Lecture - 2
MATHEMATICAL PRELIMINARIES

Reference: Book Chapter 2

VECTORS AND GEOMETRY

- A *scalar* is a quantity that is solely defined by its magnitude.
- A *vector* is a quantity that is determined by both its magnitude and its direction, and is a directed line segment.
- A space vector can be expressed in terms of its components in a Cartesian Coordinate system as $a = a_1 i + a_2 j + a_3 k$. The magnitude of this vector is given as

$$||\mathbf{a}|| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

- The length of a vector *x* is also called the norm (or Euclidean norm) of the vector, denoted as $|x|$. The position vector of a given point A : {x, y, z} is the vector *from the origin of the axes to the point* A.

DOT PRODUCT

- The *dot product* or *inner product* gives a scalar, as the product of two similarly-sized vectors.

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$$

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3$$

- Two vectors are said to be *orthogonal* if their dot product is equal to zero.

EQUATION OF A LINE

- The *slope-intercept form* of the equation of a line that has a slope of *m* and a *y*-intercept *c* is:

$$y = mx + c$$

- The *point-slope form* of the equation of a line that passes through a point P : (x_1, y_1), and has a slope *m* is

$$(y - y_1) = m(x - x_1)$$

- The *two-point form* of the equation of a line that passes through two points P1 : (x_1, y_1) and P2 : (x_2, y_2) is

$$(y - y_1)(x_2 - x_1) = (x - x_1)(y_2 - y_1)$$

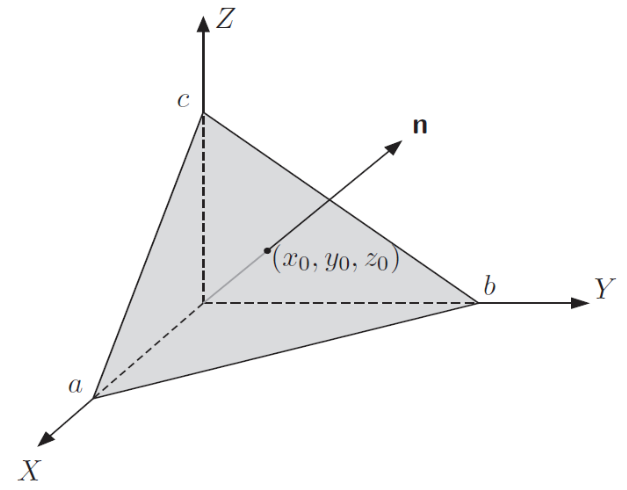
- The *two-intercept form of* a line with *x* intercept *a* and *y* intercept *b* is given as

$$\frac{x}{a} + \frac{y}{b} = 1$$

EQUATION OF A PLANE

- The equation of a plane with a nonzero normal vector $n = [a, b, c]^T$ passing through the point $V_0=(x_0, y_0, z_0)$ is given as

$$n \cdot (V - V_0) = 0$$



- The general form is:

$$ax + by + cz + d = 0$$

where $d = -ax_0 - by_0 - cz_0$.

- The plane specified in this form has its X, Y and Z intercepts at

$$p = \frac{-d}{a},\ q = \frac{-d}{b} \text{ and } r = \frac{-d}{c}.$$

EQUATION OF A PLANE (CONTINUED)

- In the intercept form, a plane passing through the points (*a, 0, 0*), (*0, b, 0*), and (*0, 0, c*) is given by

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1$$

- The plane through (x_1, y_1, z_1) and (x_2, y_2, z_2), and parallel to the direction $[a, b, c]^T$ is given by

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ a & b & c \end{vmatrix} = 0$$

EQUATION OF A PLANE (CONTINUED)

- The plane that goes through three points (x_1, y_1, z_1), (x_2, y_2, z_2), and (x_3, y_3, z_3) is given by

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0$$

MATRIX OPERATIONS

- **Addition**: $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \end{bmatrix}_{2\times 3}$ $B = \begin{bmatrix} 5 & 6 & 1 \\ 3 & 0 & 0 \end{bmatrix}_{2\times 3}$ $A + B = \begin{bmatrix} 6 & 6 & 3 \\ 3 & 3 & -1 \end{bmatrix}_{2\times 3}$

- **Subtraction**: $A - B = \begin{bmatrix} -4 & -6 & 1 \\ -3 & 3 & -1 \end{bmatrix}_{2\times 3}$

Only matrices of the same order can undergo addition/subtraction operations.

- **Scalar Multiplication**: $4A = \begin{bmatrix} 4 & 0 & 8 \\ 0 & 12 & -4 \end{bmatrix}$

- **Matrix Multiplication**:

  The product of these two matrices, *AB*, is defined only if the number of columns in *A* is equal to the number of rows in *B*. If *A* is a $m \times n$ matrix, and *B* is a $n \times p$ matrix, then the product of *A* and *B* has an order of $m \times p$.

MATRIX MULTIPLICATION

1. $AB \neq BA$ in general
2. $AI = IA = A$
3. $AB = 0$ does not necessarily imply $A = 0$ or $B = 0$
4. $k(AB) = (kA)B = A(kB)$
5. $A(BC) = (AB)C$
6. $(A + B)C = AC + BC$
7. $A(B + C) = AB + AC$

TRANSPOSE

- The transpose of a matrix is obtained by interchanging its rows and columns.
- Example:

$$C = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 4 \end{bmatrix}_{2\times 3} \qquad C^T = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 4 \end{bmatrix}_{3\times 2}$$

1. If $A^T = A$, then A is called a <u>Symmetric Matrix</u>. If $A^T = -A$, then A is called a <u>Skew Symmetric Matrix</u>.
2. $(A^T)^T = A$
3. $(A + B)^T = A^T + B^T$
4. $(AB)^T = B^T A^T$
5. $(cA)^T = cA^T$

DETERMINANTS

$$
\begin{aligned}
det(A) &= \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \\
&= \sum_{i=1}^{n} a_{ij} C_{ij} \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{i+j} M_{ij}
\end{aligned}
$$

C_{ij} is called the *cofactor* of *A*, and M_{ij} is the corresponding *minor*.

DETERMINANTS (CONTINUED)

- Example:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$$

$$\begin{aligned} det(A) &= \begin{vmatrix} 2 & 1 \\ 3 & -1 \end{vmatrix} \\ &= (2)(-1) - (3)(1) \\ &= -5 \end{aligned}$$

DETERMINANTS (CONTINUED)

- $\det(A^T) = \det(A)$
- $\det(I) = 1$
- If two rows or columns of a matrix are identical,

  then $\det(A) = 0$.
- If all entries of a row or column are all zeros,

  then $\det(A) = 0$.
- If B is obtained by interchanging any two rows or columns of A,

  then $\det(B) = -\det(A)$

DETERMINANTS (CONTINUED)

- If a row or column of a matrix is a linear combination of two or more columns, then det(A) = 0
- det(AB) = det(A)det(B)
- c is a scalar, and A is a n×n matrix. det(cA) = c^ndet(A)
- If the determinant of a matrix is *0*, the matrix is said to be *singular*. Otherwise, the matrix is said to be *non-singular*.
- If a matrix is upper triangular or lower triangular, the determinant of that matrix is simply equal to the product of the elements in the principal diagonal.

INVERSE

- For a square matrix A, if there exists a matrix B, such that, $AB = BA = \mathrm{I}$, then B is the inverse of A.

$$A^{-1} = \frac{1}{det(A)}[Adjoint(A)]$$

where Adjoint(A) = C^T , and C is the matrix of co-factors of A.

1. $AA^{-1} = I$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^T)^{-1} = (A^{-1})^T$

INVERSE (CONTINUED)

- Example:

$$
\begin{aligned}
A &= \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix} \\
A^{-1} &= \frac{1}{3 \cdot 4 - 2 \cdot 1} \begin{bmatrix} 4 & -1 \\ -2 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix}
\end{aligned}
$$

EIGENVALUES

- For an $n \times n$ matrix, find all scalars λ such that the equation

$$Ax = \lambda x$$

  has a nonzero solution x. Such a scalar λ is called the eigenvalue of A, and any nonzero $n \times 1$ vector, x, is called the eigenvector corresponding to λ.

- Given a matrix A, the eigenvalues of A, denoted by λ, can be computed by finding the roots of the equation $|A - \lambda I| = 0$.

EIGENVECTORS

Given a matrix A, consider the equation $(A - \lambda I)x = 0$, where λ is an eigenvalue of A. The non zero values of x that satisfy the above equation for each eigenvalue are called Eigenvectors of A.

EIGENVECTORS (CONTINUED)

- Example:

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}_{2\times 2}$$

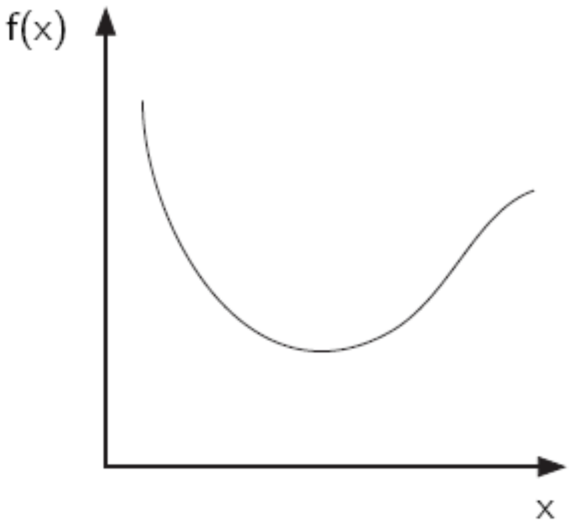
$$\begin{aligned}
A - \lambda I &= \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \\
&= \begin{bmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{bmatrix} \\
|A - \lambda I| &= \begin{vmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} \\
&= (-5-\lambda)(-2-\lambda) - 4 \\
&= \lambda^2 + 7\lambda + 6 \\
&= 0
\end{aligned}$$

$$\lambda_1 = -1 \text{ and } \lambda_2 = -6$$

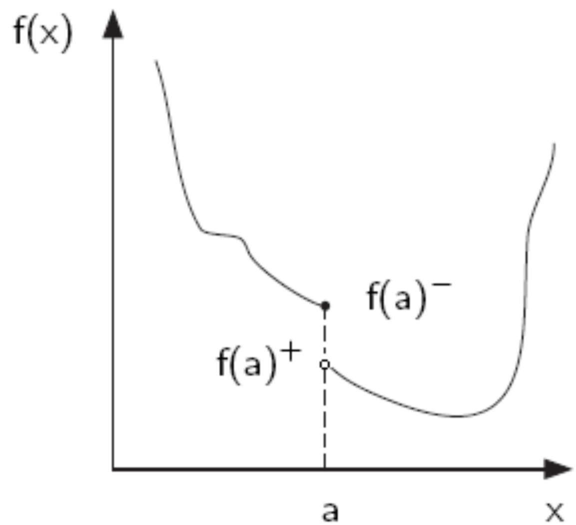
POSITIVE DEFINITENESS

- A matrix is said to be positive definite if all its eigenvalues are positive.
- A matrix is said to positive semi definite if all its eigenvalues are nonnegative (that is, including zero).
- A matrix is said to be negative definite if all its eigenvalues are negative.
- A matrix is said to negative semi definite if all its eigenvalues are non-positive (that is, including zero).

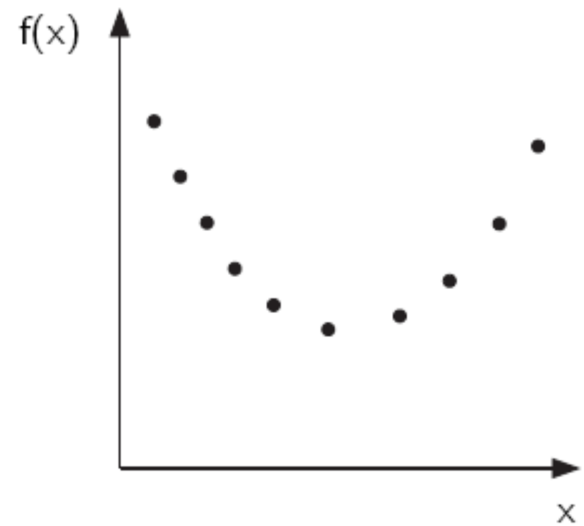
TYPES OF FUNCTIONS: CONTINUITY



Continuous function

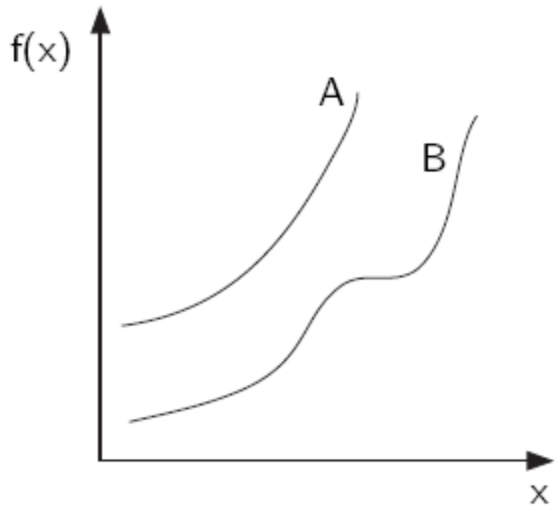


Discontinuous function

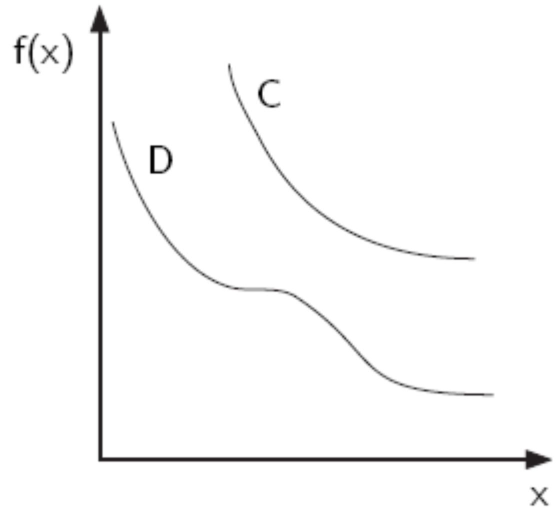


Discrete function

TYPES OF FUNCTIONS: MONOTONICITY

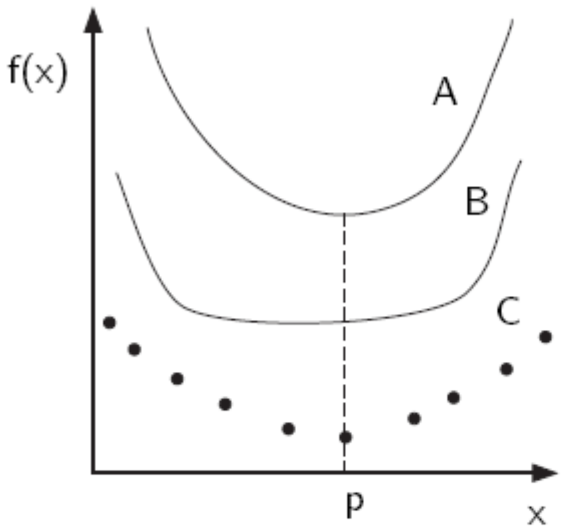


Monotonically increasing

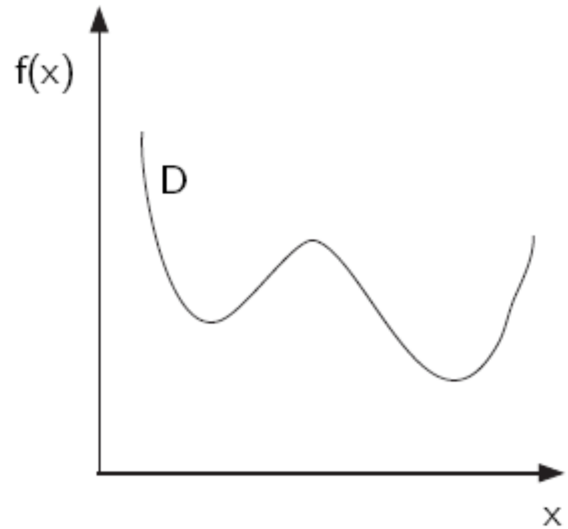


Monotonically decreasing

TYPES OF FUNCTIONS: MODALITY

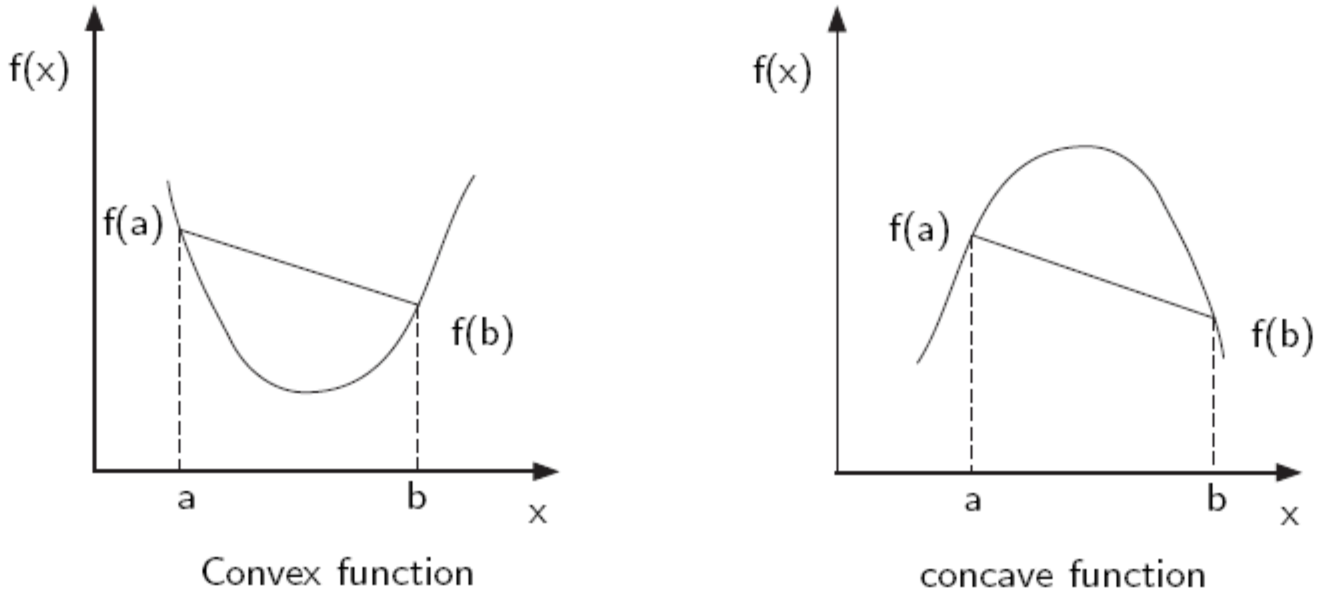


Unimodal



Multimodal

TYPES OF FUNCTIONS: CONVEXITY



- f is called **convex** if:

$$\forall x_1, x_2 \in X, \forall t \in [0,1] : \qquad f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

- f is called **strictly convex** if:

$$\forall x_1 \neq x_2 \in X, \forall t \in (0,1) : \qquad f(tx_1 + (1-t)x_2) < tf(x_1) + (1-t)f(x_2).$$

- A function f is said to be (strictly) concave if $-f$ is (strictly) convex.

LIMITS OF FUNCTIONS

- The function *f*(*x*) reaches a value *L* as *x* approaches the value x_0.

$$\lim_{x \longrightarrow x_0} f(x) = L$$

- Example:

Let the function be $f(x) = ((a+x)^2 - a^2)/x$. We wish to find the limit of this function as x tends to 0.

$$\begin{aligned}
& \lim_{x \longrightarrow x_0} \frac{(a+x)^2 - a^2}{x} \\
= & \lim_{x \longrightarrow x_0} \frac{a^2 + 2xa + x^2 - a^2}{x} \\
= & \lim_{x \longrightarrow x_0} \frac{2xa + x^2}{x} \\
= & \lim_{x \longrightarrow x_0} 2a + x
\end{aligned}$$

Since, we want to find the limit of $f(x)$ as x tends to 0, we simply substitute $x = 0$ in the above equation, and we obtain

$$\lim_{x \longrightarrow 0} f(x) = 2a$$

DERIVATIVE

- The derivative of a function *f*(*x*) at a certain point *x* is a measure of the rate at which that function is changing as the variable *x* changes at that point.

$$\frac{d\ f(x)}{dx} = \lim_{h \longrightarrow 0} \frac{f(x+h) - f(x)}{h}$$

PARTIAL DERIVATIVE

- Partial derivatives are defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the differentiation.

$$\frac{\partial f}{\partial x} = \lim_{h \longrightarrow 0} \frac{f(x+h,y) - f(x,y)}{h}$$

$$\frac{\partial f}{\partial y} = \lim_{h \longrightarrow 0} \frac{f(x,y+h) - f(x,y)}{h}$$

INDEFINITE INTEGRATION

- Let us assume that the derivative of a function *f*(*x*) is *F*(*x*).

- Example:

$$f(x) = \int F(x)dx + C$$

$$\int F(x)dx = \int (x^2 + 2)dx = \frac{x^3}{3} + 2x + C$$

DEFINITE INTEGRATION

- Definite integration is the integration of a function, *F*(*x*), over a particular range of the variable *x*.

$$\int_a^b F(x)dx = [f(x)]_a^b = f(b) - f(a)$$

- Example:

$$\begin{aligned}\int_1^2 (x^2+2)dx &= \left[\frac{x^3}{3} + 2x + C\right]_1^2 \\ &= \left[\left(\frac{2^3}{3} + 2(2) + C\right) - \left(\frac{1^3}{3} + 2(1) + C\right)\right] \\ &= \left[\frac{8}{3} + 4 + C - \frac{1}{3} - 2 - C\right] = \frac{13}{3}\end{aligned}$$

TAYLOR SERIES

- The Taylor series is a way of approximating a function in terms of its derivatives. The approximation is accurate around a chosen point, and become progressively inaccurate as we move away from that point.

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)\frac{(x-x_0)}{1!} \\ &+ f''(x_0)\frac{(x-x_0)^2}{2!} + f'''(x_0)\frac{(x-x_0)^3}{3!} + \ldots \end{aligned}$$

- The 2nd order Taylor series expansion:

$$f(x) = f(x_0) + f'(x_0)\frac{(x-x_0)}{1!} + f''(x_0)\frac{(x-x_0)^2}{2!}$$

TAYLOR SERIES (CONTINUED)

- Example:

$$\sin(x) = \sin(\pi/2) + \cos(\pi/2)\frac{(x-\pi/2)}{1!} - \sin(\pi/2)\frac{(x-\pi/2)^2}{2!}$$

Since, $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$, we have

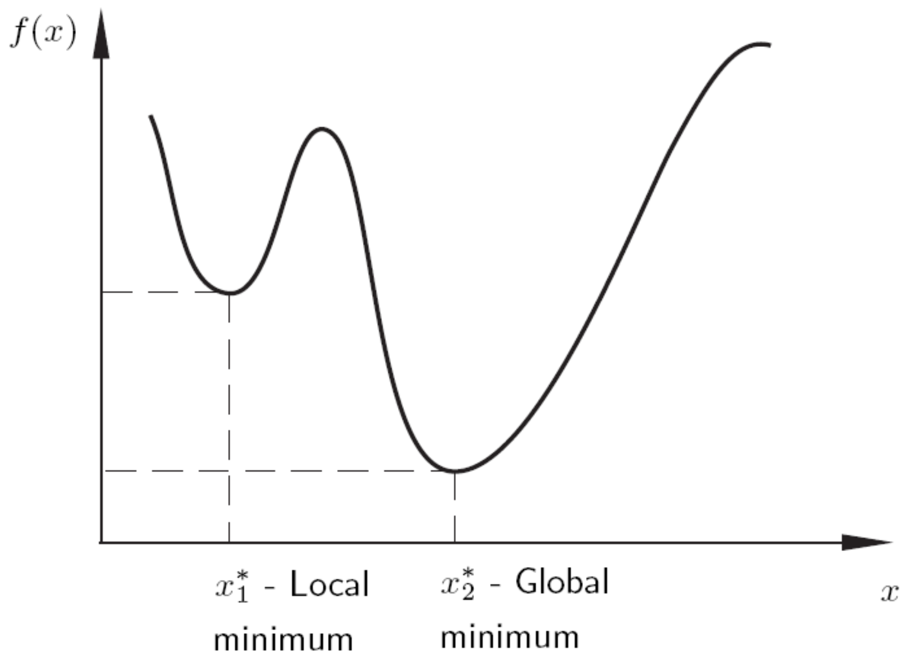
$$\sin(x) = 1 - \frac{(x-\pi/2)^2}{2!}$$

OPTIMIZATION BASICS

- Unconstrained Optimization: no constraints on the values x.
- Constrained Optimization: constraints on the values x.

OPTIMIZATION BASICS: LOCAL VS. GLOBAL

- The function $f(x)$ is said to be at a *global minimum* at a point $x^* \in S$ if $f(x^*) \leq f(x)$, for all $x \in S$.
- The function $f(x)$ is said to be at a *local minimum* at a point $x^* \in S$ if $f(x^*) \leq f(x)$ for all x within a distance ϱ from x^*, that is, there exists a $\varrho > 0$ such that for all x satisfying $|x - x^*| < \varrho, f(x^*) \leq f(x)$.



NECESSARY CONDITIONS FOR LOCAL OPTIMUM

- Assuming that the first and second derivatives of the function *f*(x) exist, the necessary conditions for x^* to be *a local minimum* of the function *f*(x) on an interval (*a, b*) are:

$$1.\quad \left.\frac{df}{dx}\right|_{x=x^*} = 0$$

$$2.\quad \left.\frac{d^2f}{dx^2}\right|_{x=x^*} \geq 0$$

- The necessary conditions for x^* to be *a local maximum* of the function *f*(x) on an interval (*a, b*) are:

$$1.\quad \left.\frac{df}{dx}\right|_{x=x^*} = 0$$

$$2.\quad \left.\frac{d^2f}{dx^2}\right|_{x=x^*} \leq 0$$

STATIONARY POINTS AND INFLECTION POINTS

- A *Stationary point* is a point x^* that satisfies the following equation:

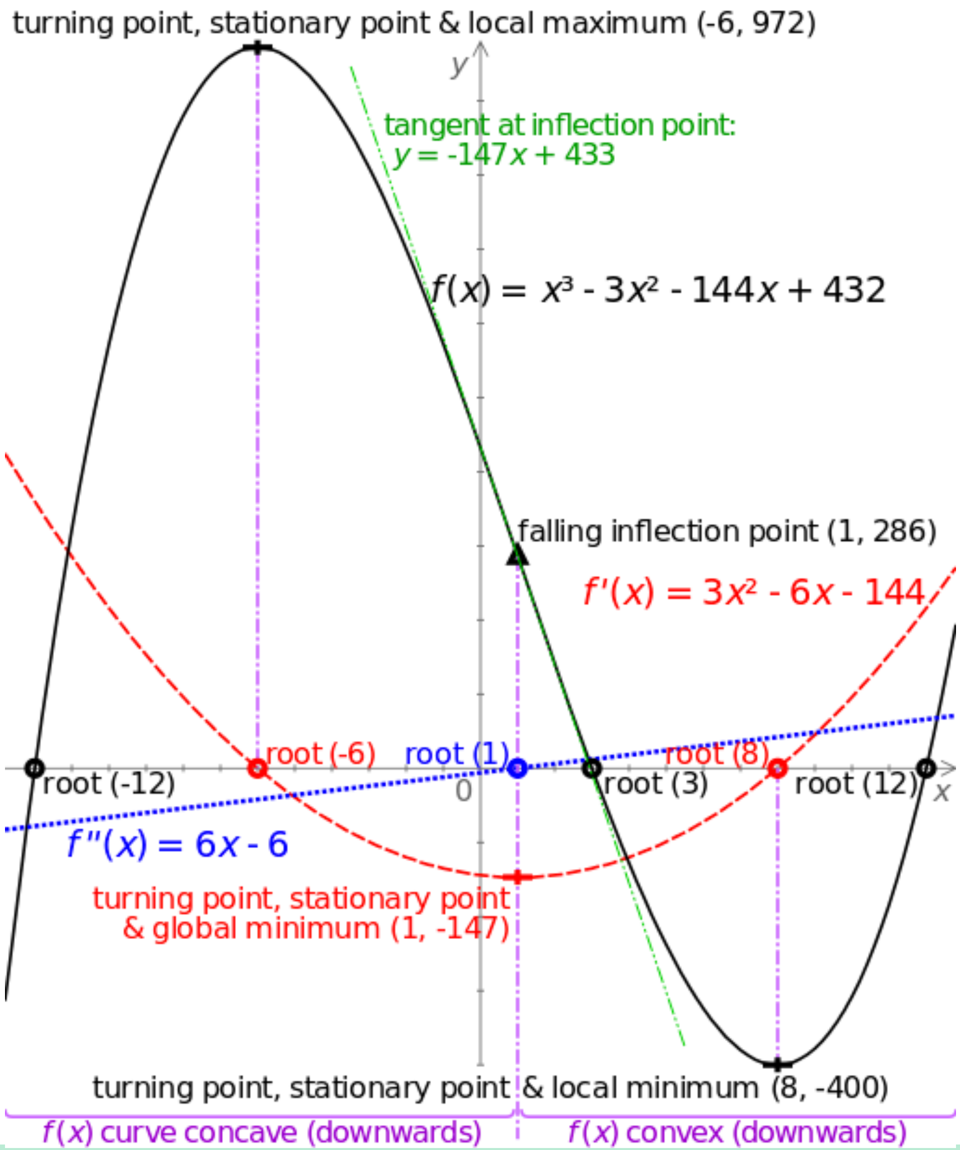
$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

- If a stationary point is not a local minimum or maximum, it is an *inflection point*. The sufficient conditions of optimality can be used to differentiate between stationary points and inflection points.
- An inflection point is a point on a continuously differentiable plane curve at which the curve crosses its tangent, that is, the curve changes from being concave (concave downward) to convex (concave upward), or vice versa.
- Points of inflection can also be categorized according to whether f′(x) is zero or not zero.
  - if f′(x) is zero, the point is a stationary point of inflection
  - if f′(x) is not zero, the point is a non-stationary point of inflection

SUFFICIENT CONDITIONS FOR LOCAL OPTIMA

Consider a point x^*, at which the first derivative of $f(x)$ is equal to zero, and the order of the first nonzero higher derivative is n. The following are the sufficient conditions for x^* to be a local optimum.

1. If n is odd, then x^* is *an inflection point*.
2. If n is even, then x^* is *a local optimum*. Also,
   a. *If the value of that derivative of $f(x)$ at x^* is positive, then the point x^* is a local minimum.*
   b. *If the value of that derivative of $f(x)$ at x^* is negative, then the point x^* is a local maximum.*



The roots, turning points, stationary points, inflection point and concavity of a cubic polynomial $x^3 - 3x^2 - 144x + 432$ (black line) and its first and second derivatives (red and blue).

GRADIENT OF A FUNCTION

For a function $f(x)$, its gradient is the vector of the 1st order partial derivatives, and is given by

$$\nabla(f(x)) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{Bmatrix}$$

Example:

$$f(x) = 2x_1^4 + 4x_2^3 - 3x_1x_2^2$$

$$\nabla(f(x)) = \begin{Bmatrix} 8x_1^3 - 3x_2^2 \\ \\ 12x_2^2 - 6x_1x_2 \end{Bmatrix}$$

HESSIAN OF A FUNCTION

The Hessian of a function *f*(*x*) is the matrix of the 2nd order partial derivatives of the function, and is given by the following matrix

$$\nabla^2(f(x)) \equiv H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

HESSIAN OF A FUNCTION (EXAMPLE)

Example:

$$f(x) = 2x_1^4 + 4x_2^3 - 3x_1x_2^2$$

$$\nabla(f(x)) = \begin{Bmatrix} 8x_1^3 - 3x_2^2 \\ \\ 12x_2^2 - 6x_1x_2 \end{Bmatrix}$$

$$\nabla^2(f(x)) = \begin{bmatrix} \frac{\partial(8x_1^3 - 3x_2^2)}{\partial x_1} & \frac{\partial(8x_1^3 - 3x_2^2)}{\partial x_2} \\ \\ \frac{\partial(12x_2^2 - 6x_1x_2)}{\partial x_1} & \frac{\partial(12x_2^2 - 6x_1x_2)}{\partial x_2} \end{bmatrix}$$

$$\nabla^2(f(x)) = \begin{bmatrix} 24x_1^2 & -6x_2 \\ \\ -6x_2 & 24x_2 - 6x_1 \end{bmatrix}$$

USING OPTIMALITY CRITERIA: EXAMPLE

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

Determine the minimum of $f(x)$ by hand. Report the values of the optimum x and the function value at this point.

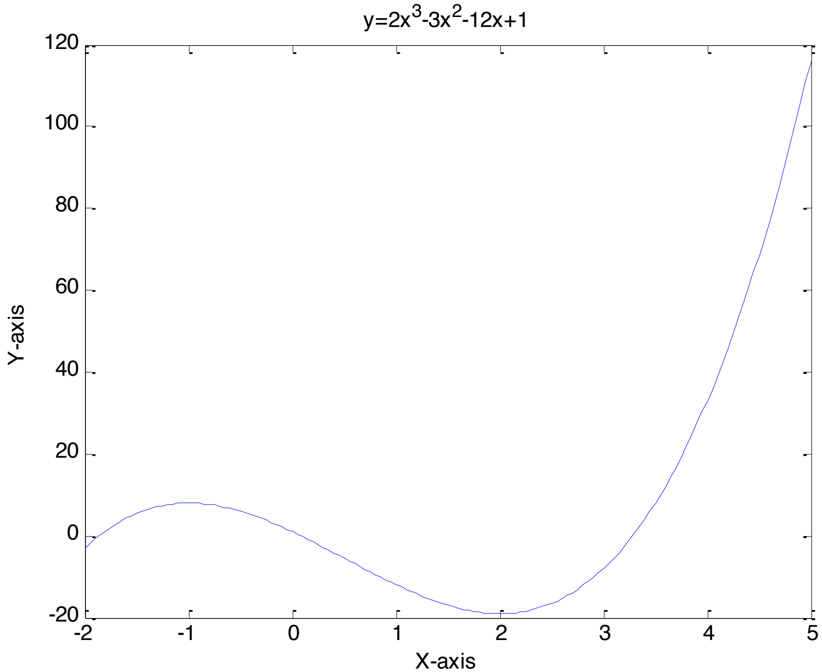
$f'(x) = 6x^2 - 6x - 12 = 6(x - 2)(x + 1) = 0$

$x_1 = 2$ and $x_2 = -1$. $f(x_1) = -19$, $f(x_2) = 8$

$f''(x) = 12x - 6$

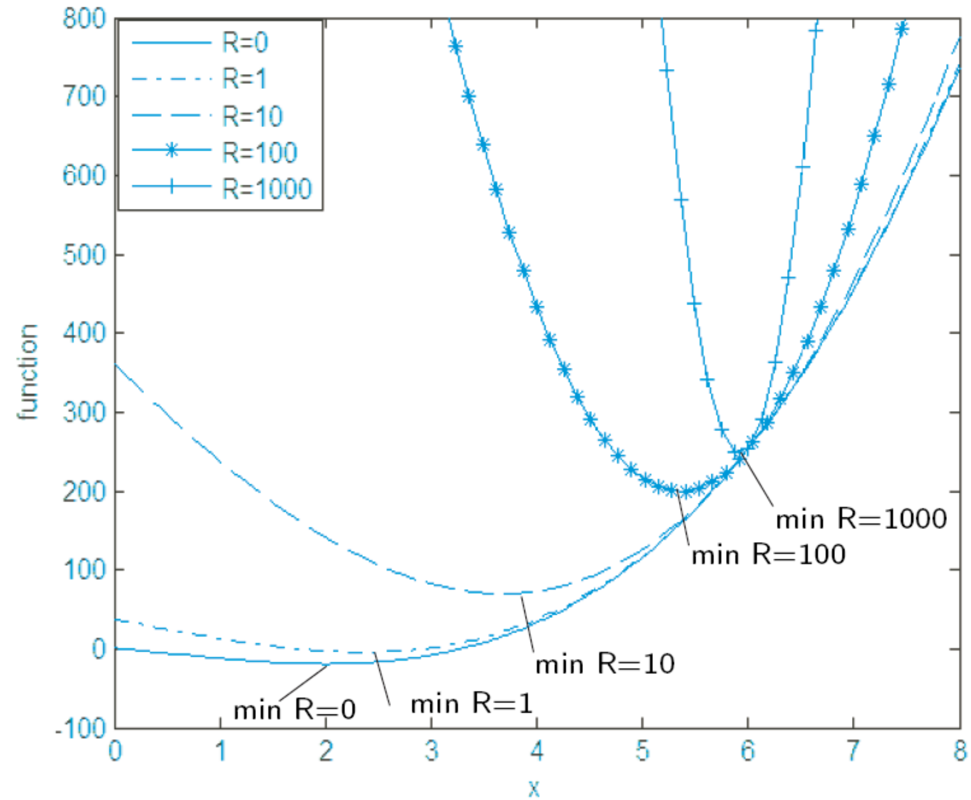
At point $x_1 = 2$, $f''(x) = 18 > 0$, $f(x) = -19$.
It is the local minimum.

At point $x_2 = -1$, $f''(x) = -18 < 0$, $f(x) = 8$.
It is the local maximum.



CONSTRAINED OPTIMIZATION: EXAMPLE

- This sub-problem is a simple demonstration of how constrained optimization is sometimes performed.
- Let us add another component to the previous function, giving it the form, $f(x) = 2x^3 - 3x^2 - 12x + 1 + R(x - 6)^2$, where R is a constant.
- Plot this new function for R = 0, 1, 10, 100, and 1000, for $0 \leq x \leq 8$ on the same figure. Use different line types for each R. Keep your vertical axis between −100 and 800.



OPTIMIZATION: EXAMPLE (CONTINUED)

- By looking at the plot, can you tell what the minimum of the new function is, for the different values of *R*? Indicate these minima on the plot, and compare them with the minimum of the original function. What is the minimum of the new function if *R* = ∞?

- The minimum of the function $f(x)$ approaches $x = 6$ as R increases. As R tends to ∞, the minimum of f(x) = f(6) = 253.