the function up to second order, gradient vector $\nabla f(x^*)$ is a zero vector. So we only have first and third term. The first term is a constant. The third term is a pure quadratic function. The quadratic term involving hessian matrix take a shape according to the curvature of function at the turning point. If it is a minimum point, the parabola will be 'upward open' (or convex). If it is a maximum point, the parabola will be 'downward open' (or concave). This in turn is indicated by the 'definiteness' of the hessian matrix evaluated at the turning point.

The turning point x^* correspond to a minimum point if $\nabla^2 f(x^*)$ is positive definite. x^*

Mathematically, at x^* function is minimum if $\nabla^2 f(x^*) \succ 0$

The turning point x^* correspond to a maximum point if $\nabla^2 f(x^*)$ is negative definite.

Mathematically, at x^* function maximum if $\nabla^2 f(x^*) \prec 0$ x^*

The turning point x^* correspond to a saddle point if $\nabla^2 f(x^*)$ is indefinite.

AMRITA VISHWA VIDYAPEETHAM UNIVERSITY Established u/s 3 of UGC Act 1956

Use of Taylor series in Newton's method to find optimum points

To move towards optimal point, the iterative formula used in Newton's method is:

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

$x_{k+1} = x_k - (J)^{-1} \bar{f}(x_k) \to$ Root
$x_{k+1} = x_k - (H)^{-1} \nabla f(x_k) \to$ min

Derivation from Taylor's expansion

$$f_{app}(x) = f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x}) + \frac{1}{2}(x - \tilde{x})^T \nabla^2 f(\tilde{x})(x - \tilde{x}) + \ldots.$$

$$\nabla f_{app}(x) = \nabla f(\tilde{x}) \quad + \frac{1}{2} \times 2 \times \nabla^2 f(\tilde{x})(x - \tilde{x})$$

$$\nabla f_{app}(x) = 0 \Rightarrow (x - \tilde{x}) = -(\nabla^2 f(\tilde{x}))^{-1} \nabla f(\tilde{x})$$

$$x_{new} = \tilde{x} - (\nabla^2 f(\tilde{x}))^{-1} \nabla f(\tilde{x})$$

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

Newton’s Method to find minimum of a function

Algorithm:

Step 1: Chose an initial starting point $x^{(0)}$ and a termination parameter ε. Set k=0.

Step 2: Compute the gradient $\nabla f\left(x^{(k)}\right)$ and Hessian $\boldsymbol{H}\left(\boldsymbol{x}^{(k)}\right)$

Step 3: Evaluate $\boldsymbol{d}^{(k)} = -[\boldsymbol{H}\left(\boldsymbol{x}^{(k)}\right)]^{-\mathbf{1}}\nabla f\left(\boldsymbol{x}^{(k)}\right)$ (Solve $\nabla f\left(\boldsymbol{x}^{(k)}\right) + \boldsymbol{H}\left(\boldsymbol{x}^{(k)}\right)\boldsymbol{d}^{(k)} = \mathbf{0}$)

Step 4: Evaluate $x^{(k+1)} = x^{(k)} + d^{(k)}$

Step 5: Compute $\left\|\nabla f\left(x^{(k+1)}\right)\right\|$. If $\left\|\nabla f\left(x^{(k+1)}\right)\right\| < \varepsilon$, stop and mention $x^{(k+1)}$ is minimum,

Else go to step 2 with k=k+1.

Exercise

1. Find a minimum for the function, $f(x,y) = (x-2)^2 + (y-3)^2$ starting from the point (-1,4), using Newton's method. Choose the termination parameter $\varepsilon = 0.1$. Evaluate the gradients and Hessian analytically.

$$\nabla f = \begin{pmatrix} 2x-4 \\ 2y-6 \end{pmatrix} \qquad H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\bar{x}^{(0)} = (-1,4); \quad \nabla f(-1,4) = \begin{pmatrix} -6 \\ 2 \end{pmatrix}; \quad H(-1,4) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow \text{+ve definite matrix.}$$

$$\bar{x}^{(1)} = \bar{x}^{(0)} - [H]^{-1} \nabla f(-1,4)$$

$$= \begin{pmatrix} -1 \\ 4 \end{pmatrix} - \underbrace{\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}}_{2\times 2} \underbrace{\begin{pmatrix} -6 \\ 2 \end{pmatrix}}_{2\times 1}$$

$$= \begin{pmatrix} -1 \\ 4 \end{pmatrix} - \begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$\nabla f(\bar{x}^{(1)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \|\nabla f(x^{(1)})\| = 0 < \varepsilon$$

and $H(x^{(1)})$ is +ve definite matrix. $\Rightarrow$ (2,3) has a min. pt.

$\boxed{\text{OR}}$ Solve:-

$$d = -H^{-1} \nabla f.$$

$$\nabla f(x^{(0)}) + H(x^{(0)})\, d = \bar{0}$$

$$\begin{pmatrix} -6 \\ 2 \end{pmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$2d_1 = 6; \quad 2d_2 = -2$$

$$d_1 = 3, \quad d_2 = -1$$

$$x^{(1)} = x^{(0)} + d.$$

$$= \begin{pmatrix} -1 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Exercise

2. Find a minimum of $f(x,y) = 100(y-x^2)^2 + (1-x)^2$ starting from $(-1,1)$ with $\varepsilon = 0.1$.

$$\nabla f(x) = \begin{pmatrix} -400x(y-x^2) - 2(1-x) \\ 200(y-x^2) \end{pmatrix} ; \quad H = \begin{bmatrix} -400y + 1200x^2 + 2 & -400x \\ -400x & 200 \end{bmatrix}$$

$$= \begin{pmatrix} -400xy + 400x^3 - 2 + 2x \\ 200y - 200x^2 \end{pmatrix}$$

$$\nabla f(-1,1) = \begin{pmatrix} -4 \\ 0 \end{pmatrix}$$

$$H(-1,1) = \begin{bmatrix} 802 & 400 \\ 400 & 200 \end{bmatrix}$$

$$\bar{x}^{(1)} = \bar{x}^{(0)} - (H)^{-1} \nabla f$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \begin{bmatrix} 802 & 400 \\ 400 & 200 \end{bmatrix}^{-1} \begin{pmatrix} -4 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} -1 \\ 1 \end{pmatrix} - \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

$$\bar{x}^{(1)} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}, \quad \nabla f(\bar{x}^{(1)}) = \begin{pmatrix} 1600 \\ -800 \end{pmatrix}, \quad \|\nabla f(x^{(1)})\| \not< \varepsilon$$

$$\bar{x}^{(2)} = \bar{x}^{(1)} - \left[H(x^{(1)})\right]^{-1} \left(\nabla f(x^{(1)})\right)$$

$$= \begin{pmatrix} 1 \\ -3 \end{pmatrix} - \begin{bmatrix} 2402 & -400 \\ -400 & 200 \end{bmatrix}^{-1} \begin{pmatrix} 1600 \\ -800 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -3 \end{pmatrix} - \begin{pmatrix} 0 \\ -4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla f(x^{(2)}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad H(x^{(2)}) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

+ve definite

$$\left.\begin{array}{l} \|\nabla f(x^{(2)})\| = 0 < \varepsilon \\ \text{Also } H(x^{(2)}) \text{ is +ve definite} \end{array}\right\} \Longrightarrow \bar{x}^{(2)} \text{ is the minimum of the function.}$$

Exercise

3. Find the optimum point of the function, $f(x,y,z) = 2xy-2x^2 - y^2 + z + yz - z^2$ starting from the point (1, 1, 1), using Newton's method.

$$\nabla f = \begin{pmatrix} 2y-4x \\ 2x-2y+z \\ 1+y-2z \end{pmatrix}; \quad H = \begin{bmatrix} -4 & 2 & 0 \\ 2 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

$$\nabla f(1,1,1) = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$$

$$\bar{x}^{(1)} = \bar{x}^{(0)} - \left(H(\bar{x}^{(0)})\right)^{-1} \nabla f(x^{(0)})$$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{bmatrix} -0.75 & -1 & -0.5 \\ -1 & -2 & -1 \\ -0.5 & -1 & -1 \end{bmatrix} \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0.5 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$$

$$\nabla f(x^{(1)}) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \|\nabla f(x^{(1)})\| = 0$$

$\Rightarrow \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$ is a stationary pt.

Hessian matrix H is negative definite.

(Prin. minors:- $-4, 4, -4$)

$\therefore \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$ is a maximum point.

Exercise

4. Find a minimum for the function, $f(x,y,z) = 2x^2 - 2xy + y^2 - z - yz + z^2$ starting from the point (1, 2, 3), using Newton's method. Evaluate the gradients and Hessian analytically.

$$\nabla f = \begin{pmatrix} 4x - 2y \\ -2x + 2y - z \\ -1 - y + 2z \end{pmatrix}; \quad H = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

$$\nabla f(x^{(0)}) = \nabla f(1,2,3) = \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix}$$

$$\bar{x}^{(1)} = x^{(0)} - \left(H(x^{(0)})\right)^{-1} \left(\nabla f(x^{(0)})\right) = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{bmatrix} 4 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}^{-1} \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 0.5 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$$

$\nabla f(x^{(1)}) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$; $\|\nabla f(x^{(1)})\| < \varepsilon$, we can stop the algorithm.

$H(x^{(1)}) = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ is +ve definite $\left(\because \text{pr. minors are } 4 > 0,\ 8 - 4 = 4 > 0,\ 4(3) + 2(-4) = 4 > 0 \right)$

$\Rightarrow \bar{x}^{(1)} = \begin{pmatrix} 0.5 \\ 1 \\ 1 \end{pmatrix}$ is a minimum point.

5. Using MATLAB write a code and find the minimum of $f(x,y)=\sin x+y^2+4x^2-10y+50$ with starting point as (-3,5).

$$\nabla f = \begin{pmatrix} \cos x + 8x \\ 2y-10 \end{pmatrix}; \quad H = \begin{bmatrix} -\sin x + 8 & 0 \\ 0 & 2 \end{bmatrix}$$

```
x=[-3;5]; % starting point

for i=1:5
GradfAtx1x2=[cos(x(1))+8*x(1);2*x(2)-10];
H=[-sin(x(1))+8,0;0,2];
xc=pinv(H)*GradfAtx1x2;
x=x-xc ;
x
Lamda=eig(H)
end
```

This code gives x value & eigenvalues of H in each iteration as output.

Output:

Iter. 1, $\bar{x}^{(1)}$

```
x = 2×1
    0.0696
    5.0000

Lamda = 2×1
    2.0000
    8.1411
```

Iter. 2, $\bar{x}^{(2)}$

```
x = 2×1
   -0.1264
    5.0000

Lamda = 2×1
    2.0000
    7.9305
```

Iter. 3, $\bar{x}^{(3)}$

```
x = 2×1
   -0.1240
    5.0000

Lamda = 2×1
    2.0000
    8.1261
```

$\bar{x}^{(4)}$

```
x = 2×1
   -0.1240
    5.0000

Lamda = 2×1
    2.0000
    8.1237
```

$\bar{x}^{(5)}$

```
x = 2×1
   -0.1240
    5.0000

Lamda = 2×1
    2.0000
    8.1237
```

λs of H are +ve.

$\Rightarrow x=-0.1240$, $y=5$ is the minimum of $f(x,y)$.