MA2000: OTML

Nachiketa Mishra

Indian Institute of Information Technology, Design & Manufacturing, Kancheepuram

Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1-\lambda) y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0,1].$$

Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1-\lambda)y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0,1].$$

Note that the point $\lambda x + (1-\lambda)y$ is called a convex combination of x and y.

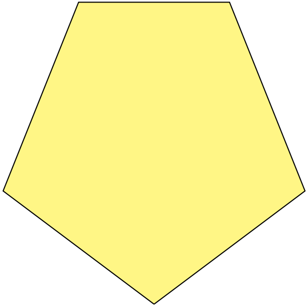
Convex sets

Definition

A set $X \subseteq \mathbb{R}^{\mathrm{n}}$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1 - \lambda) y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0, 1].$$

Note that the point $\lambda x + (1 - \lambda) y$ is called a convex combination of x and y.



Example of a convex set (left) and non-convex set (right).

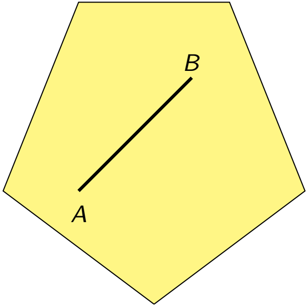
Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1-\lambda) y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0, 1].$$

Note that the point $\lambda x + (1-\lambda) y$ is called a convex combination of x and y.



Example of a convex set (left) and non-convex set (right).

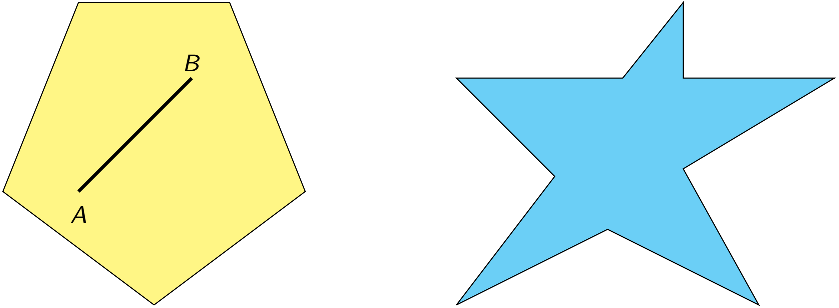
Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1-\lambda)y \in X, \qquad \forall (x,y,\lambda) \in X \times X \times [0,1].$$

Note that the point $\lambda x + (1-\lambda)y$ is called a convex combination of x and y.



Example of a convex set (left) and non-convex set (right).

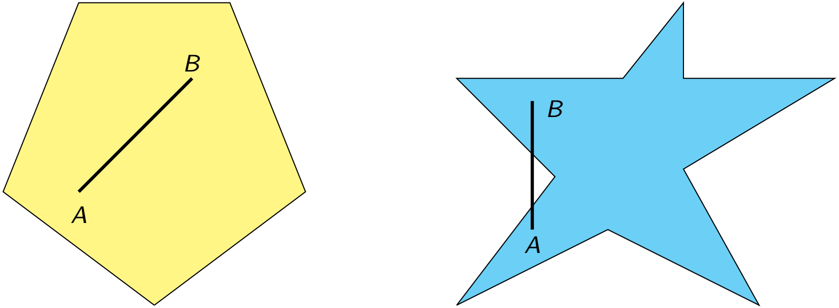
Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1-\lambda) y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0,1].$$

Note that the point $\lambda x + (1-\lambda) y$ is called a convex combination of x and y.



Example of a convex set (left) and non-convex set (right).

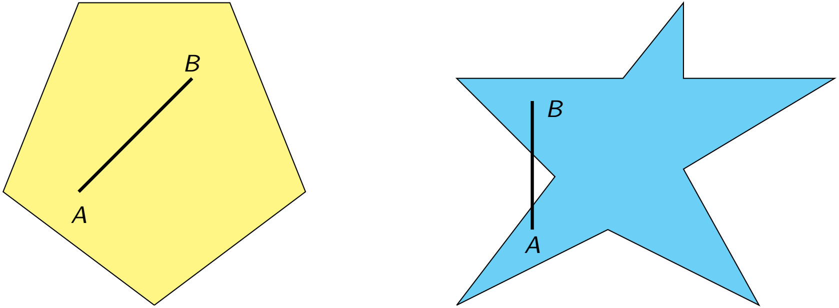
Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is said to be convex if it contains all of its segments, that is

$$\lambda x + (1 - \lambda) y \in X, \qquad \forall (x, y, \lambda) \in X \times X \times [0, 1].$$

Note that the point $\lambda x + (1 - \lambda) y$ is called a convex combination of x and y.



Example of a convex set (left) and non-convex set (right).

Definition (Extreme point or vertex of a convex set)

An Extreme point (vertex) of a convex set is a point of the set which does not lie on any segment joining two other point of the set

Definition (convex combination of vectors)

Given a set of vectors $\{x_1, x_2, \cdots, x_k\}$, a linear combination

$$x = \lambda_1 x_1 + \lambda_2 x_2, \cdots + \lambda_k x_k$$

is called convex combination of given vectors, if

$$\lambda_1, \lambda_2, \cdots, \lambda_k \geq 0, \text{ and } \sum_{i=1}^{k} \lambda_i = 1$$

Theorem

The Set of all convex combination of finite number of points of $S \subset \mathbb{R}^n$ is a convex set

Proof.

Let

$$S = \left\{ x : x = \sum_{i=1}^{m} \lambda_i x_i, \sum_{i=1}^{m} \lambda_i = 1 \right\}$$

we have to show that S is convex. Let x' and x'' be in S, so that

$$x' = \sum_{i=1}^{m} \lambda_i' x_i, \text{ where } \lambda_i' \geq 0, \sum_{i=1}^{m} \lambda_i' = 1$$

$$x'' = \sum_{i=1}^{m} \lambda_i'' x_i, \text{ where } \lambda_i'' \geq 0, \sum_{i=1}^{m} \lambda_i'' = 1$$

Consider now the vector

$$x = \lambda x' + (1 - \lambda) x'', \quad 0 \leq \lambda \leq 1$$

proof continue ⋯

$$
\begin{aligned}
&= \lambda \sum_{i=1}^{m} \lambda_i' x_i + (1-\lambda) \sum_{i=1}^{m} \lambda_i'' x_i, \\
&= \sum_{i=1}^{m} [\lambda \lambda_i' + (1-\lambda)\lambda_i''] x_i = \sum_{i=1}^{m} \mu_i x_i
\end{aligned}
$$

$$\text{where } \mu_i = \lambda\lambda_i' + (1-\lambda)\lambda_i', \quad i = 1, 2, \cdots, m.$$

Since $0 \leq \lambda \leq 1, \quad \lambda_i' \geq 0, \quad \lambda_i'' \geq 0$ therefore $\mu_i \geq 0$. Also

$$
\begin{aligned}
\sum_{i=1}^{m} \mu_i &= \sum_{i=1}^{m} [\lambda\lambda_i' + (1-\lambda)\lambda_i'] \\
&= \lambda \sum_{i=1}^{m} \lambda_i' + (1-\lambda) \sum_{i=1}^{m} \lambda_i'' \\
&= \lambda + (1-\lambda) = 1
\end{aligned}
$$

□

proof continue $\cdots$

- We have proved that $\mu_i \geq 0, \ \forall \ i$ and $\sum_{i=1}^{m} \mu_i = 1$
- x is the convex combination of vectors $x_1, x_2, \cdots, x_k$ or $x \in S$.
- Thus each pair of points $x', x'' \in S$ that we consider
- The line segment joining them is connected in the set.
- Hence S is convex set

$\square$

Example-: 01

Prove that $C = \{(x_1, x_2) : 2x_1 + 3x_2 = 7\} \subset \mathbb{R}^2$ is a convex set.

SOLUTION:

Assume that $X, Y \in C$, where $X = (x_1, x_2)$, $Y = (y_1, y_2)$. The line segment connecting X and Y is the set.

From the definition of convex sets, we can write the following:

$$W = W : W = \theta X + (1-\theta)Y, 0 \le \theta \le 1$$

For some $0 \le \theta \le 1$, assume that $W = (w_1, w_2)$ is the point of set W. Hence, we can write

$$w_1 = \theta x_1 + (1-\theta)y_1$$
$$w_2 = \theta x_2 + (1-\theta)y_2$$

As $x, y \in C$, we can write

$$2x_1 + 3x_2 = 7$$
$$2y_1 + 3y_2 = 7$$

But, from the formula,

$$2w_1 + 3w_2 = 2[\theta x_1 + (1-\theta)y_1] + 3[\theta x_2 + (1-\theta)y_2]$$

Now, take the common terms outside, we get

$$= \theta[2x_1 + 3x_2] + (1-\theta)[2y_1 + 3y_2]$$

$$= \theta \times 7 + (1-\theta) \times 7 = 7 \qquad [2x_1 + 3x_2 = 7]$$

Hence, $W = (w_1, w_2)$ belongs to C Since W is any point of C, and $X, Y \in C$ This can be written as $[X : Y] \subset C$. Therefore, set C is convex.

Example-: 02

Show that the following set is convex

$$S = \{(x_1, x_2) : 3x_1^2 + 2x_2^2 \leq 6\}$$

SOLUTION:

Let $\mathbf{x}, \mathbf{y} \in S$ where $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$.

The line segment joining $\mathbf{x}$ and $\mathbf{y}$ is the set:

$$\{\mathbf{u} : \mathbf{u} = \lambda\mathbf{x} + (1 - \lambda)\mathbf{y}, \quad 0 \leq \lambda \leq 1\}$$

For some $\lambda, 0 \leq \lambda \leq 1$, let $\mathbf{u} = (u_1, u_2)$ be a point of this set, so that

$$u_1 = \lambda x_1 + (1 - \lambda)y_1 \quad \textit{and} \quad u_2 = \lambda x_2 + (1 - \lambda)y_2$$

Now,

$$
\begin{aligned}
3u_1^2 + 2u_2^2 =& 3[\lambda x_1 + (1-\lambda)y_1]^2 + 2[\lambda x_2 + (1-\lambda)y_2]^2 \\
=& \lambda^2(3x_1^2 + 2x_2^2) + (1-\lambda)^2(3y_1^2 + 2y_2^2) + 2\lambda(3x_1y_1 + 2x_2y_2) \\
\leq& 6\lambda^2 + 6(1-\lambda)^2 + 12\lambda(1-\lambda)
\end{aligned}
$$

Since $(3x_1y_1 + 2x_2y_2) \leq \sqrt{(x_1\sqrt{3})^2 + (x_2\sqrt{2})^2}\sqrt{(y_1\sqrt{3})^2 + (y_2\sqrt{2})^2}$

Thus, $3u_1^2 + 2u_2^2 \leq 6$ and hence $\mathbf{u} = (u_1, u_2)$ is a point on S

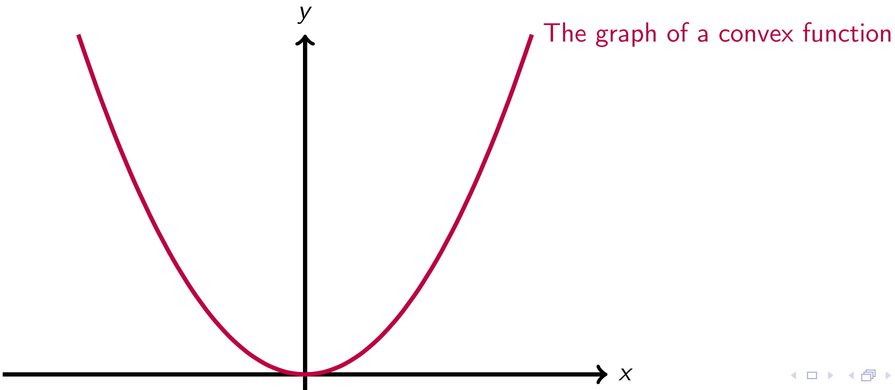
Hence S is convex set

Convex Functions

Definition

Let $X \subseteq \mathbb{R}^{n}$ be a given convex set. A function $f : X \Rightarrow \mathbb{R}$ is said to be convex if it always lies below its chords , that is

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \forall (x, y, \lambda) \in X \times X \times [0, 1].$$

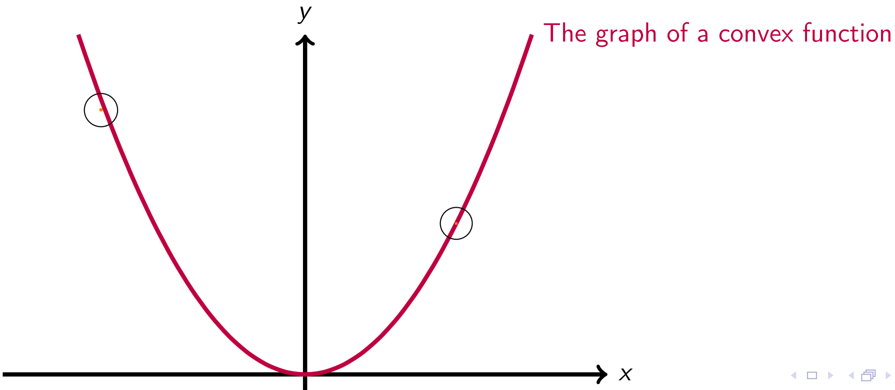


Convex Functions

Definition

Let $X \subseteq \mathbb{R}^n$ be a given convex set. A function $f : X \Rightarrow \mathbb{R}$ is said to be convex if it always lies below its chords , that is

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \forall (x,y,\lambda) \in X \times X \times [0,1].$$

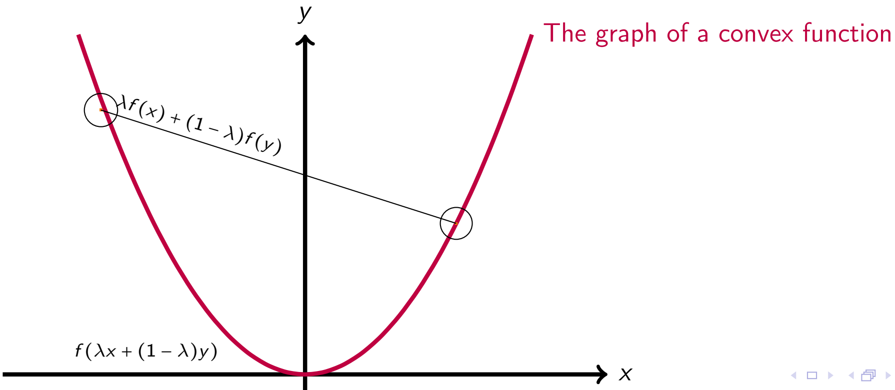


Convex Functions

Definition

Let $X \subseteq \mathbb{R}^n$ be a given convex set. A function $f : X \Rightarrow \mathbb{R}$ is said to be convex if it always lies below its chords , that is

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \forall (x, y, \lambda) \in X \times X \times [0, 1].$$

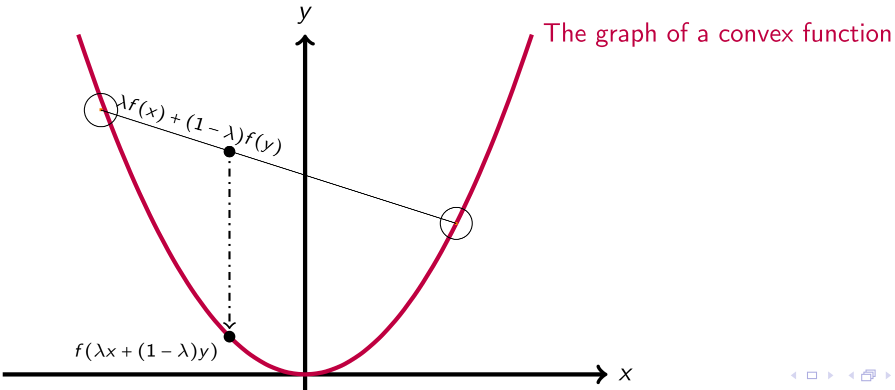


Convex Functions

Definition

Let $X \subseteq \mathbb{R}^n$ be a given convex set. A function $f : X \Rightarrow \mathbb{R}$ is said to be convex if it always lies below its chords , that is

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \forall (x, y, \lambda) \in X \times X \times [0, 1].$$



We say a function is strictly convex if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ holds with strict inequality for any $x \neq y$ and $\lambda \in (0,1)$. We say that f is concave if $-f$ is convex, and similarly that f is strictly concave if $-f$ is strictly convex.

Some examples of convex functions are given as follows.

1. Exponential, $f(x) = exp(ax)$; for any $a \in \mathbb{R}^{\mathrm{n}}$.
2. Negative logarithm, $f(x) = -\log x$ with $x > 0$
3. Affine functions, $f(x) = w^T x + b$
4. Quadratic functions, $f(x) = \frac{1}{2}X^T AX$ with $A \in S_+^n, A \geq 0$
5. Norms $f(x) = ||x||$
6. Non-negative weighted sums of convex functions. Let $f_1, f_2, \cdots, f_k$ be convex functions and $w_1, w_2, \cdots, w_k$ be non-negative real numbers.Then $f(x) = \sum_{i=1}^n w_i f_i(x)$

Example

Is the function, $f(x) = |x|$ a convex function?

SOLUTION:

To prove this we need to check the definition:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \forall (x, y, \lambda) \in X \times X \times [0, 1]$$

Furthermore, these inequalities have to be true for all $x, y \in \mathbb{R}^n$ and every $\lambda \in [0, 1]$. It is not enough to simply pick a few values randomly and check the equations. So, we have to work symbolically. In this case,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= |\lambda x + (1-\lambda)y| \\ &\leq |\lambda x| + |(1-\lambda)y| \\ &= \lambda|x| + (1-\lambda)|y|, \qquad \lambda, 1-\lambda \geq 0 \\ &= \lambda f(x) + (1-\lambda)f(y) \end{aligned}$$

$\Rightarrow f$ is convex.

Example

Show that $f(x) = x^2, x \in \mathbb{R}$ is strictly convex.

SOLUTION:

Pick x_1, x_2 so that $x_1 \neq x_2$, and pick $\lambda \in (0,1)$.

$$\begin{aligned} f((1-\lambda)x_1 + \lambda x_2) &= ((1-\lambda)x_1 + \lambda x_2)^2 \\ &= (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + 2(1-\lambda)\lambda x_1 x_2 \end{aligned}$$

Since, $x_1 \neq x_2, (x_1 - x_2)^2 > 0 \Rightarrow x_1^2 + x_2^2 > 2x_1x_2$

Thus,

$$\begin{aligned} (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + 2(1-\lambda)\lambda x_1 x_2 &< (1-\lambda)^2 x_1^2 + \lambda^2 x_2^2 + (1-\lambda)\lambda(x_1^2 + x_2^2) \\ &= (1 - 2\lambda - \lambda^2 + \lambda + \lambda^2)x_1^2 + (\lambda - \lambda^2 + \lambda^2)x_2^2 \\ &= (1-\lambda)x_1^2 + \lambda x_2^2 \\ &= (1-\lambda)f(x_1) + \lambda f(x_2) \end{aligned}$$

which proves strict convexity.

Example

Verify the $f(x, y) = x + y, \quad \forall x, y \in \mathbb{R}$ is convex or concave.

SOLUTION:

For any two points $A = (x_1, y_1)$, $B = (x_2, y_2)$ and pick $\lambda \in (0, 1)$ we have

$$
\begin{aligned}
f((1-\lambda)A + \lambda B) &= f\left((1-\lambda)\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \lambda \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) \\
&= f\begin{pmatrix} (1-\lambda)x_1 + \lambda x_2 \\ (1-\lambda)y_1 + \lambda y_2 \end{pmatrix} \\
&= \{(1-\lambda)x_1 + \lambda x_2\} + \{(1-\lambda)y_1 + \lambda y_2\} \\
&= (1-\lambda)(x_1 + x_2) + \lambda(y_1 + y_2) \\
&= (1-\lambda)f(A) + \lambda f(B)
\end{aligned}
$$

Equality implies the function is both convex and concave

UnConstrained Optimization

Definition

Let $f : \mathrm{I} \to \mathbb{R}$, I an interval. A point $x_0 \in I$ is a local maximum of f if there is a $\delta > 0$ such that $f(x) \leq f(x_0)$ whenever $x \in I \cap (x_0 - \delta, x_0 + \delta)$. Similarly, we can define local minimum.

Necessary Condition for local extrema:

First derivative test

Theorem

Suppose $f : [a, b] \to \mathbb{R}$ and suppose f has either a local maximum or a local minimum at $x_0 \in (a, b)$. If f is differentiable at x_0 then $f'(x_0) = 0$.

Proof: Necessary Condition for extrema

- Suppose f has a local maximum at $x_0 \in (a, b)$
- For small h we have $f(x_0 + h) \leq f(x_0)$.
- If $h > 0$ then
$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0$$
- If $h < 0$ then
$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0$$
- Given f is differentiable at x_0, hence the following
$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$
exist and unique at $x = x_0$
- From above two inequalities we have
$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = 0$$

Sufficient Condition: local maximization

Second derivative test

The sufficient conditions only for local maximum and the sufficient conditions for local minimum are similar. In the following results we assume $f : (a, b) \to \mathbb{R}$.

Theorem (A)

Let $c \in (a, b)$ and f be continuous at c. If for some $\delta > 0$, f is increasing on $(c - \delta, c)$ and decreasing on $(c, c + \delta)$, then f has a local maximum at c.

Theorem (B)

Let $c \in (a, b)$ and f be continuous at c. If $f'(x) \geq 0$ for all $x \in (c - \delta, c)$ and $f'(x) \leq 0$ for all $x \in (c, c + \delta)$ then f has a local maximum at c.

Theorem (C)

Let $c \in (a, b)$. If $f'(c) = 0$ and $f''(c) < 0$ then f has a local maximum at c.

Proof: Sufficient Condition for maximum

Theorem: A

- Choose any x_1 and x such that $c - \delta < x_1 < x < c$.
- Then $f(x_1) \leq f(x)$ and by the continuity of f at c we have

$$f(x_1) \leq \lim_{x \to c^-} f(x) = f(c)$$

- Similarly, if $c < x_2 < c + \delta$ then $f(x_2) \geq \lim_{x \to c^+} f(x) = f(c)$.

Theorem: C

Given: $f'(c) = 0$ & $f''(c) < 0$

Claim: f has a local maximum at c

$$\lim_{x \to c} \frac{f'(x)}{x - c} = \lim_{x \to c} \frac{f'(x) - f'(c)}{x - c} = f''(c) < 0$$

- $f'(x) > 0$ for $x \in (c - \delta, c)$, hence increasing.
- $f'(x) < 0$ for $x \in (c, c + \delta)$, hence decreasing.

Converse of theorem not true

- If f is continuous at c and f has a local maximum at c, then f need not be increasing on $(c-\delta, c)$ or decreasing on $(c, c+\delta)$ for any $\delta > 0$.

  Example:

  $$f(x) = -(x\sin(1/x))^2 \text{ if } x \neq 0 \text{ and } f(0) = 0 \text{ for } c = 0$$

- If f has a maximum at c and f is twice differentiable at c, then $f''(c)$ need not be less than 0.

  Example:

  $$f(x) = -x^4 \text{ for } c = 0$$

Convexity & Concavity of a Function

- If the first derivative of a function $f(x)$ at x is $f'(x_0)$.
- Convexity and Concavity: The smile test for maximum/minimum ☹ or ☺.

- If $f''(x) < 0$ for all x, then strictly concave.
  Critical points are global maxima ☹

- If $f''(x) > 0$ for all x, then strictly convex.
  Critical points are global minima ☺

N^{th} Derivative test:

- If the first nonzero derivative value at x_0 encountered in successive derivation is that of the N^{th} derivative, $f^{(n)}(x) \neq 0$,
- Then the stationary value $f(x_0)$ will be:

  1. A relative max if N is even and $f^{(n)}(x_0) < 0$
  2. A relative min if N is even and $f^{(n)}(x_0) > 0$
  3. An inflection point if N is odd

Taylor series analysis 1D

x is for all possible points in the neighbourhood of x_0.

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

Necessary condition $f'(x_0) = 0$

$$f(x) - f(x_0) \approx \frac{f''(x_0)}{2!}(x - x_0)^2$$

- If x_0 is a local minima: $f(x) - f(x_0) > 0 \Rightarrow f''(x_0) > 0$
- If x_0 is a local maxima: $f(x) - f(x_0) < 0 \Rightarrow f''(x_0) < 0$

Taylor series analysis 2D

x is for all possible points in the neighbourhood of x_0.

$$
\begin{aligned}
f(x,y) - f(x_0,y_0) \approx \; & \frac{f_x(x_0,y_0)}{1!}(x-x_0) + \frac{f_y(x_0,y_0)}{1!}(y-y_0) \\
& + \frac{f_{xx}(x_0,y_0)}{2!}(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \frac{f_{yy}(x_0,y_0)}{2!}(y-y_0)^2 \\
= \; & h^T \begin{pmatrix} f_x \\ f_y \end{pmatrix}_{(x_0,y_0)} + \frac{1}{2} h^T H h, \text{ for } h = \begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix} \text{ and } H = \begin{pmatrix} f_{xx} & f_{yx} \\ f_{xy} & f_{yy} \end{pmatrix}_{(x_0,y_0)} \qquad (1)
\end{aligned}
$$

Necessary condition for Local Extrema: $f_x(x_0,y_0) = 0$ and $f_y(x_0,y_0) = 0$

$$f(x,y) - f(x_0,y_0) \approx \frac{1}{2} h^T H h$$

- If (x_0,y_0) is a local minima: $f(x,y) - f(x_0,y_0) > 0 \Rightarrow \frac{1}{2} h^T H h > 0 \quad \forall h \neq 0$
- If (x_0,y_0) is a local maxima: $f(x,y) - f(x_0,y_0) < 0 \Rightarrow \frac{1}{2} h^T H h < 0 \quad \forall h \neq 0$

A real symmetric matrix is positive definite iff all its eigenvalue are positive

- ($\Longrightarrow$):
  - Let $\lambda \in \mathbb{R}$ be an eigenvalue of $A \in \mathbb{R}^{n \times n}$, and $x \in \mathbb{R}^n$ be the corresponding eigenvector, i.e.,

$$Ax = \lambda x. \tag{2}$$

  - Also given that A is positive definite, i.e., $x^T Ax > 0, \forall x \in \mathbb{R}^n$.
  - **Claim:** All eigenvalues are A, is positive, i.e., $\lambda_i > 0$.
  - Multiplying x^T both sides of (2), we get

$$x^T Ax = \lambda x^T x = \lambda \|x\|^2.$$

  - From the above, the left side is positive and $\|x\|^2$ is positive. Hence λ is real-positive.
- ($\Longleftarrow$)
  - Assume that all eigenvalues are positive, i.e., $\lambda_i > 0$.
  - **Claim:** The matrix A is positive definite.i.e., $x^T Ax > 0, \forall x \in \mathbb{R}^n$
  - We know, the real symmetric matrix is diagonalizable by an orthogonal matrix. So there exists an orthogonal matrix Q, such that $Q^T AQ = D$.
  - Here $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix, whose all diagonal entries are positive real no.

A real symmetric matrix is positive definite iff all its eigenvalue are positive

-
  - Let $x \in \mathbb{R}^n$ be any nonzero vector.
  - Now, $x^T A x = x^T Q D Q^T x$. Putting $y = Q^T x$, we get

$$x^T A x = y^T D y.$$

  - Let $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, then we have

$$\begin{aligned} x^T A x &= y^T D y \\ &= \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & \\ 0 & \lambda_2 & 0 & \cdots \\ \vdots & & & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2 > 0. \end{aligned}$$

  - Since x is a nonzero vector and Q is invertible, $y = Q^T x$ is not a zero vector. Therefore A is positive definite.