

Machine Learning

CMPT 726

Mo Chen
SFU School of Computing Science
2021-09-22

Linear Algebra and Calculus Review (cont'd)

Convexity/Concavity

Convex function:

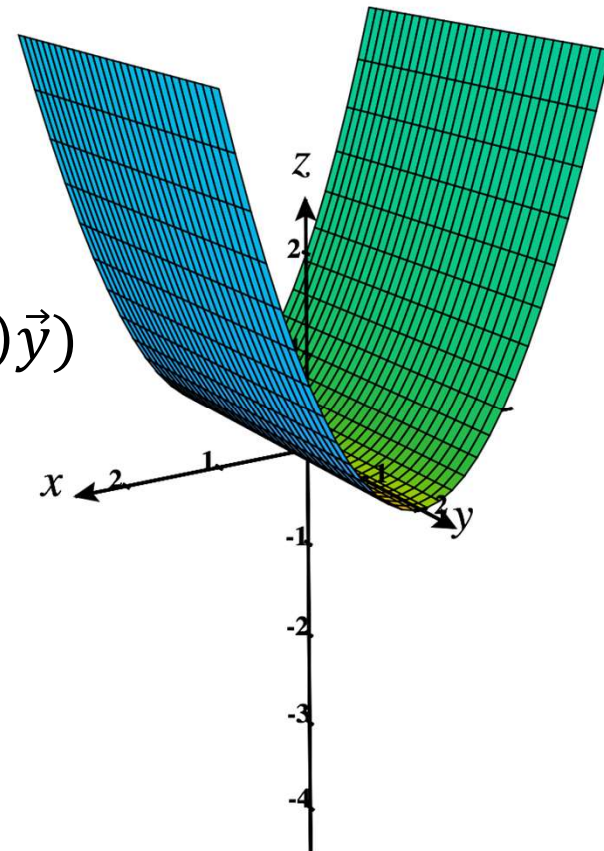
A line segment between **any** two points on the surface lies **on or above** the surface:

$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) \geq f(\alpha\vec{x} + (1 - \alpha)\vec{y})$$

where $0 \leq \alpha \leq 1$

Or equivalently: Hessian of function is **positive semi-definite** everywhere, e.g.:

$$f(\vec{x}) = \vec{x}^\top \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \vec{x} = x_1^2$$



Convexity/Concavity

Strictly convex function:

A line segment between **any** two points on the surface lies **above** the surface:

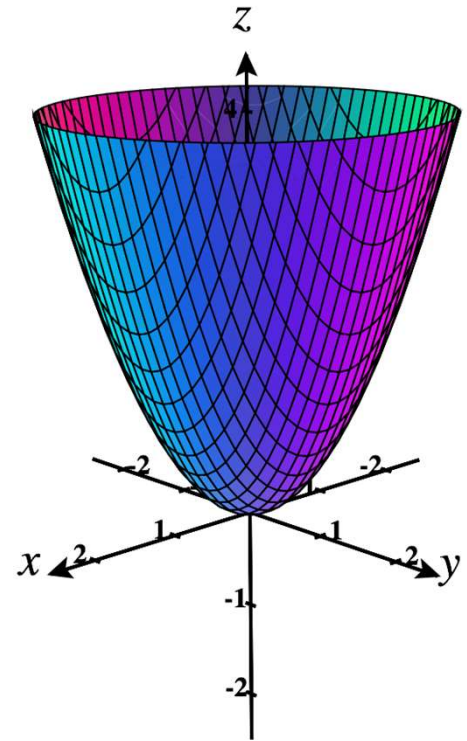
$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) > f(\alpha\vec{x} + (1 - \alpha)\vec{y})$$

where $0 < \alpha < 1$

If the Hessian is **positive definite** everywhere, e.g.:

$$f(\vec{x}) = \vec{x}^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \vec{x} = x_1^2 + x_2^2$$

Then the function is strictly convex (but not the other way around!)



Convexity/Concavity

Concave function:

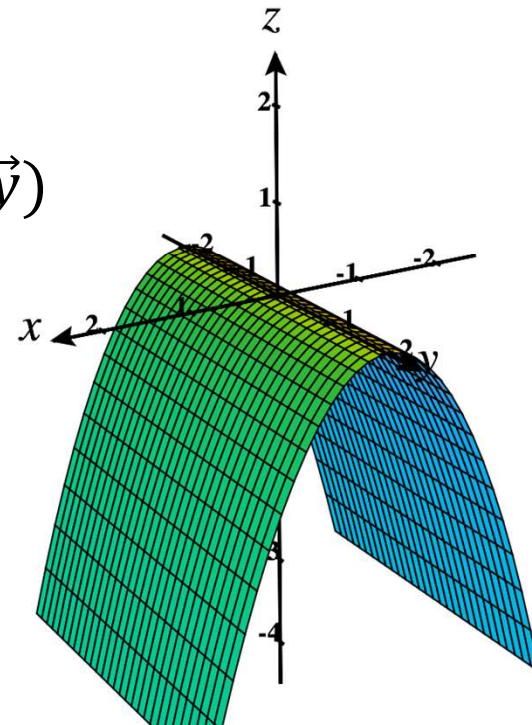
A line segment between **any** two points on the surface lies **on or below** the surface:

$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) \leq f(\alpha\vec{x} + (1 - \alpha)\vec{y})$$

where $0 \leq \alpha \leq 1$

Or equivalently: Hessian of function is **negative semi-definite** everywhere, e.g.:

$$f(\vec{x}) = \vec{x}^\top \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \vec{x} = -x_1^2$$



Convexity/Concavity

Strictly concave function:

A line segment between **any** two points on the surface lies **below** the surface:

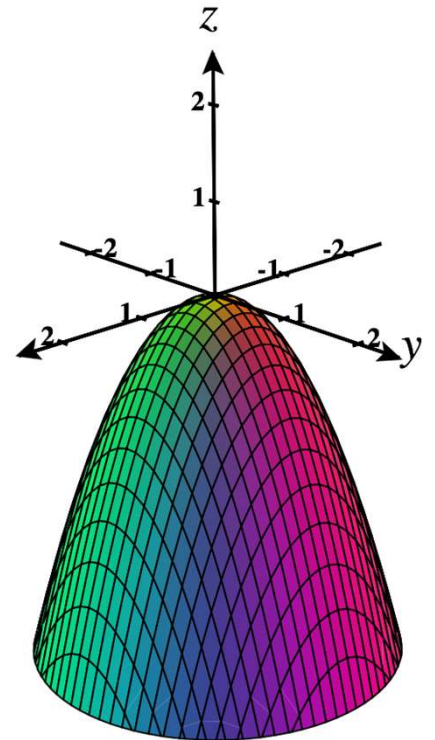
$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) < f(\alpha\vec{x} + (1 - \alpha)\vec{y})$$

where $0 < \alpha < 1$

If the Hessian is **negative definite** everywhere, e.g.:

$$f(\vec{x}) = \vec{x}^\top \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \vec{x} = -x_1^2 - x_2^2$$

Then the function is strictly concave (but not the other way around!)



Jensen's Inequality

Definition of convex functions:

$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) \geq f(\alpha\vec{x} + (1 - \alpha)\vec{y}), \text{ where } 0 \leq \alpha \leq 1$$

Jensen's inequality generalizes this to convex combinations of many points:

$$\sum_{i=1}^n \alpha_i f(\vec{x}_i) \geq f\left(\sum_{i=1}^n \alpha_i \vec{x}_i\right), \text{ where } \alpha_i \geq 0 \forall i, \text{ and } \sum_{i=1}^n \alpha_i = 1$$

Jensen's Inequality

Definition of concave functions:

$$\alpha f(\vec{x}) + (1 - \alpha)f(\vec{y}) \leq f(\alpha\vec{x} + (1 - \alpha)\vec{y}), \text{ where } 0 \leq \alpha \leq 1$$

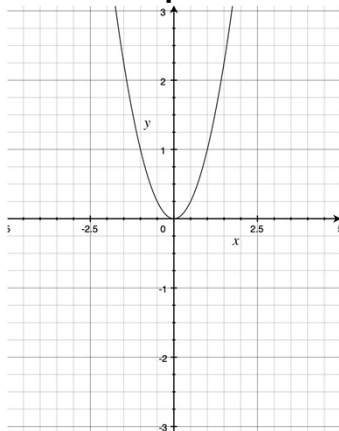
Jensen's inequality generalizes this to convex combinations of many points:

$$\sum_{i=1}^n \alpha_i f(\vec{x}_i) \leq f\left(\sum_{i=1}^n \alpha_i \vec{x}_i\right), \text{ where } \alpha_i \geq 0 \forall i, \text{ and } \sum_{i=1}^n \alpha_i = 1$$

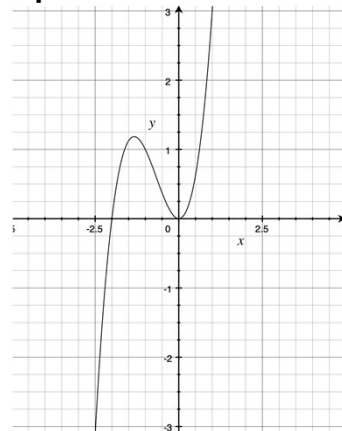
Optimality Conditions

Consider a univariate function $f: \mathbb{R} \rightarrow \mathbb{R}$ that is everywhere twice differentiable.
How can we find the points where f is minimized or maximized?

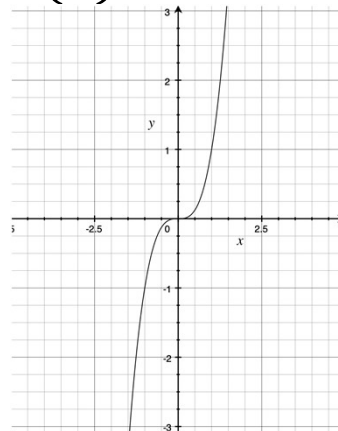
Critical points: All points x where $f'(x) = 0$.



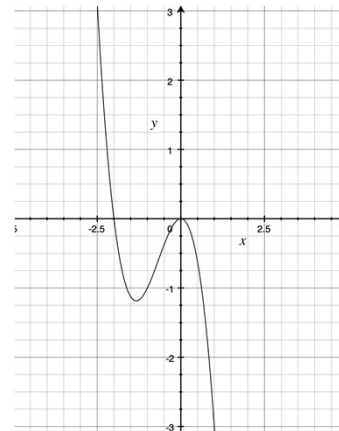
Global minimum



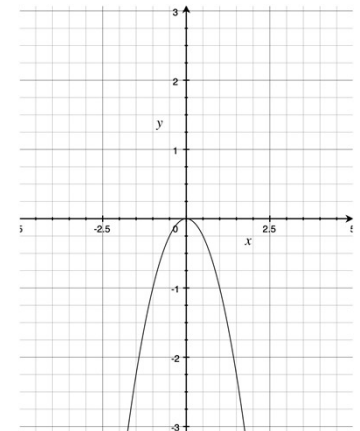
Local minimum



Saddle point



Local maximum



Global maximum

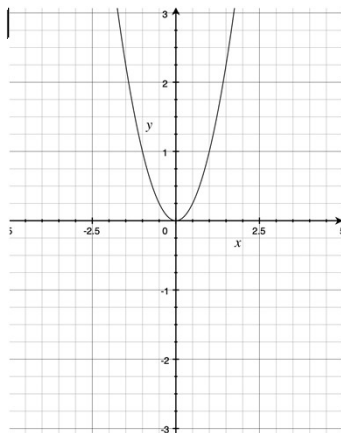
Optimality Conditions

To differentiate these cases, we can check the second derivative at the critical points.

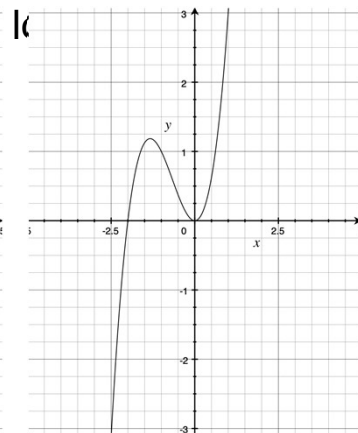
If $f''(x) > 0$, local/global minimum

If $f''(x) < 0$, local/global maximum

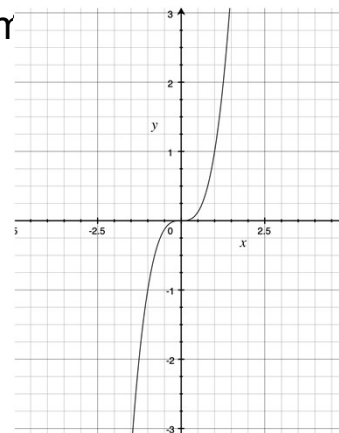
If $f''(x) = 0$, could be a local/global minimum, a local/global maximum, or a saddle point



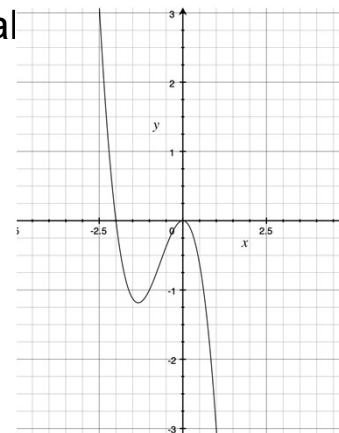
Global minimum



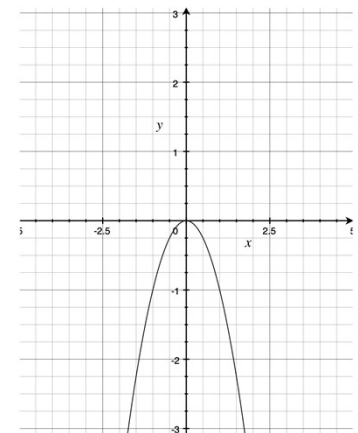
Local minimum



Saddle point



Local maximum

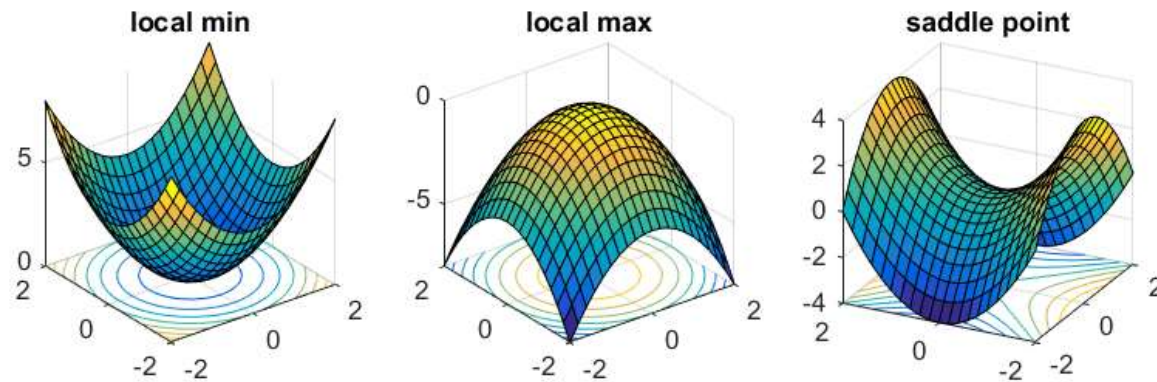


Global maximum

Optimality Conditions

Consider a multivariate function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that is everywhere twice differentiable. How can we find the points where f is minimized or maximized?

Critical points: All points \vec{x} where $\nabla f(\vec{x}) = \vec{0}$.



Optimality Conditions

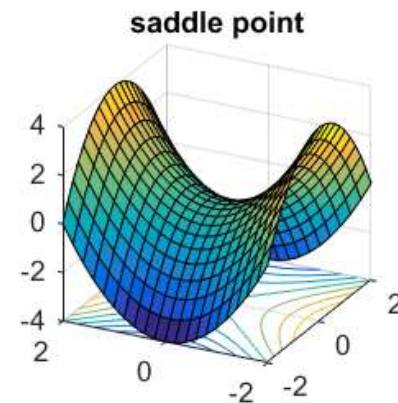
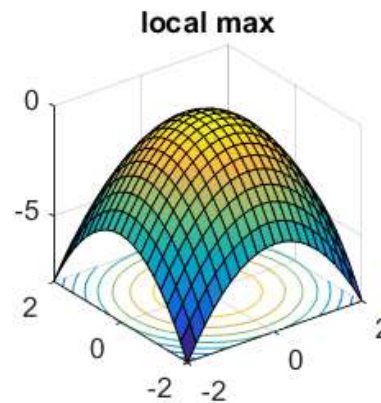
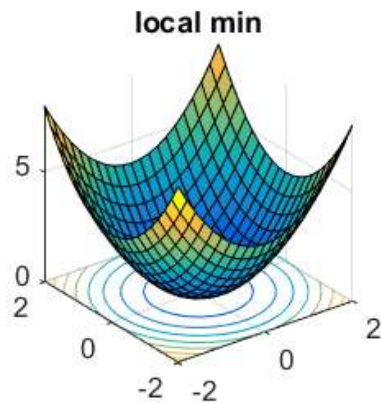
To differentiate these cases, we can check the Hessian evaluated at the critical points.

If $Hf(\vec{x}) \succ 0$ (positive definite), local/global minimum

If $Hf(\vec{x}) \prec 0$ (negative definite), local/global maximum

If $Hf(\vec{x})$ is indefinite, saddle point

If $Hf(\vec{x}) \succeq 0$ (positive semi-definite) or $Hf(\vec{x}) \preceq 0$ (negative semi-definite), could be a local/global minimum, a local/global maximum, or a saddle point



Global Optimality and Convexity/Concavity

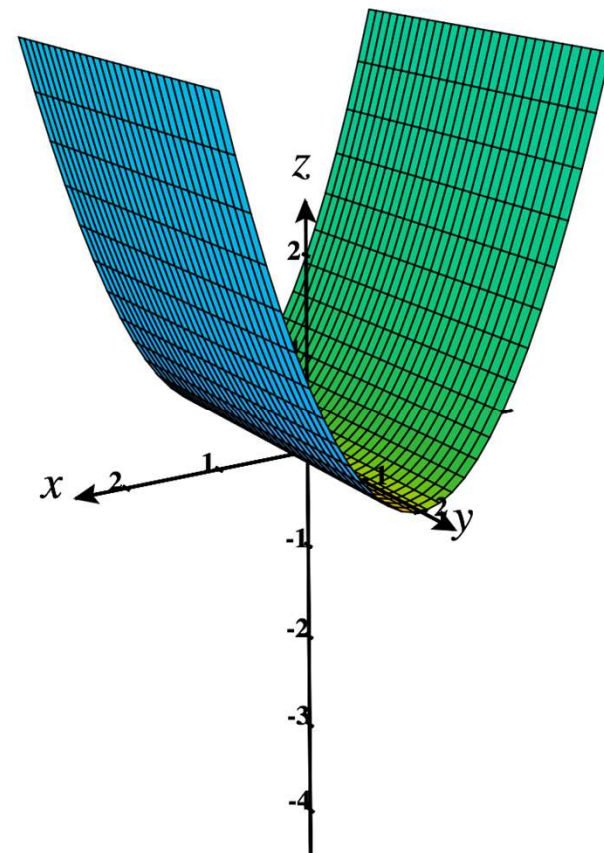
Recall: it is hard to distinguish local optima (that are not global optima) from global optima.

Convexity/concavity is a sufficient (but not necessary) condition for every local optimum to be a global optimum.

In a **convex** function:

Every **critical point** is a **local minimum**.

Every **local minimum** is a **global minimum**.



Global Optimality and Convexity/Concavity

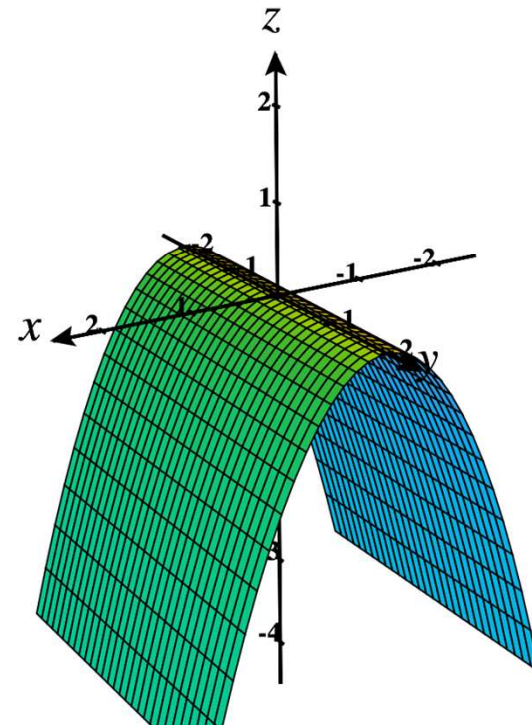
Recall: it is hard to distinguish local optima (that are not global optima) from global optima.

Convexity/concavity is a sufficient (but not necessary) condition for every local optimum to be a global optimum.

In a **concave** function:

Every **critical point** is a **local maximum**.

Every **local maximum** is a **global maximum**.

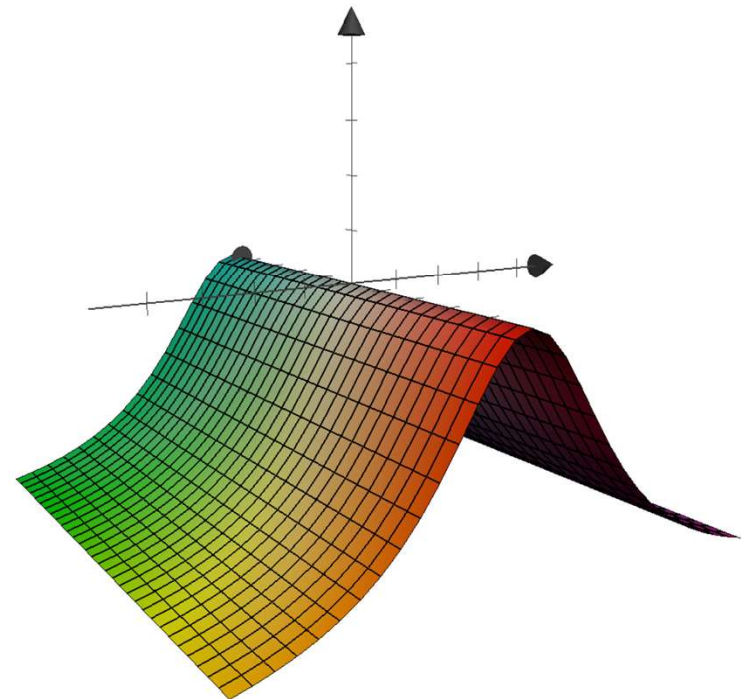


Global Optimality and Convexity/Concavity

Recall: it is hard to distinguish local optima (that are not global optima) from global optima.

Convexity/concavity is a sufficient (but not necessary) condition for every local optimum to be a global optimum.

Note: There are **non-concave** functions where every **local maximum** is a **global maximum**.



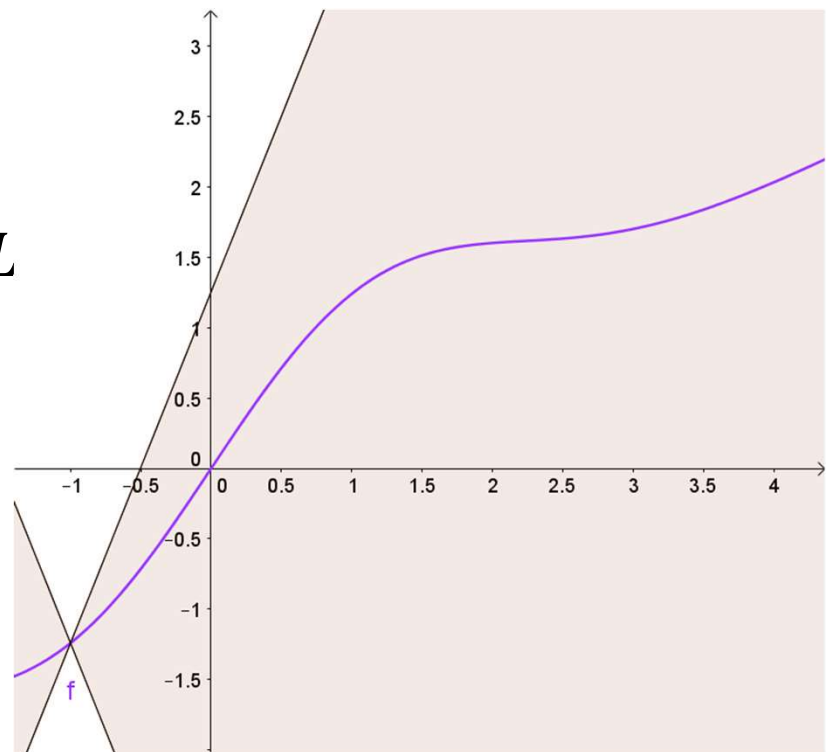
Lipschitz Continuity

A function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is L -Lipschitz if for all \vec{x}_1, \vec{x}_2 ,

$$|f(\vec{x}_1) - f(\vec{x}_2)| \leq L$$

Intuitively, an L -Lipschitz function cannot grow too quickly.

An everywhere differentiable function is L -Lipschitz if and only if $\|\nabla f(\vec{x})\| \leq L$



Quiz Practice

Q1: Which of the following is **always** a valid mathematical operation? (An operation that results in complex numbers is not considered valid.)

- (A) Computing A^{-1} , where A is full-rank but non-square
- (B) Computing $\sqrt{\vec{x}^\top A^\top A \vec{x}}$, where A is non-square
- (C) Computing $\sqrt{\vec{x}^\top A \vec{x}}$, where A is symmetric
- (D) Computing the eigendecomposition of A , where A is non-square
- (E) Computing $\frac{A}{B}$, where A and B are positive definite
- (F) None of the above
- (G) Don't know