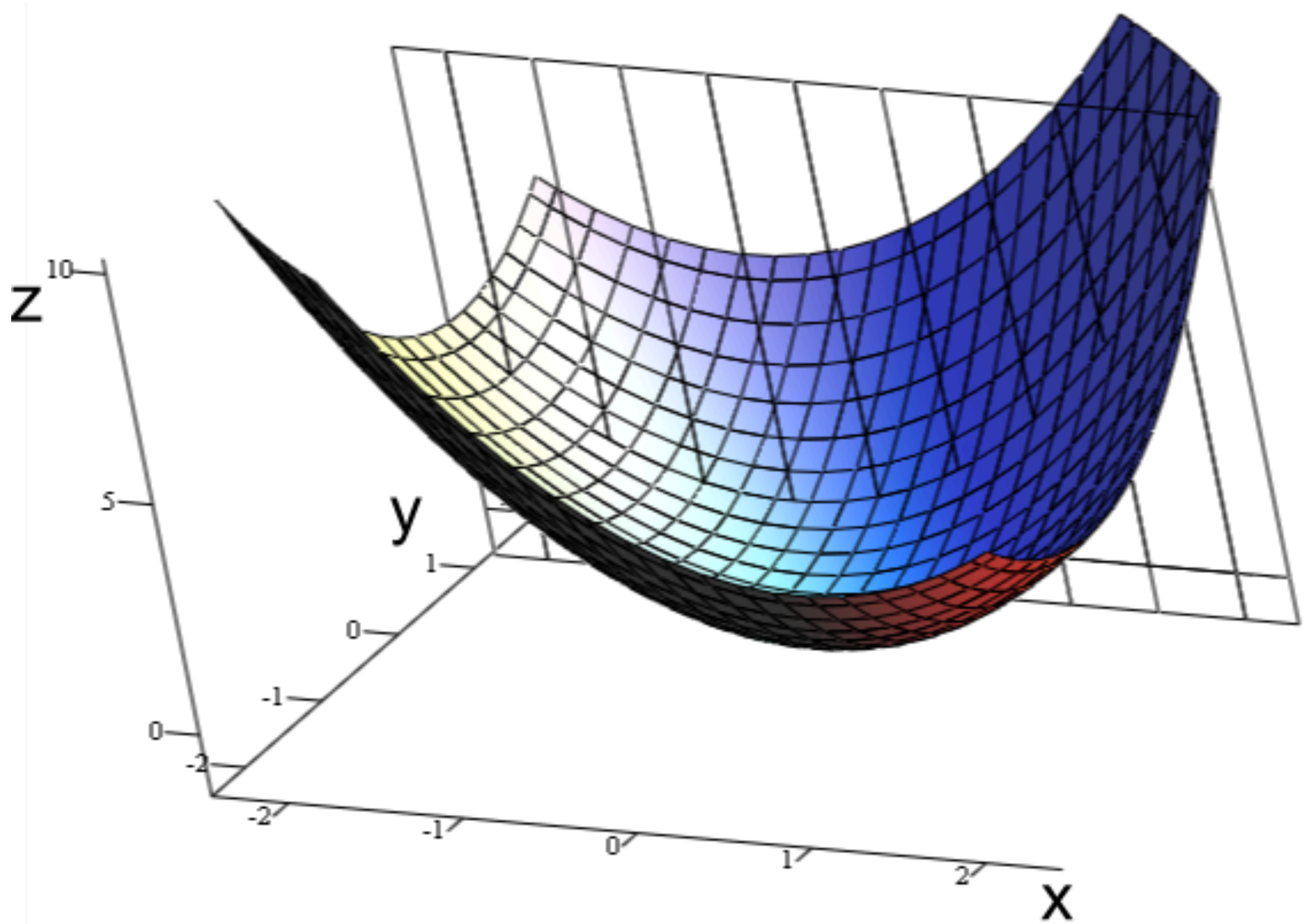


Convexity

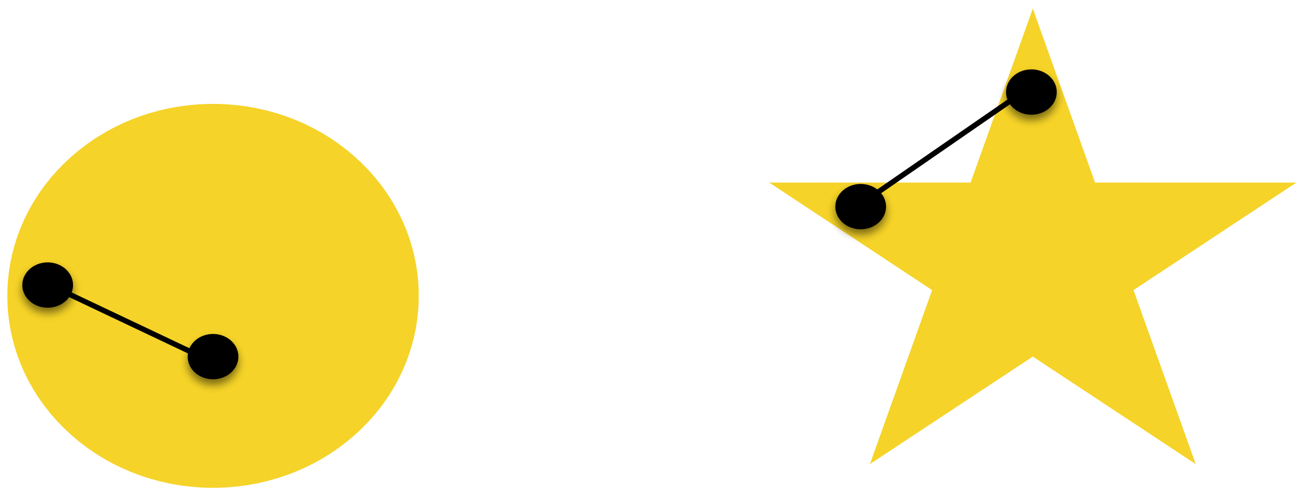
Leandro L. Minku

Convex Sets

A set C is convex if the line segment between any two points in C lies in C.

For any two points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in C$ and any $\lambda \in (0,1)$, we have:

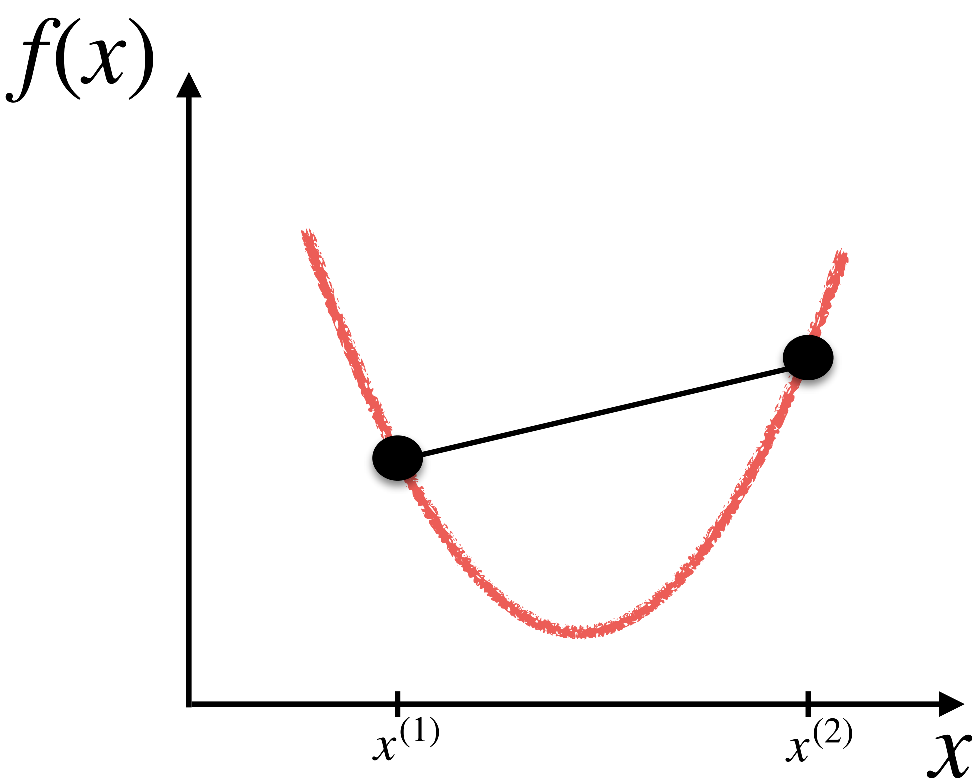
$$\lambda \mathbf{x}^{(1)} + (1 - \lambda)\mathbf{x}^{(2)} \in C$$



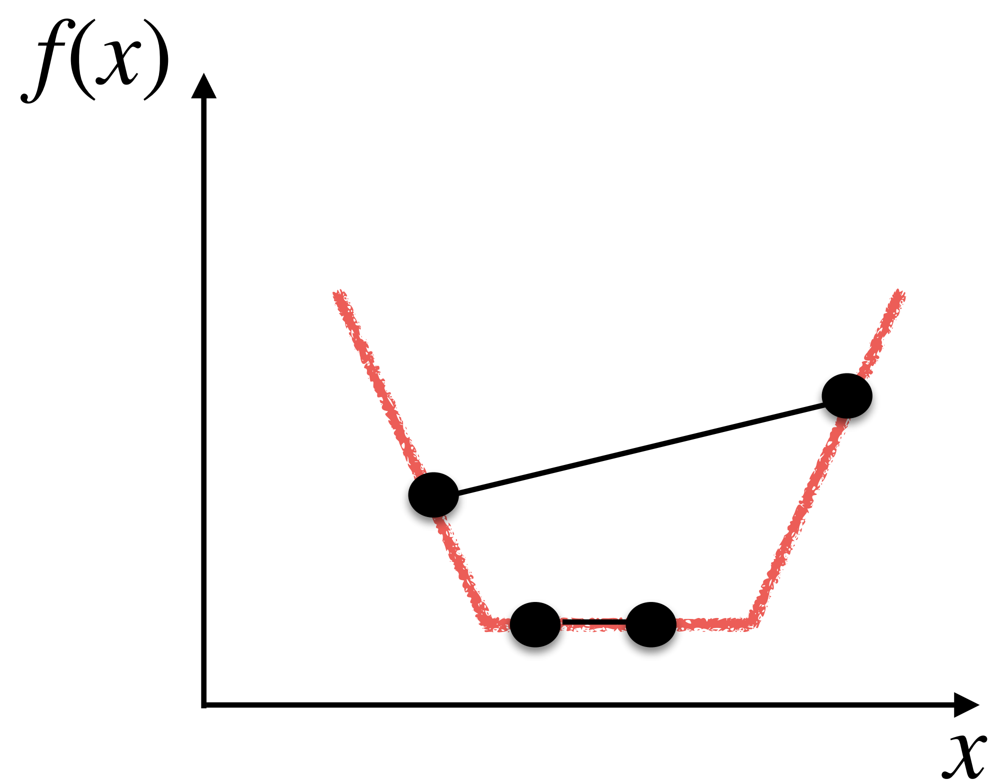
Convex Functions

A convex function $f(\mathbf{x})$ is a function with a convex domain C that satisfies the following condition for any $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in C$ and $\lambda \in (0,1)$:

$$f(\lambda \mathbf{x}^{(1)} + (1-\lambda)\mathbf{x}^{(2)}) \leq \lambda f(\mathbf{x}^{(1)}) + (1-\lambda) f(\mathbf{x}^{(2)})$$



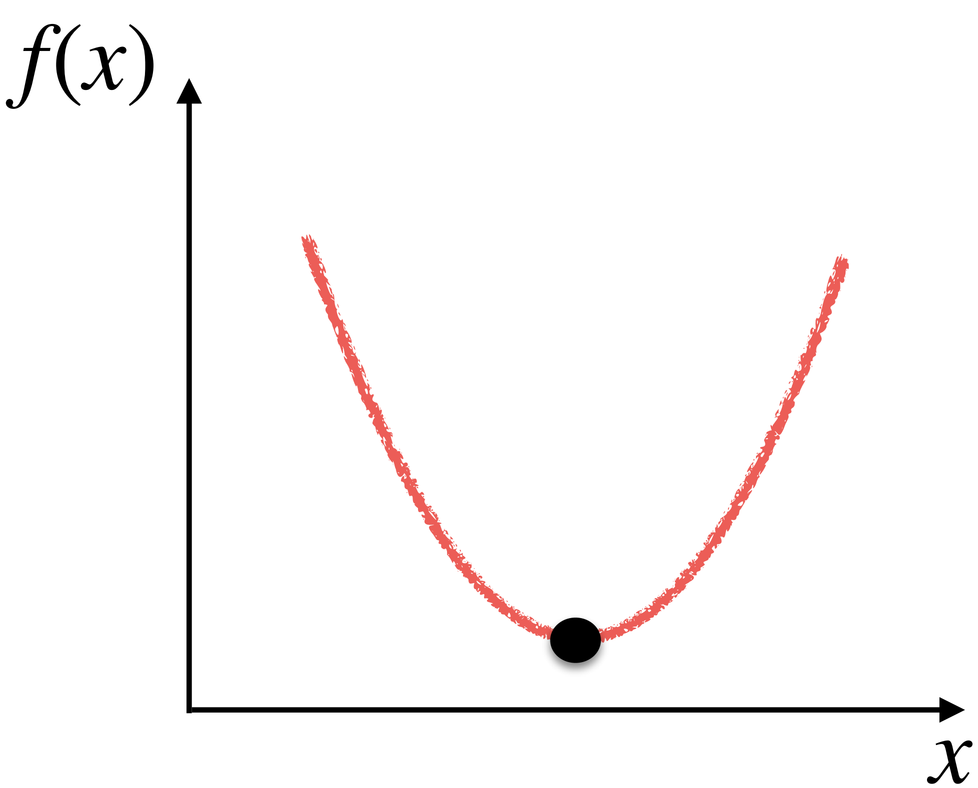
Strictly convex:
satisfies the condition with $<$ for any $\mathbf{x}^{(1)} \neq \mathbf{x}^{(2)}$

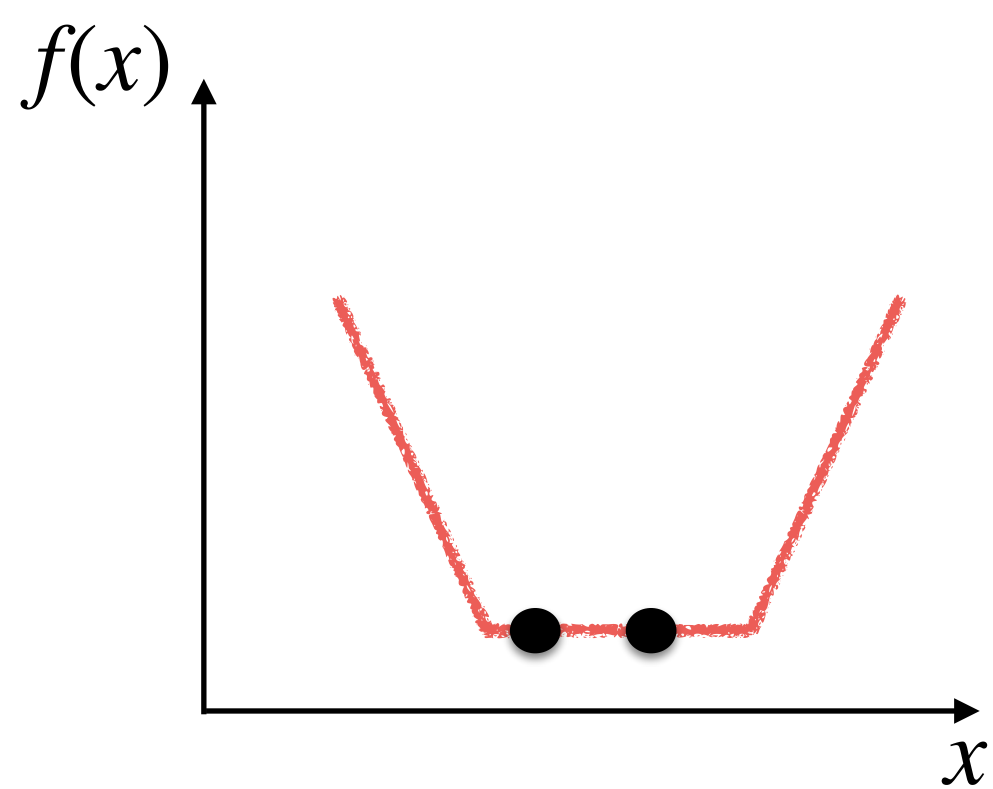


Importance of Convexity in Machine Learning / Optimisation

Any minimum in a convex function is a global minimum.

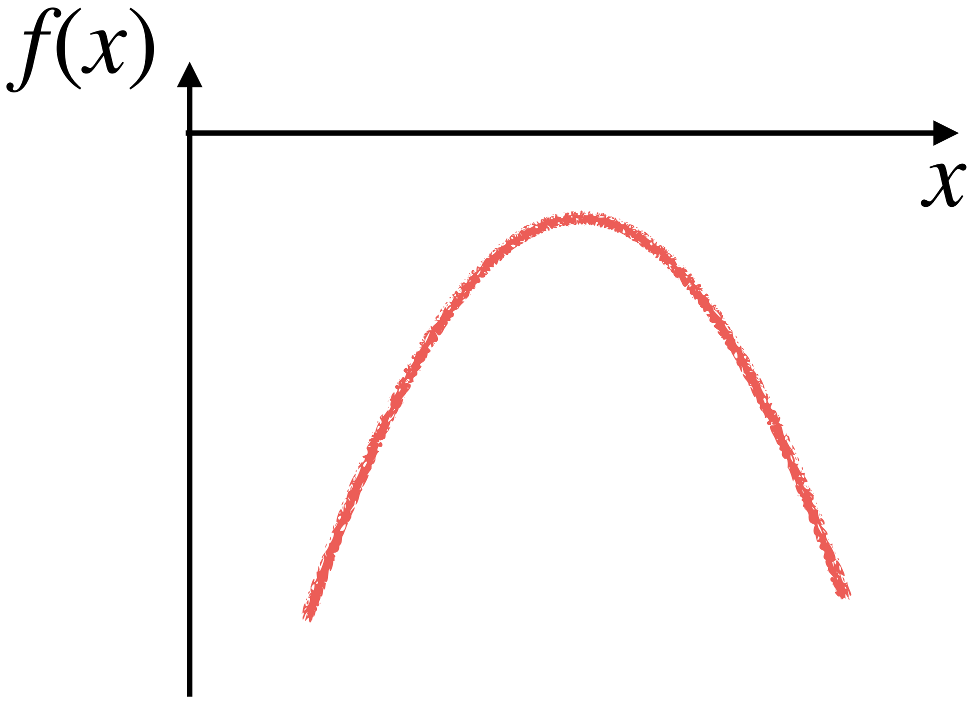
A strictly convex function has at most one stationary (critical) point. If such a point exists, it is a global minimum.

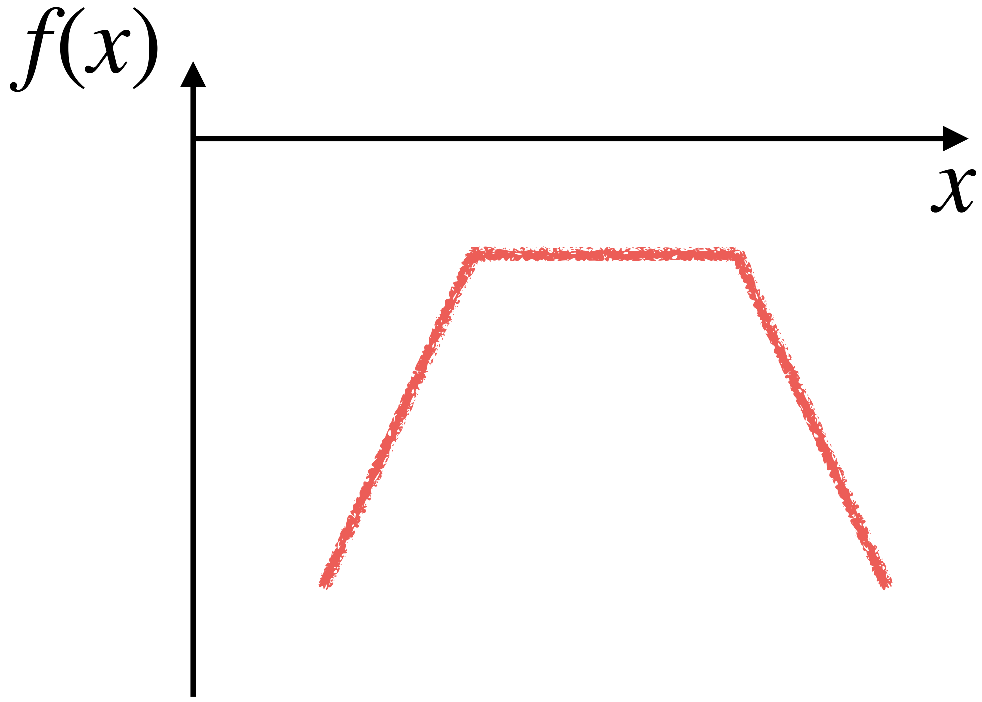




Concave

A function $f(\mathbf{x})$ is concave if $-f(\mathbf{x})$ is convex.

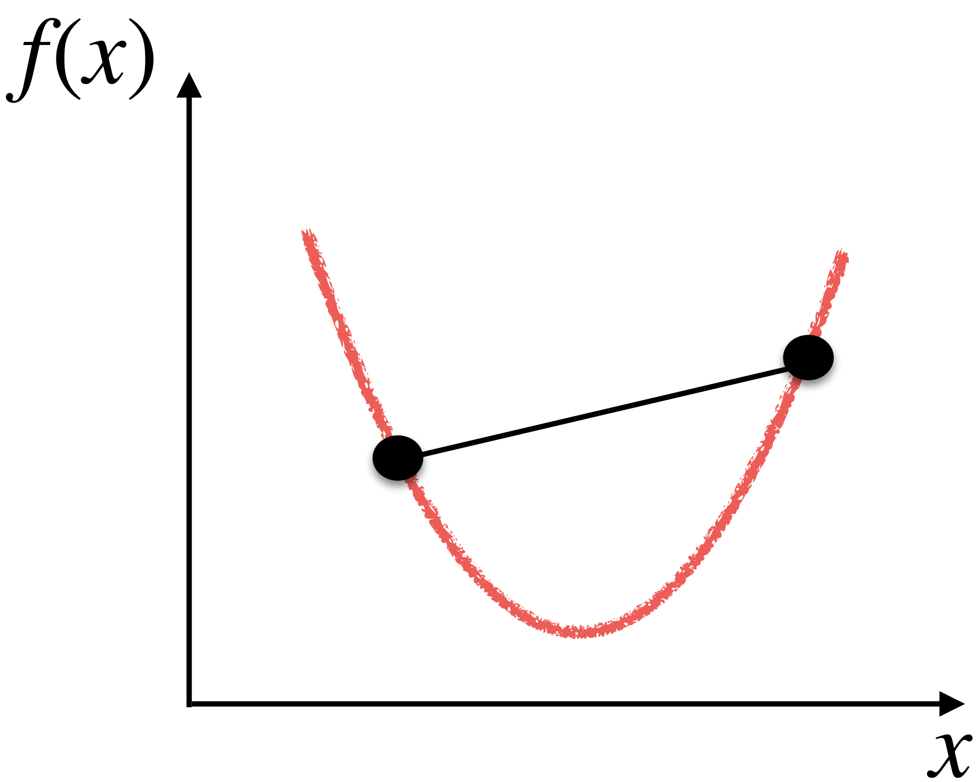




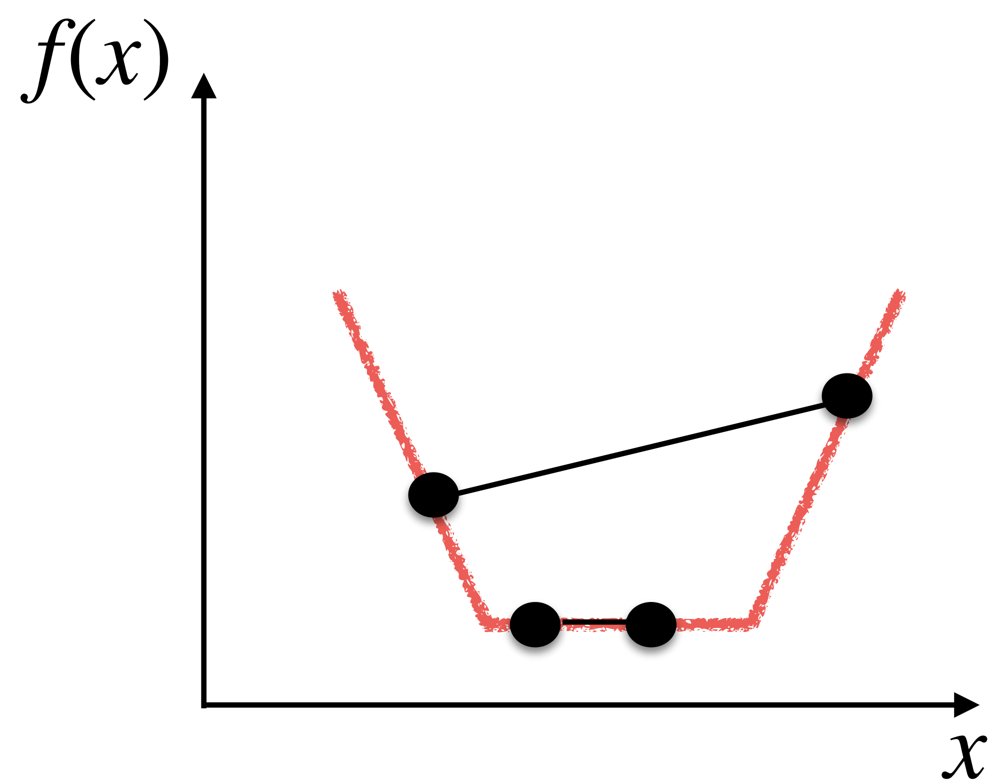
Convex Functions

A convex function $f(\mathbf{x})$ is a function with a convex domain C that satisfies the following condition for any $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in C$ and $\lambda \in (0,1)$:

$$f(\lambda \mathbf{x}^{(1)} + (1-\lambda)\mathbf{x}^{(2)}) \leq \lambda f(\mathbf{x}^{(1)}) + (1-\lambda) f(\mathbf{x}^{(2)})$$



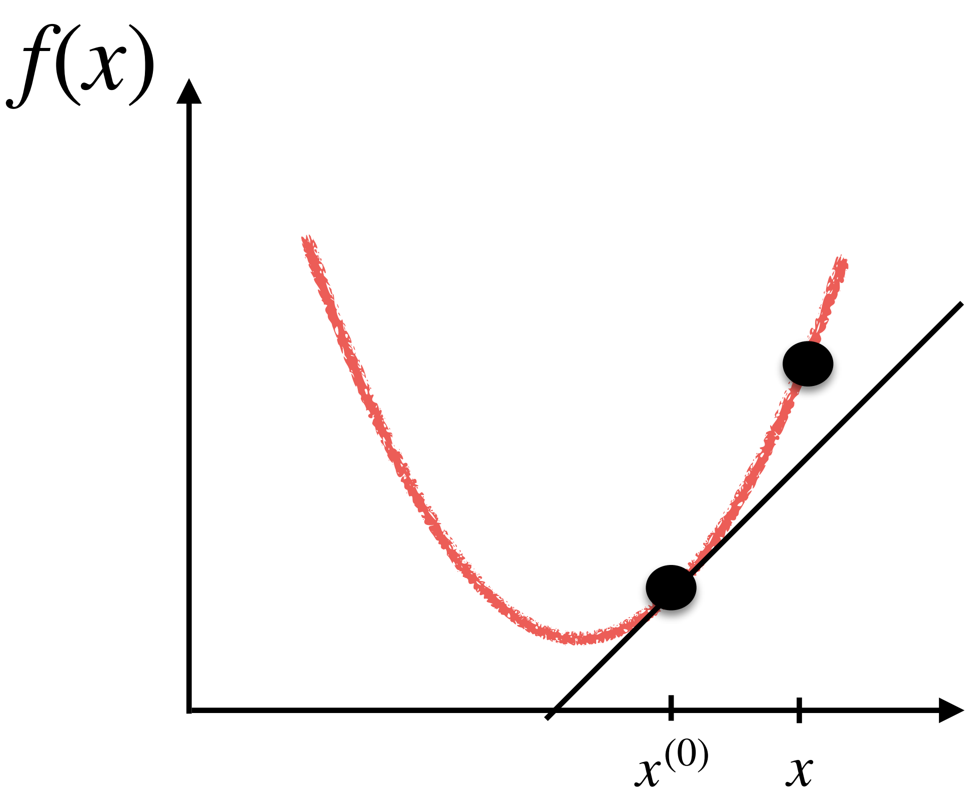
Strictly convex:
satisfies the condition with <



First-Derivative Characterisation of Convexity

A differentiable function $f(\mathbf{x})$ is convex *iif* its domain C is convex and it satisfies the following condition for any pair $\mathbf{x}^{(0)}, \mathbf{x} \in C$:

$$f(\mathbf{x}) \geq \underbrace{f(\mathbf{x}^{(0)}) + \nabla f(\mathbf{x}^{(0)}) \cdot (\mathbf{x} - \mathbf{x}^{(0)})}_{\text{Equation of the tangent line}}$$



A convex function always lies entirely on or above any tangent to the surface.

Strictly convex:
satisfies the condition with > for any $\mathbf{x}^{(1)} \neq \mathbf{x}^{(2)}$

Second-Derivative Characterisation of Convexity

A twice differentiable function $f(\mathbf{x})$ is convex *iif*:

- its domain C is a convex set and
- its Hessian $H_f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in C$.

If a twice differentiable function $f(\mathbf{x})$:

- has a convex set C as its domain and
- its Hessian $H_f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in C$

it is a strictly convex function. (sufficient but not necessary condition)

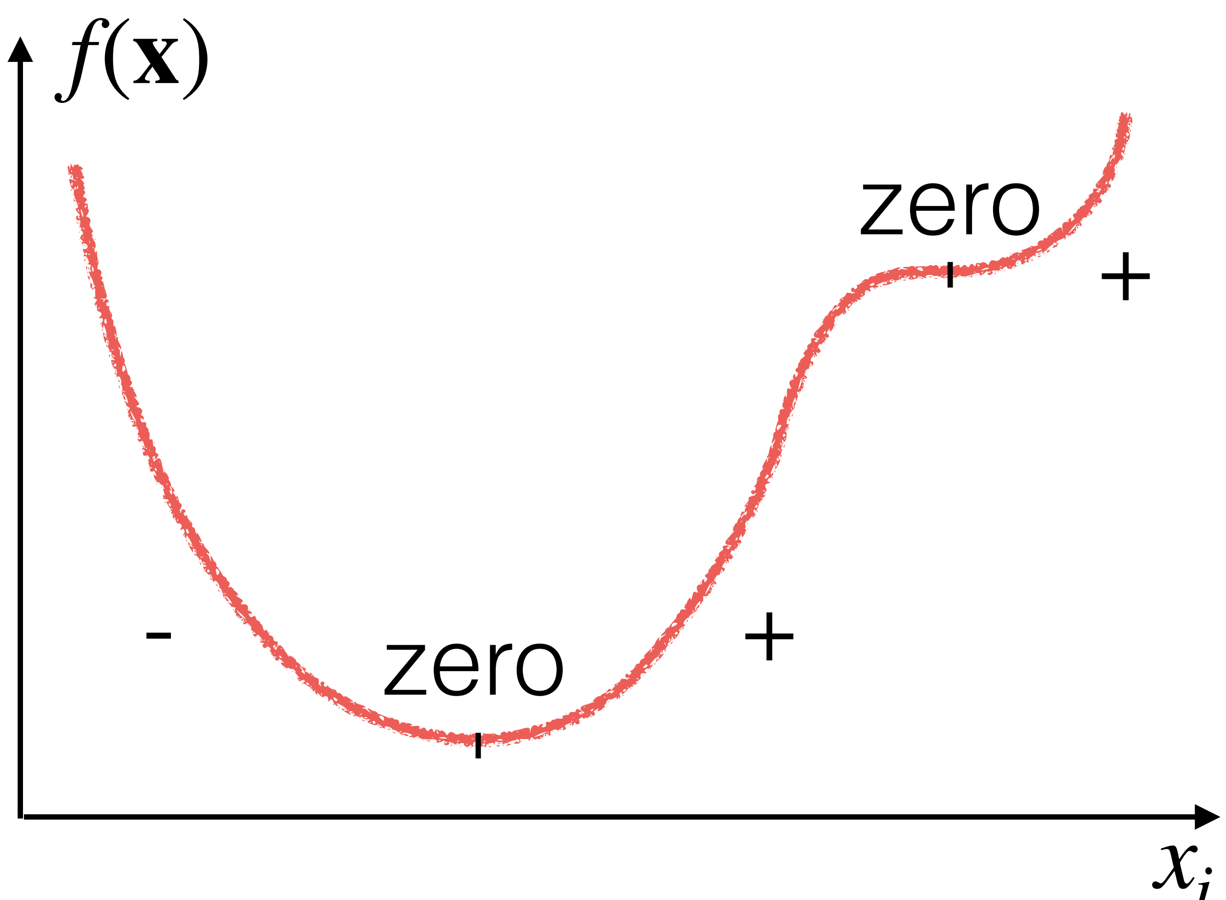
First-Order (Partial) Derivatives

$$\frac{d}{dx}f(x) = \frac{df}{dx} = f'(x) = f^{(1)}(x)$$

(First-order) derivatives tell us the rate of change of $f(x)$ as we increase x.

$$\frac{\partial f}{\partial x_i}$$

(First-order) partial derivatives tell us the rate of change of $f(\mathbf{x})$ as we increase a specific variable x_i.



(Partial) derivatives tell us whether $f(\mathbf{x})$ is increasing / decreasing (along a specific axis) and how rapidly.

Second-Order (Partial) Derivatives

$$\frac{d^2}{dx^2} f(x) = \frac{d}{dx}\left(\frac{df}{dx}\right) = \frac{d^2 f}{dx^2} = f''(x) = f^{(2)}$$

Second-order derivative is the derivative of the derivative of $f(x)$, i.e., gives the rate of change of the slope $f'(x)$.

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right), \quad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

Second-order partial derivative is the partial derivative of the partial derivative of $f(\mathbf{x})$, i.e., gives the rate of change of the slope along a given axis, with respect to the same or another axis.

You can create even higher order derivatives using the same idea.

Hessian — Matrix of Second-Order Partial Derivatives

Consider $f(\mathbf{x})$, where $\mathbf{x} = (x_0, x_1, \cdots, x_d)^T$

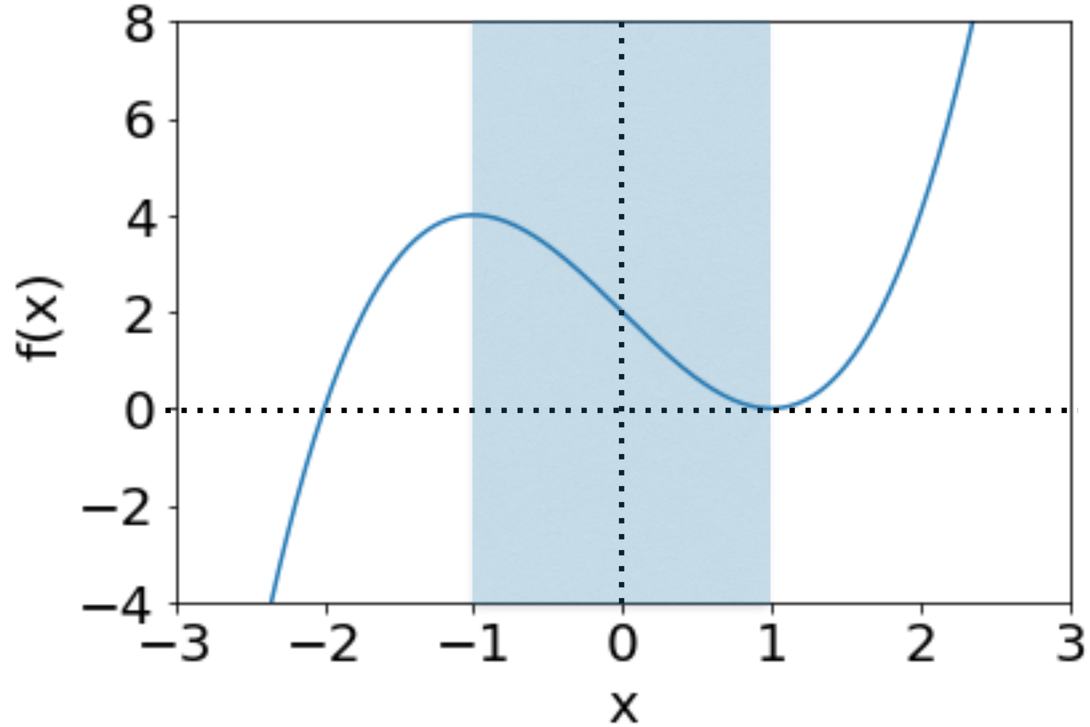
$$H(f(\mathbf{x})) = H_f(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_0^2}, & \dfrac{\partial^2 f}{\partial x_0 \partial x_1}, & \cdots, & \dfrac{\partial^2 f}{\partial x_0 \partial x_d} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_0}, & \dfrac{\partial^2 f}{\partial x_1^2}, & \cdots, & \dfrac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_d \partial x_0}, & \dfrac{\partial^2 f}{\partial x_d \partial x_1}, & \cdots, & \dfrac{\partial^2 f}{\partial x_d^2} \end{pmatrix}$$

Second Derivative Characterisation of Convexity:

Univariate Case

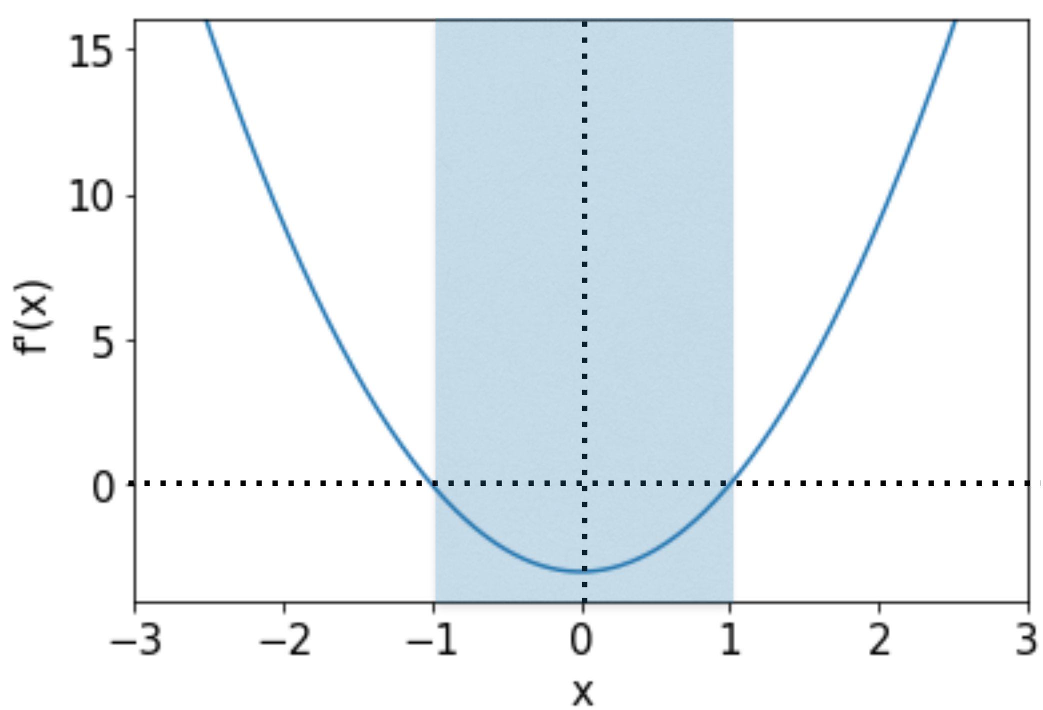
Second-Order Derivatives

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



decreasing

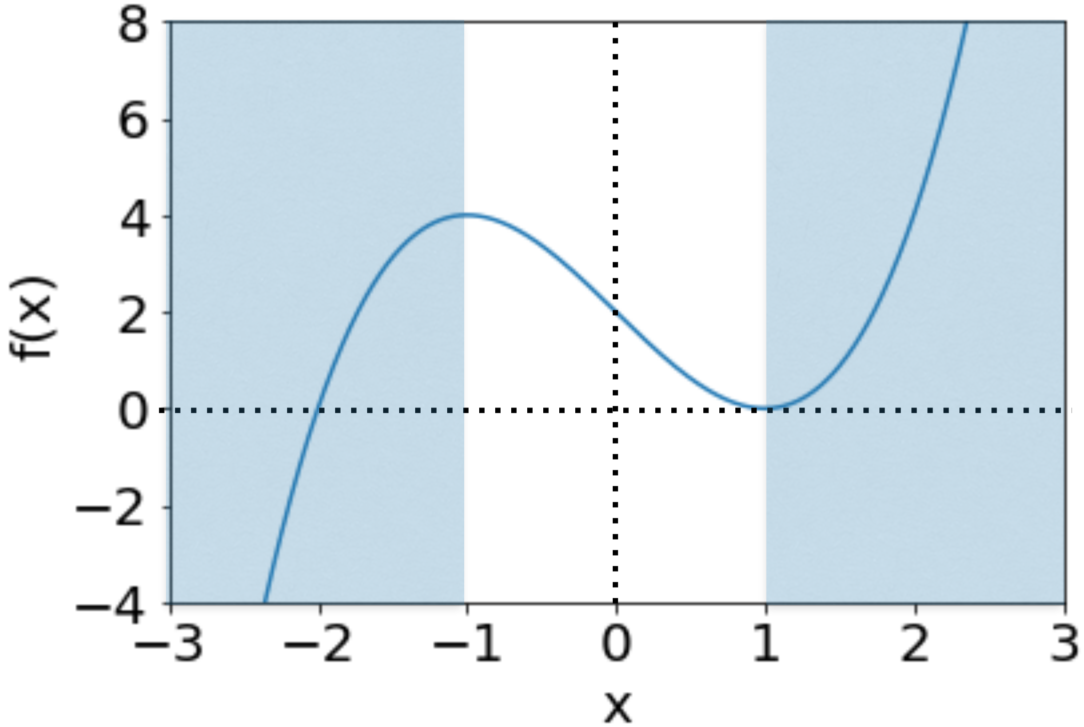
$$f'(x) = 3x^2 - 3$$



-

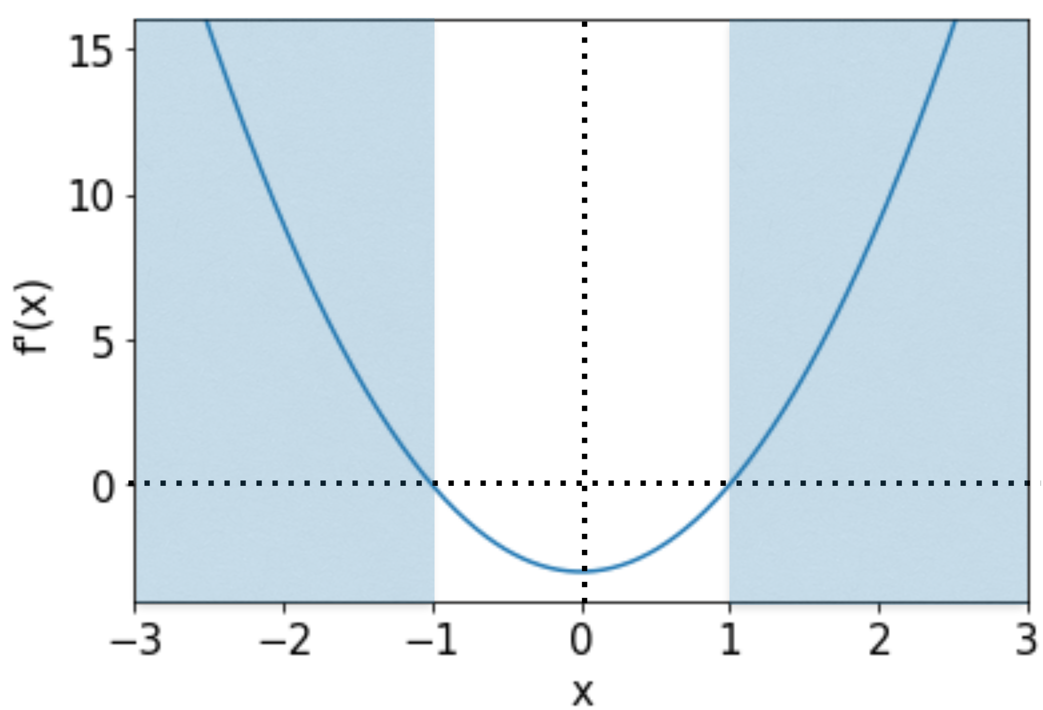
Second-Order Derivatives

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



increasing

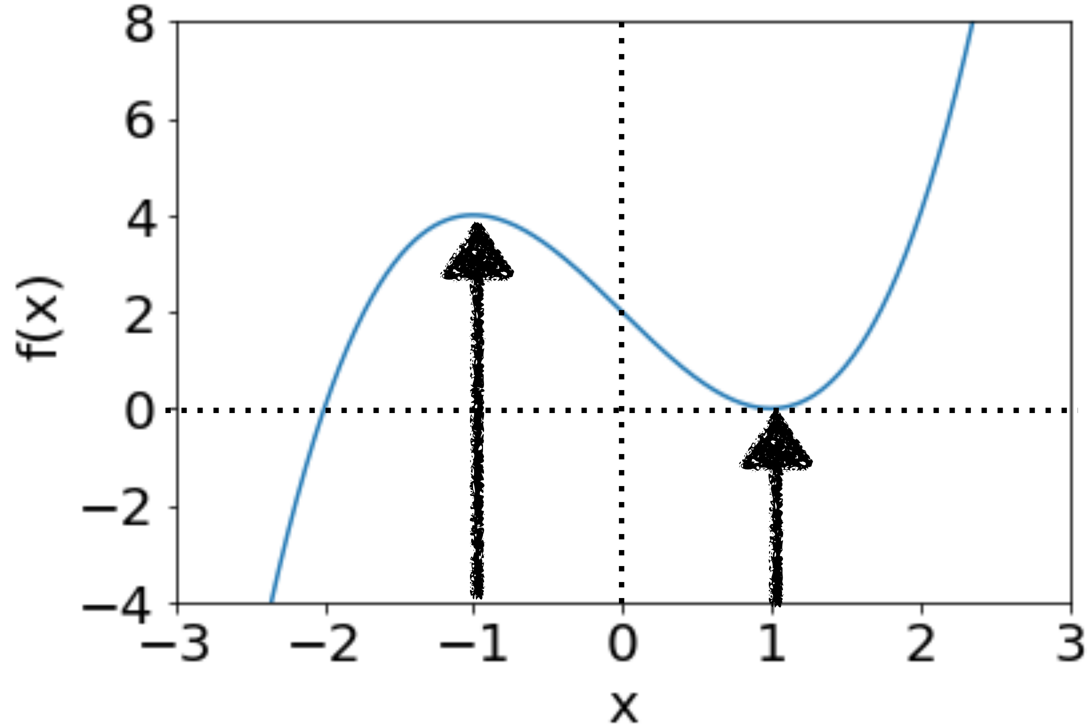
$$f'(x) = 3x^2 - 3$$



+

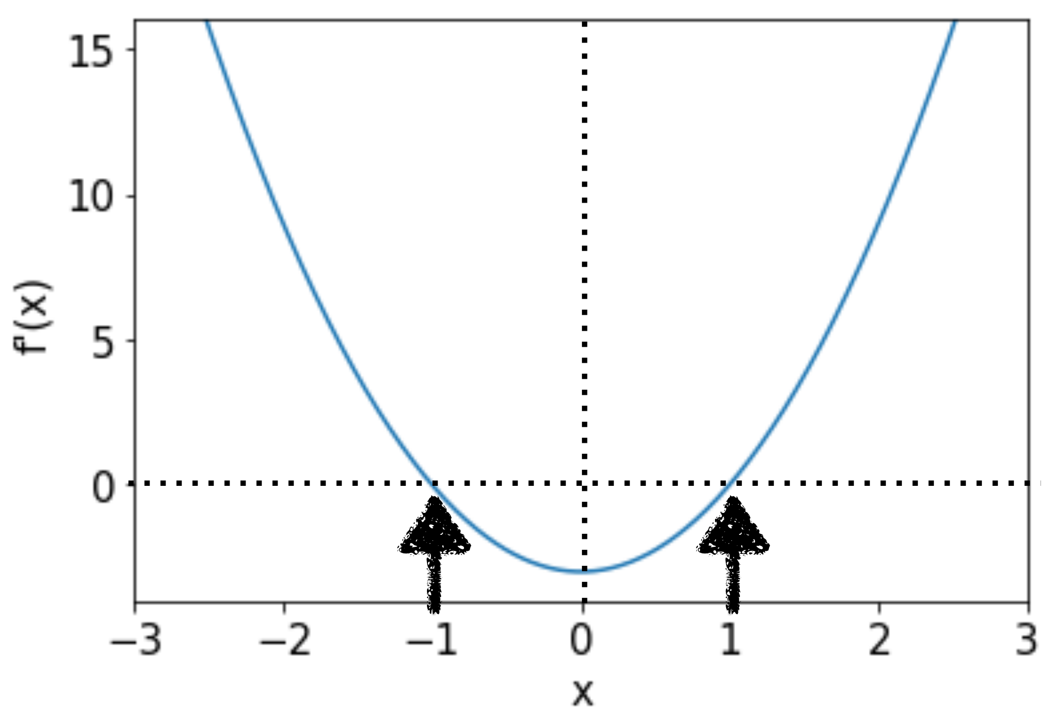
Second-Order Derivatives

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



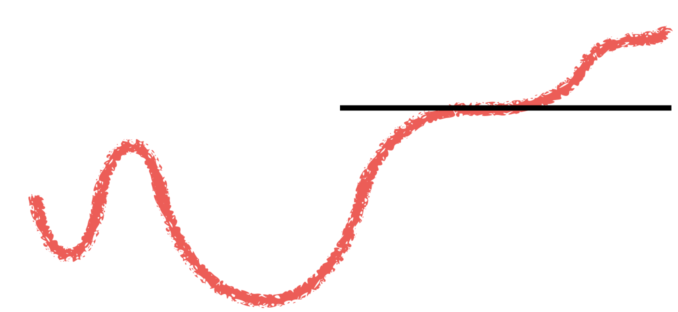
stationary

$$f'(x) = 3x^2 - 3$$



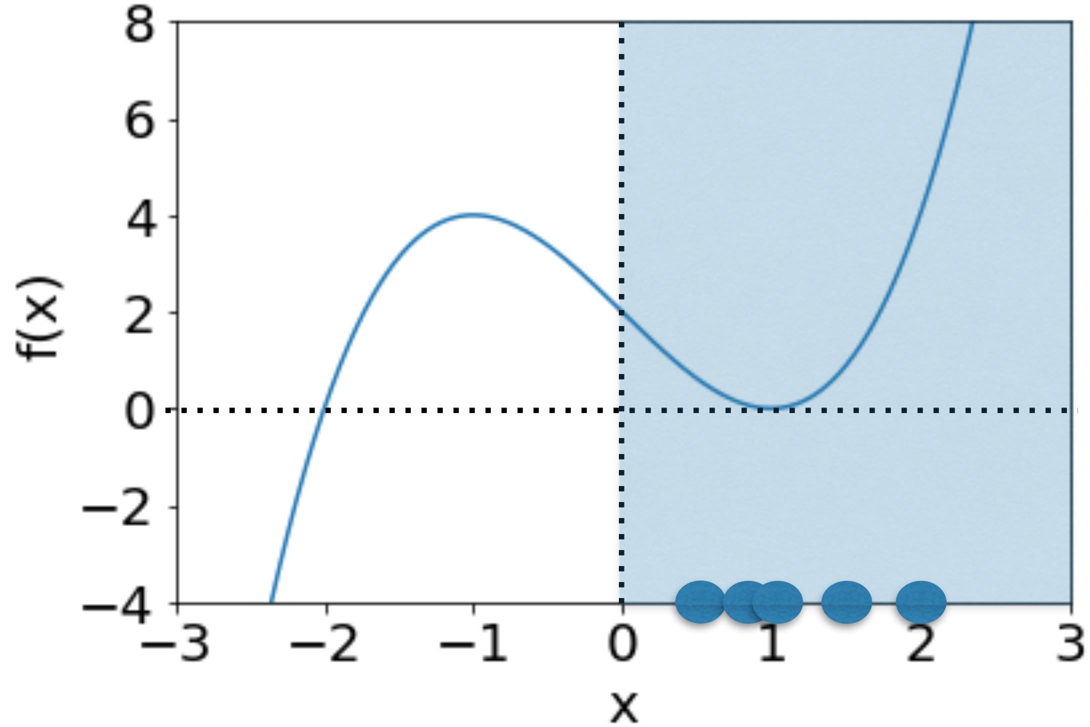
zero

Could be a minimum or maximum, but not necessarily.



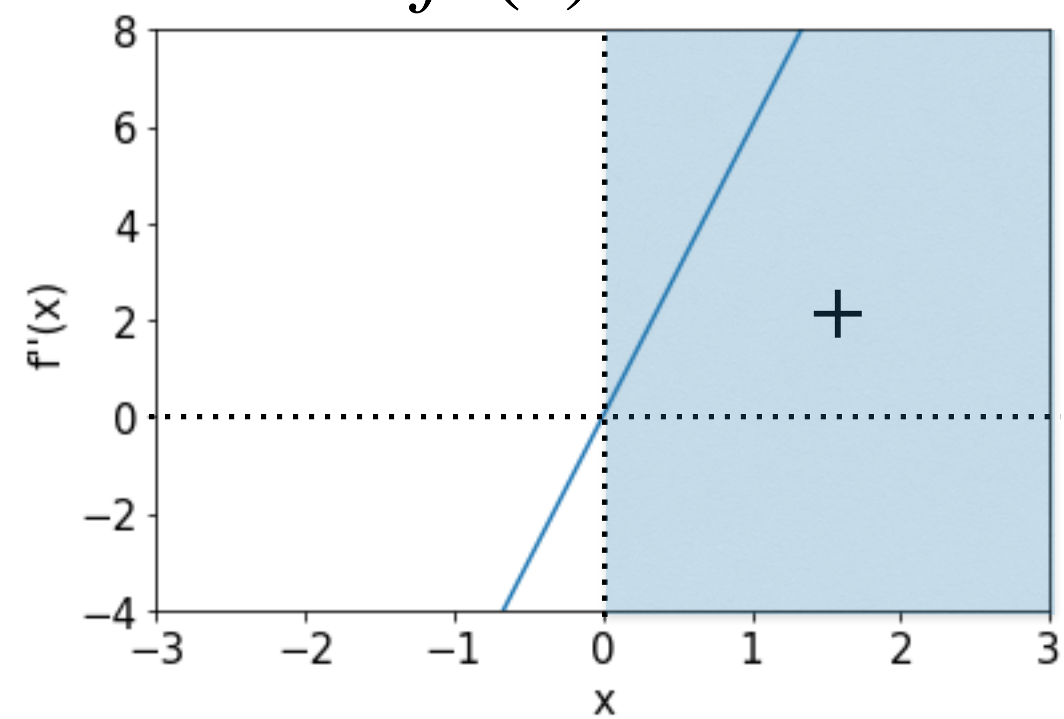
Second-Order Derivatives

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

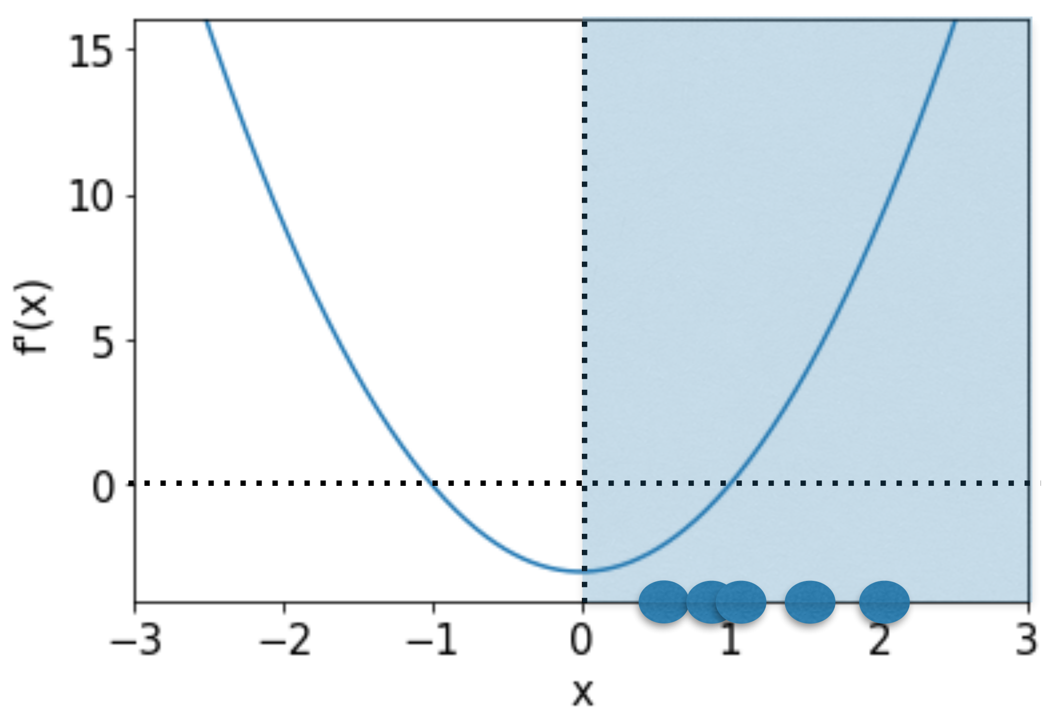


convex region

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



$$f'(x) = 3x^2 - 3$$



increasing

$f'(2) = 9$

$f'(1.5) = 3.5$

$f'(1) = 0$

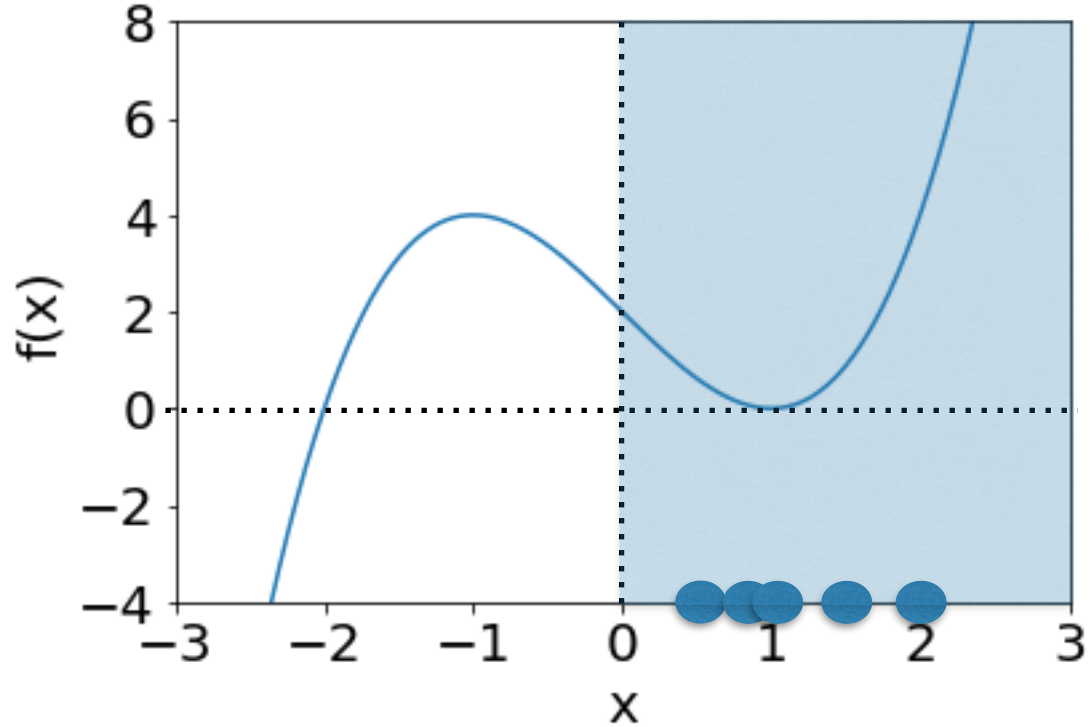
$f'(0.9) = -0.57$

$f'(0.5) = -2.25$

Slope was downwards, negative. While negative, it got less and less steep, increasing to zero. Then, got steeper and steeper, more and more positive.

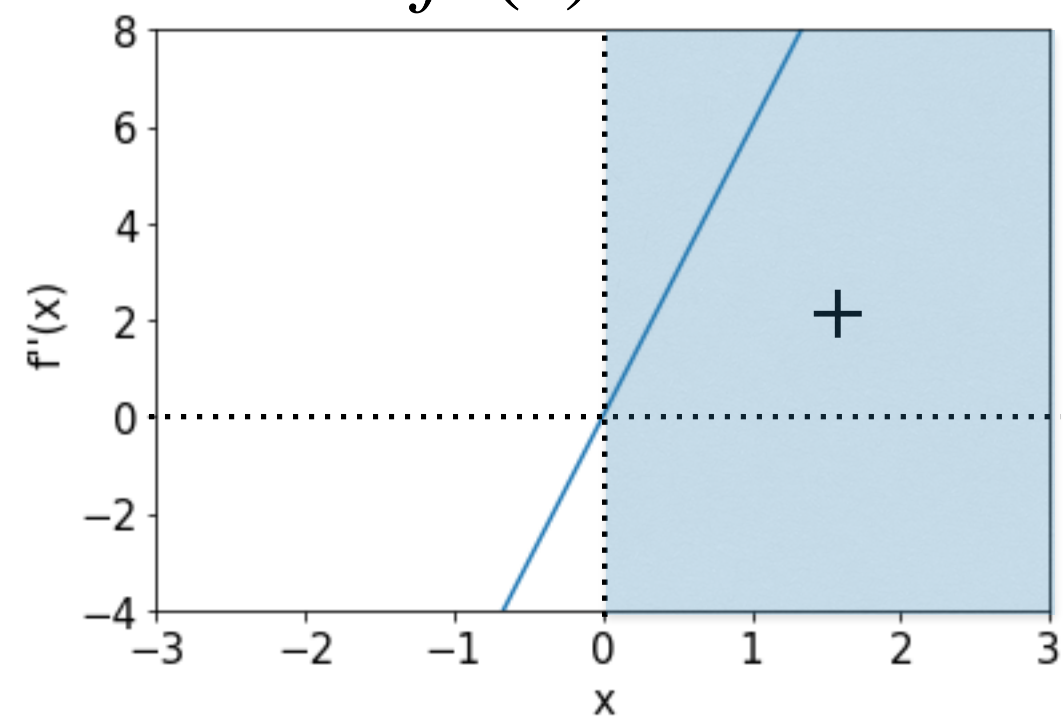
Second-Order Derivatives

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

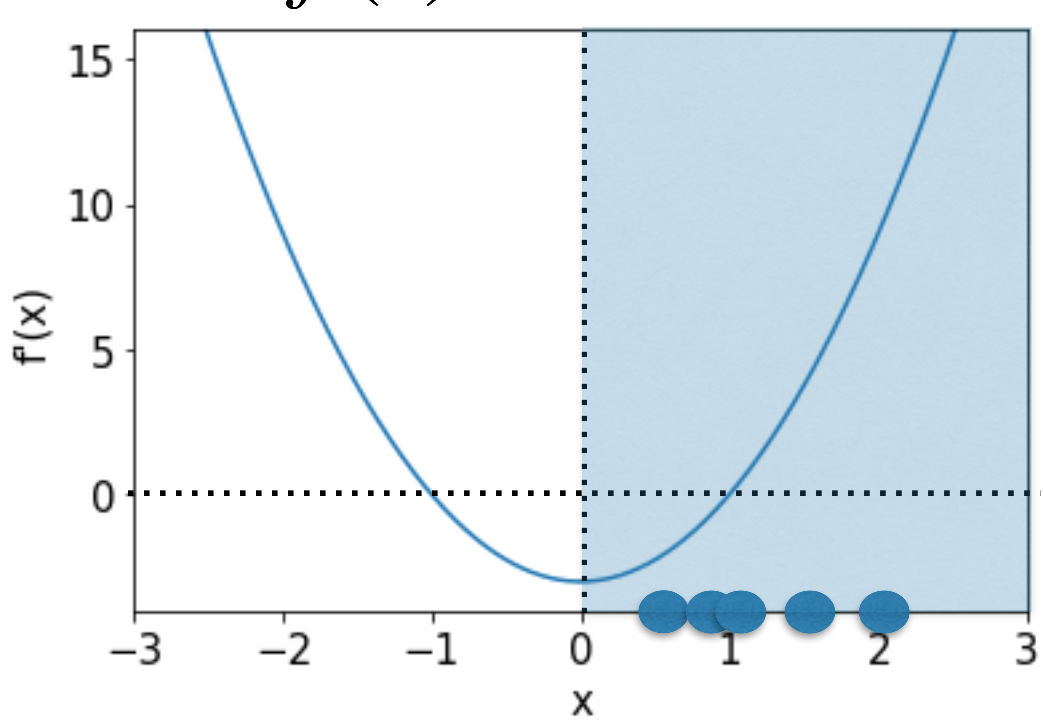


convex region

$f''(x) = 6x$



$f'(x) = 3x^2 - 3$



increasing

$f'(2) = 9$

$f'(1.5) = 3.5$

$f'(1) = 0$

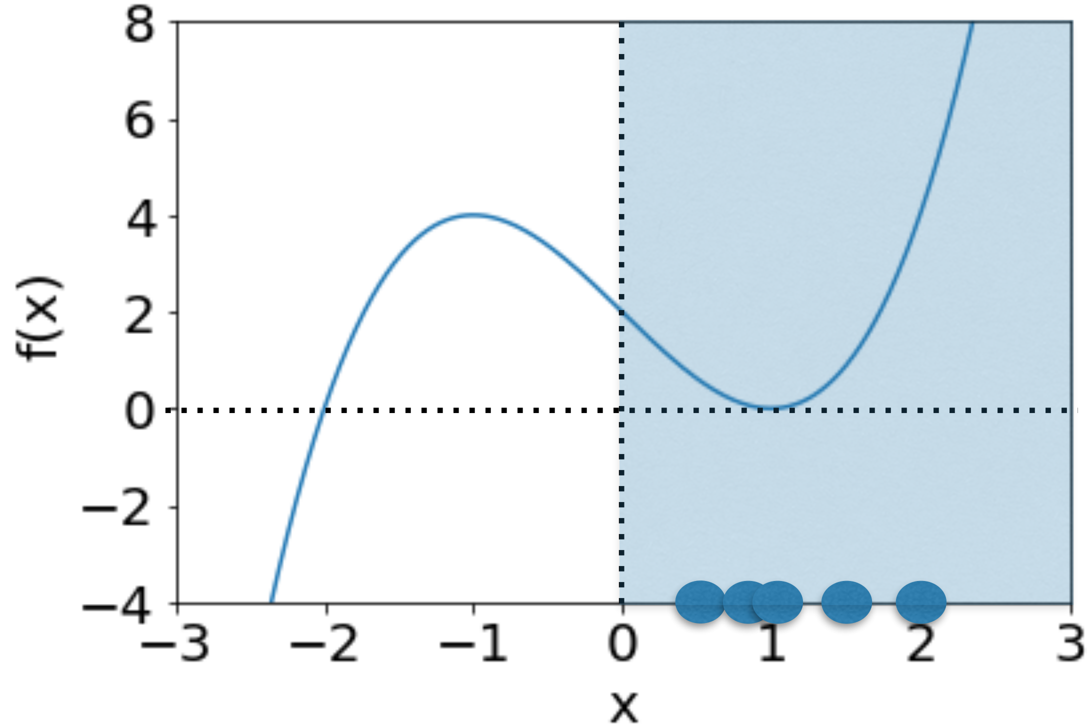
$f'(0.9) = -0.57$

$f'(0.5) = -2.25$

The function is convex *iif* $f''(x) \geq 0$ for all x.

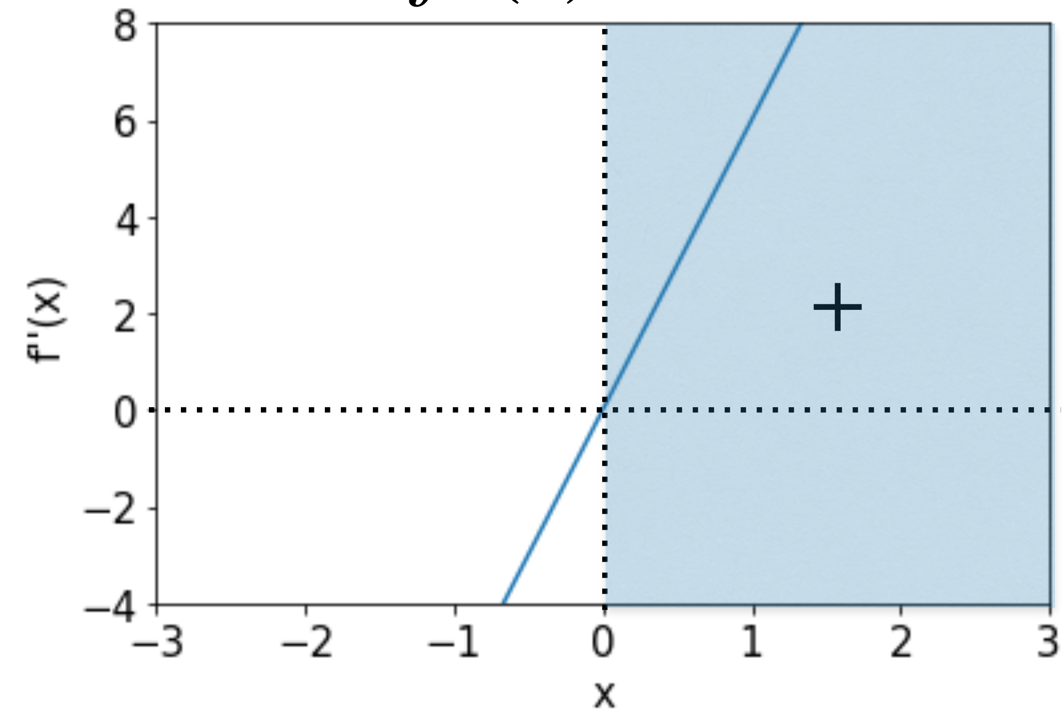
Second-Order Derivatives

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

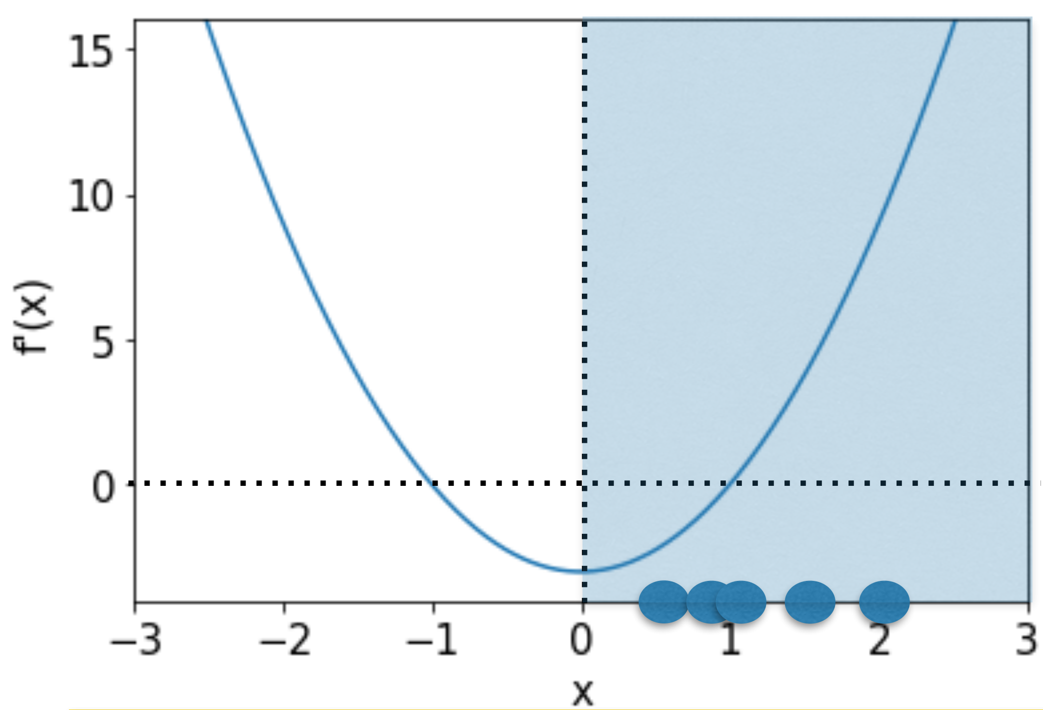


convex region

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



$$f'(x) = 3x^2 - 3$$



increasing

$f'(2) = 9$

$f'(1.5) = 3.5$

$f'(1) = 0$

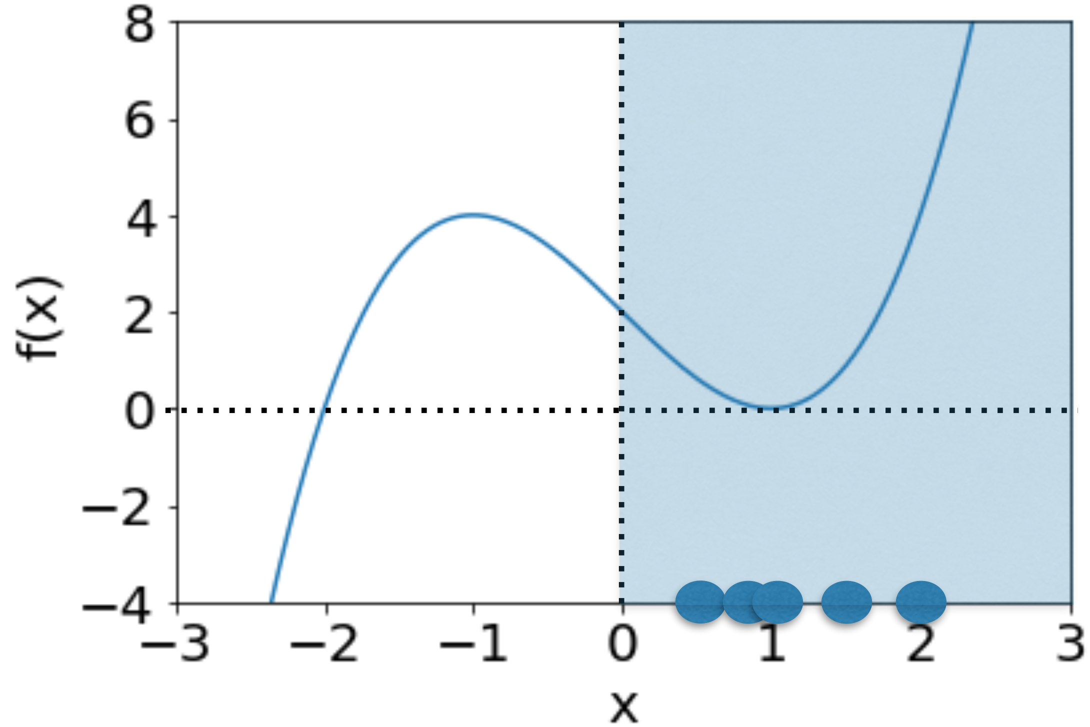
$f'(0.9) = -0.57$

$f'(0.5) = -2.25$

If $f''(x) > 0$ for all x, a function is strictly convex (sufficient but not necessary condition).

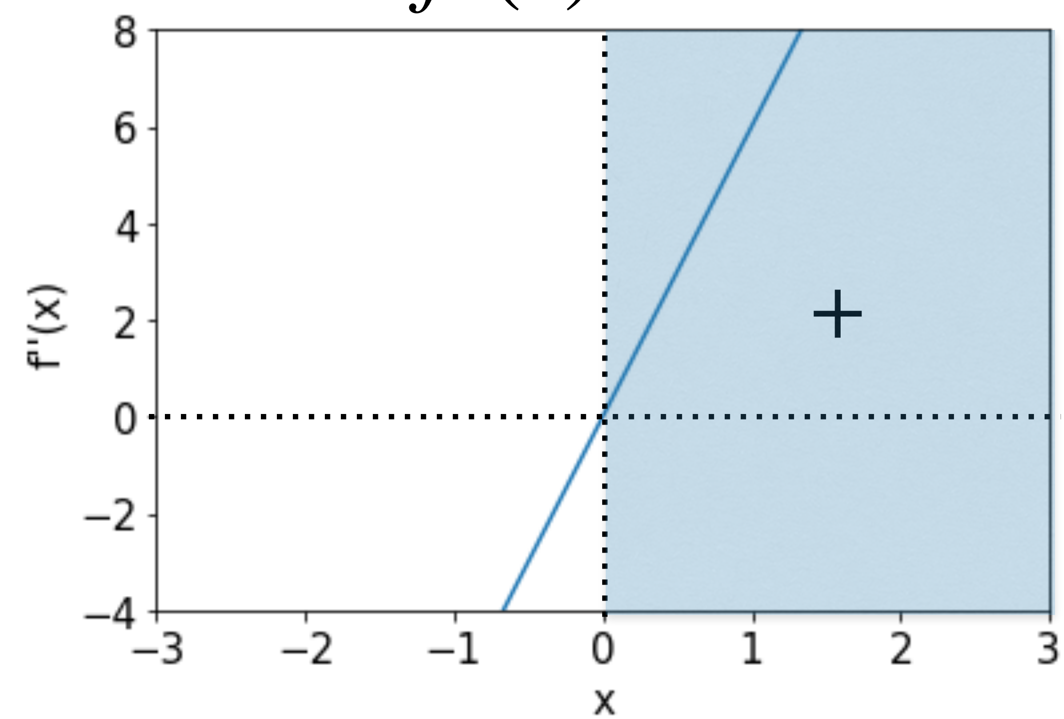
Second Derivative Test of Optimality

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

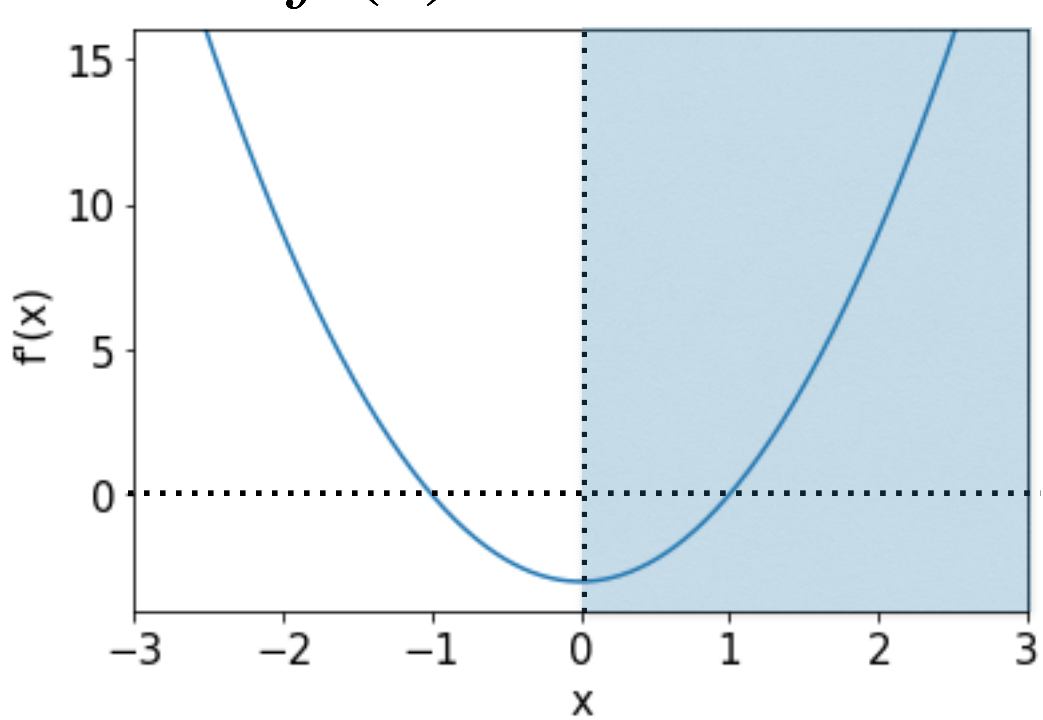


convex region

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



$$f'(x) = 3x^2 - 3$$



increasing

$f'(2) = 9$

$f'(1.5) = 3.5$

$f'(1) = 0$

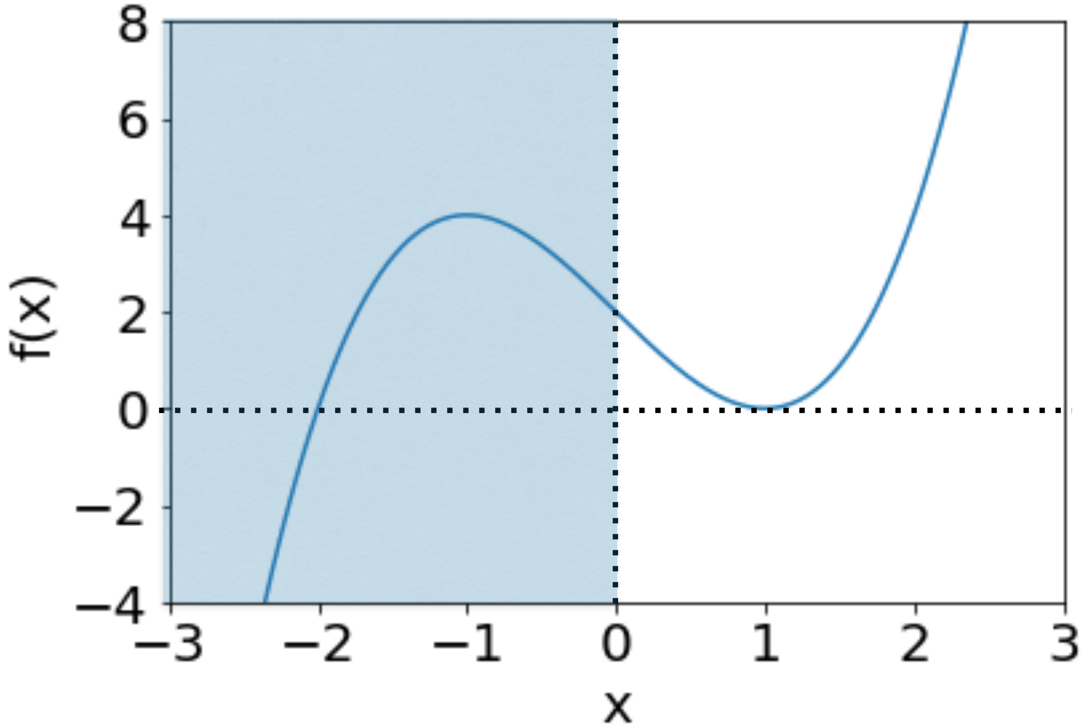
$f'(0.9) = -0.57$

$f'(0.5) = -2.25$

If $f'(x) = 0$ and $f''(x) > 0$, then x is a (local) minimum (sufficient but not necessary condition).

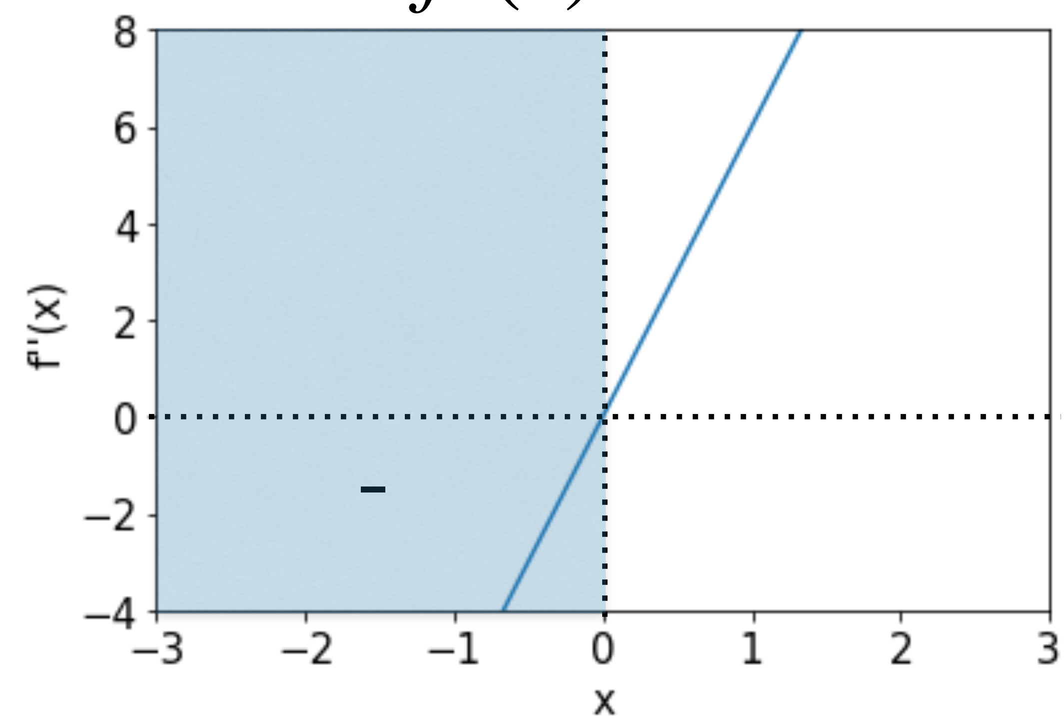
Second-Order Derivatives

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

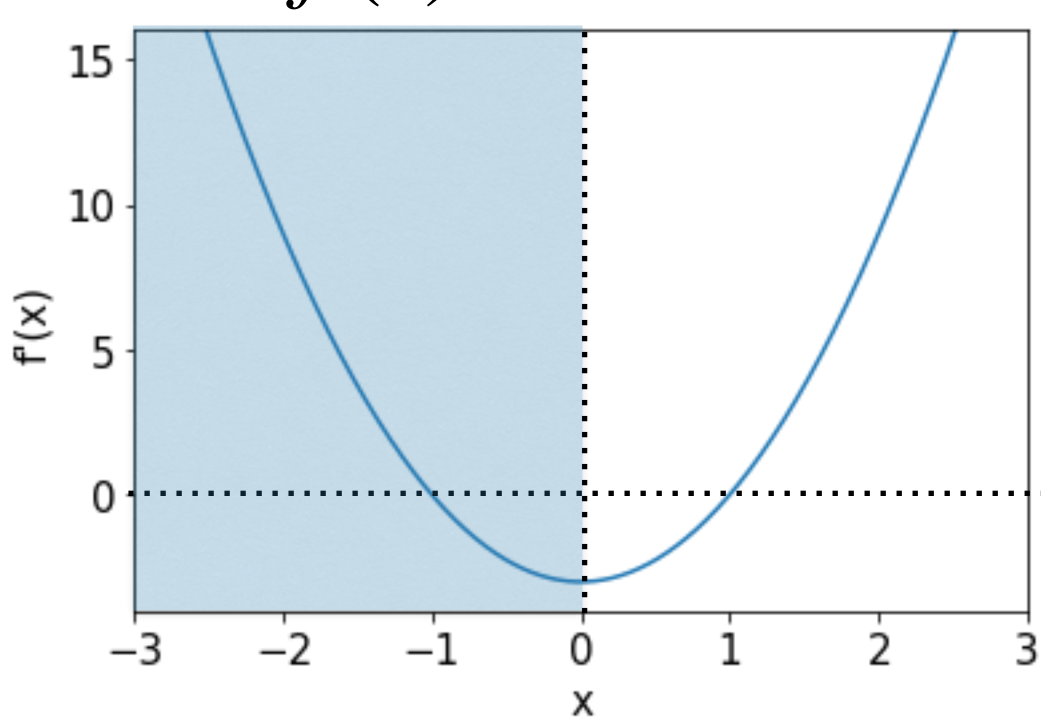


concave region

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



$$f'(x) = 3x^2 - 3$$

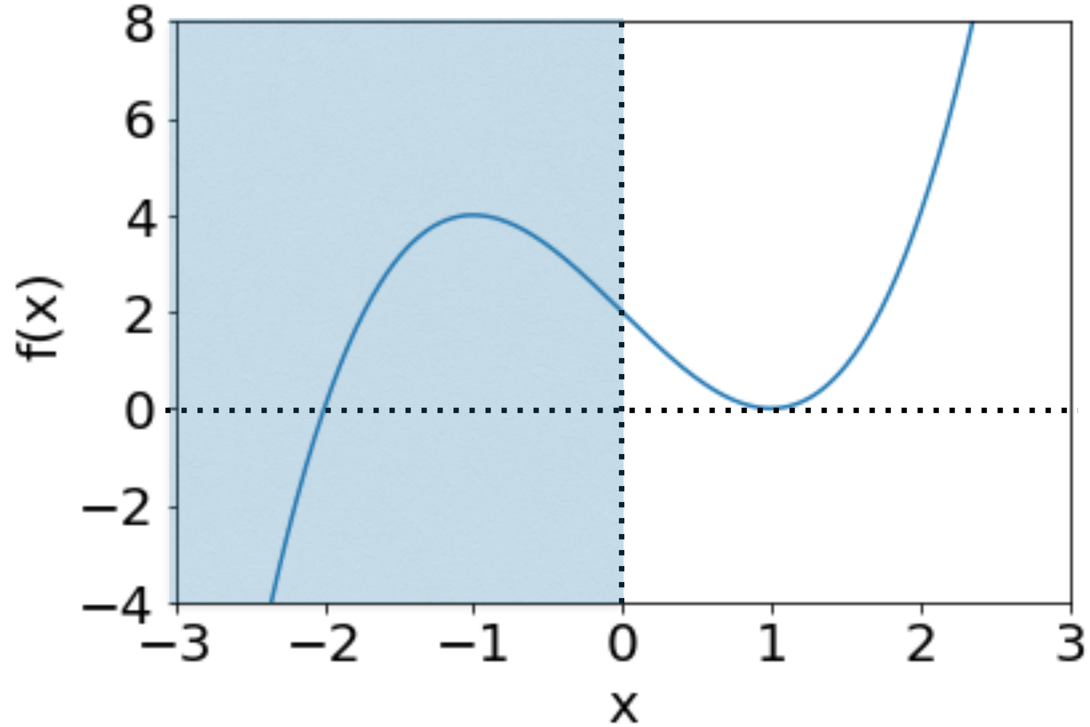


decreasing

Slope was upwards, positive. While positive, it got less and less steep, decreasing to zero. Then, got steeper and steeper, more and more negative.

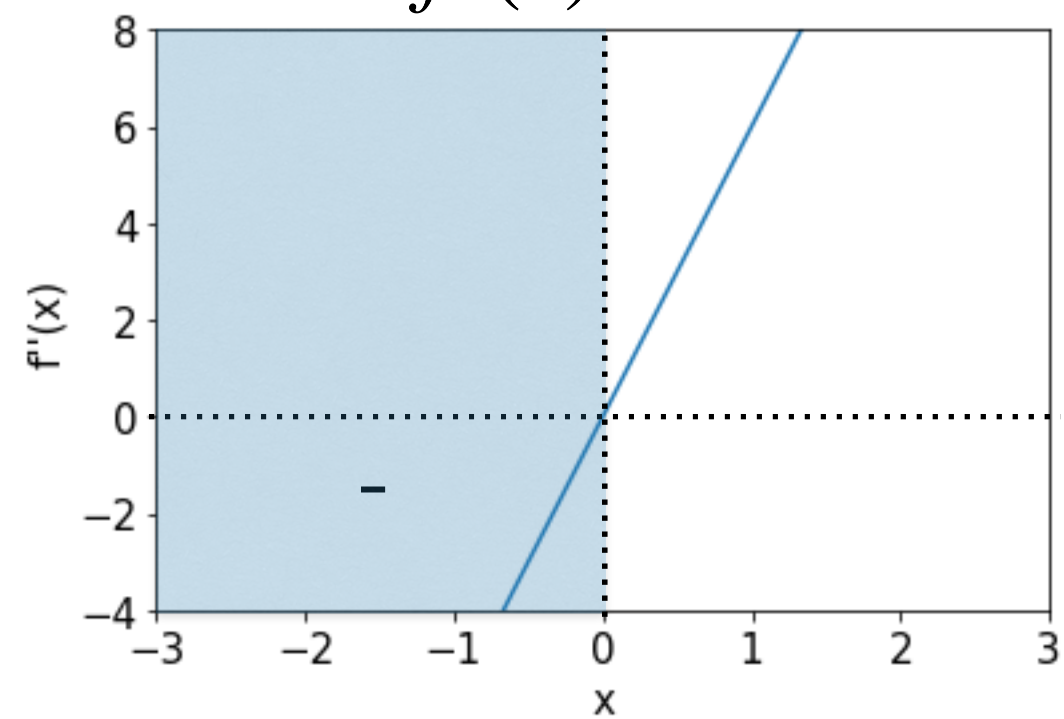
Second-Order Derivatives

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

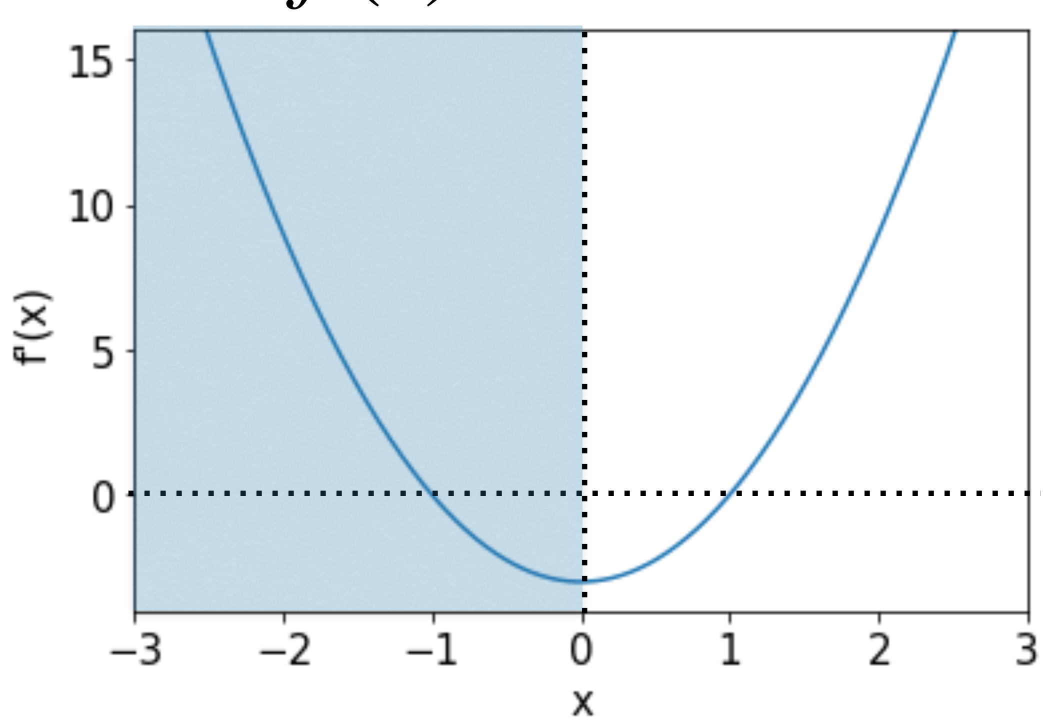


concave region

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



$$f'(x) = 3x^2 - 3$$

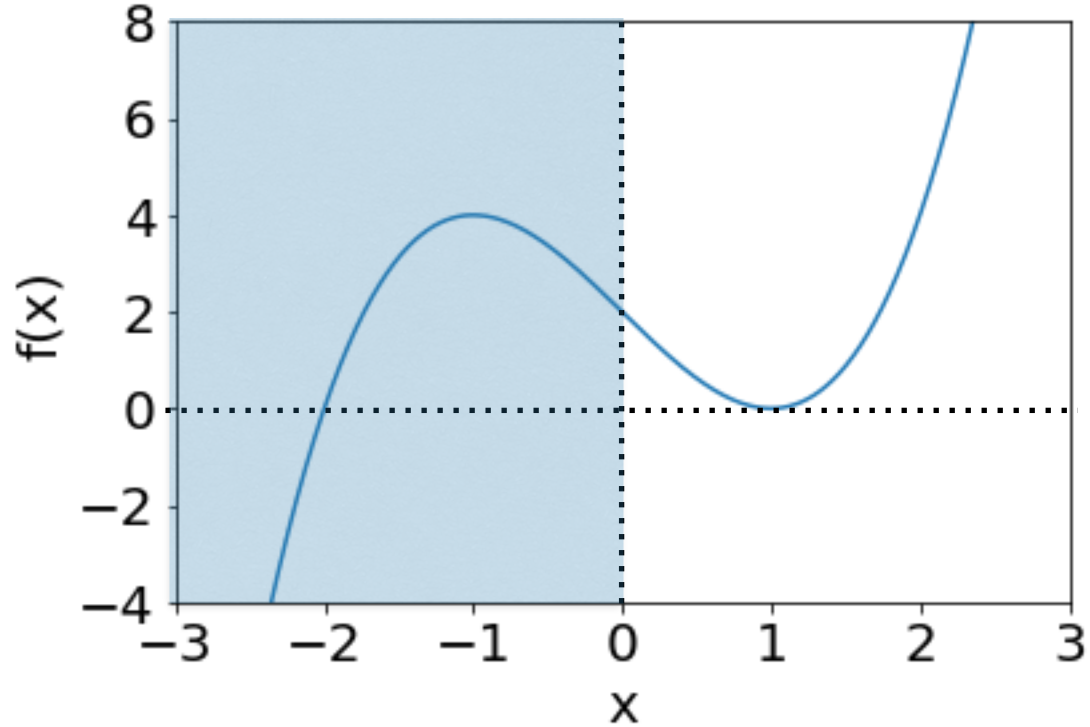


decreasing

The function is concave *iif* $f''(x) \leq 0$ for all x.

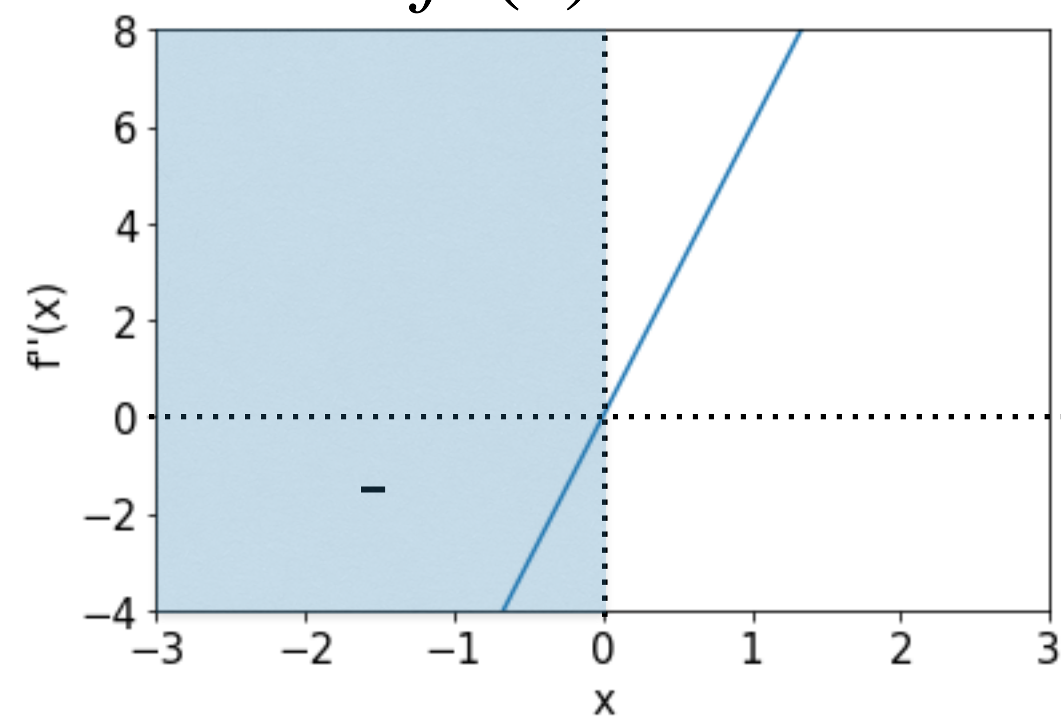
Second-Order Derivatives

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

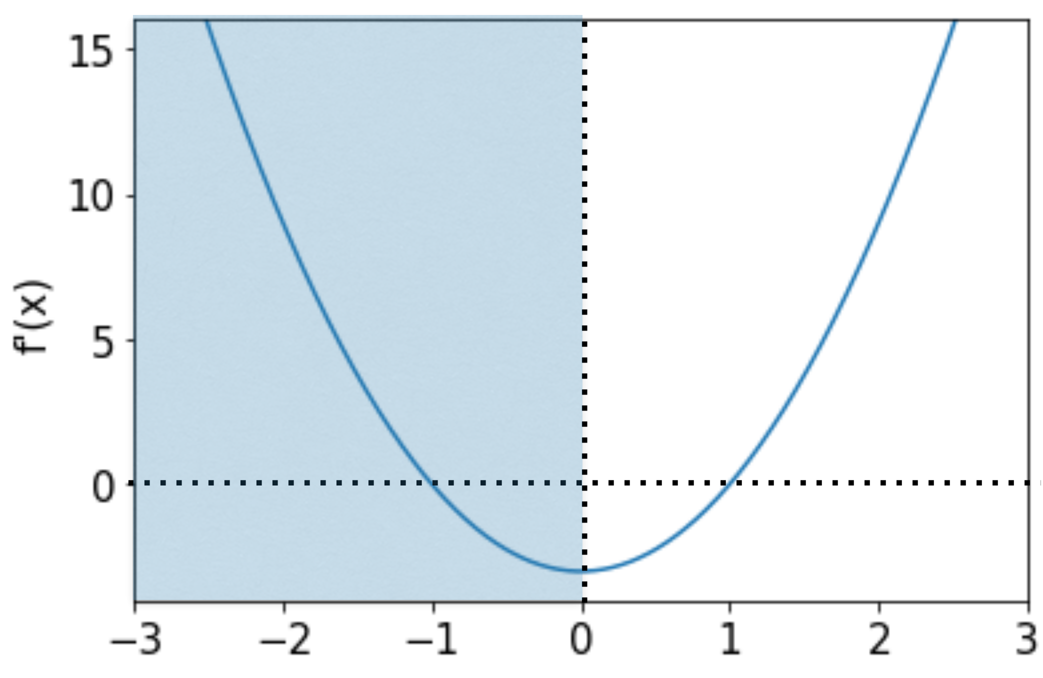


concave region

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



$$f'(x) = 3x^2 - 3$$

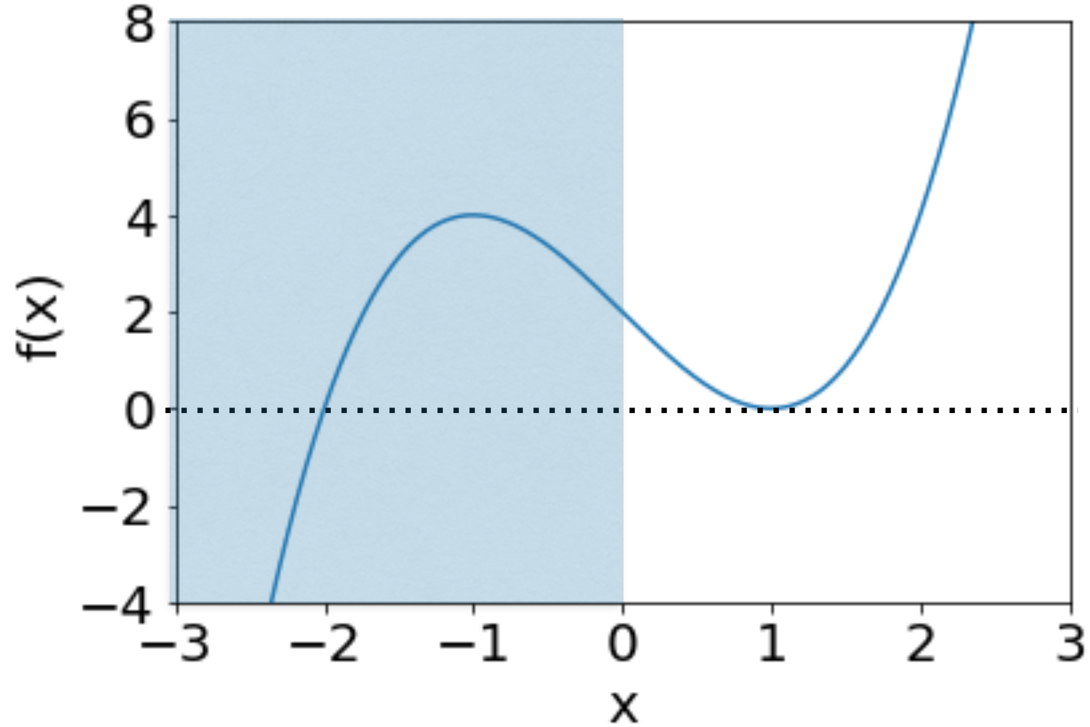


decreasing

If $f''(x) < 0$ for all x, the function is strictly concave (sufficient but not necessary condition).

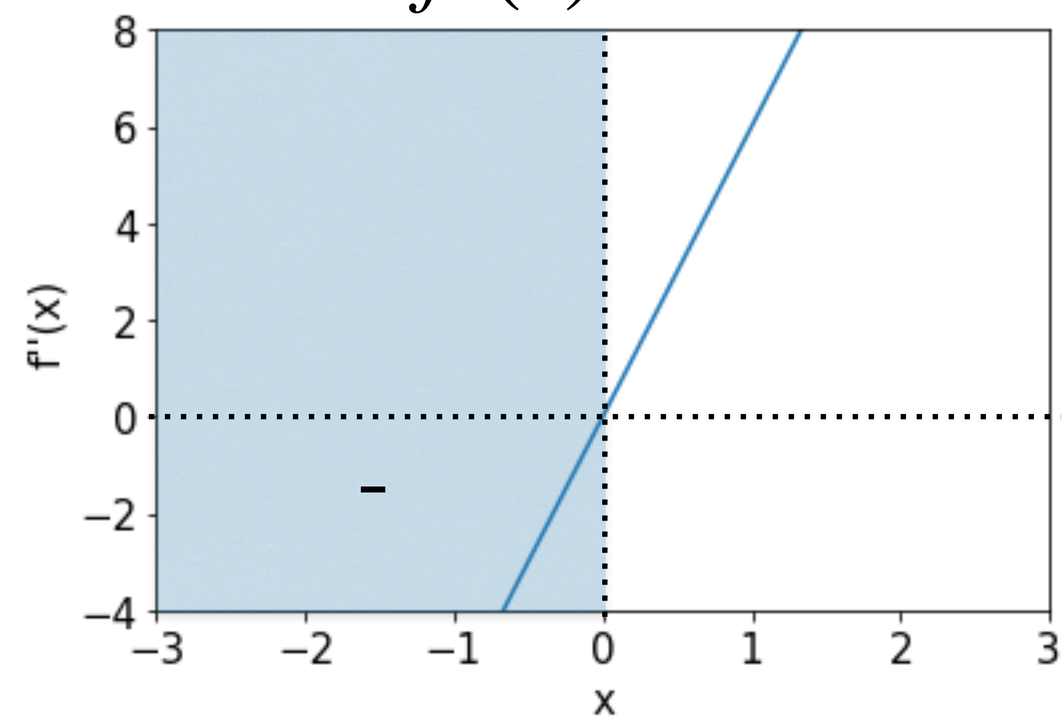
Second-Order Derivatives

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

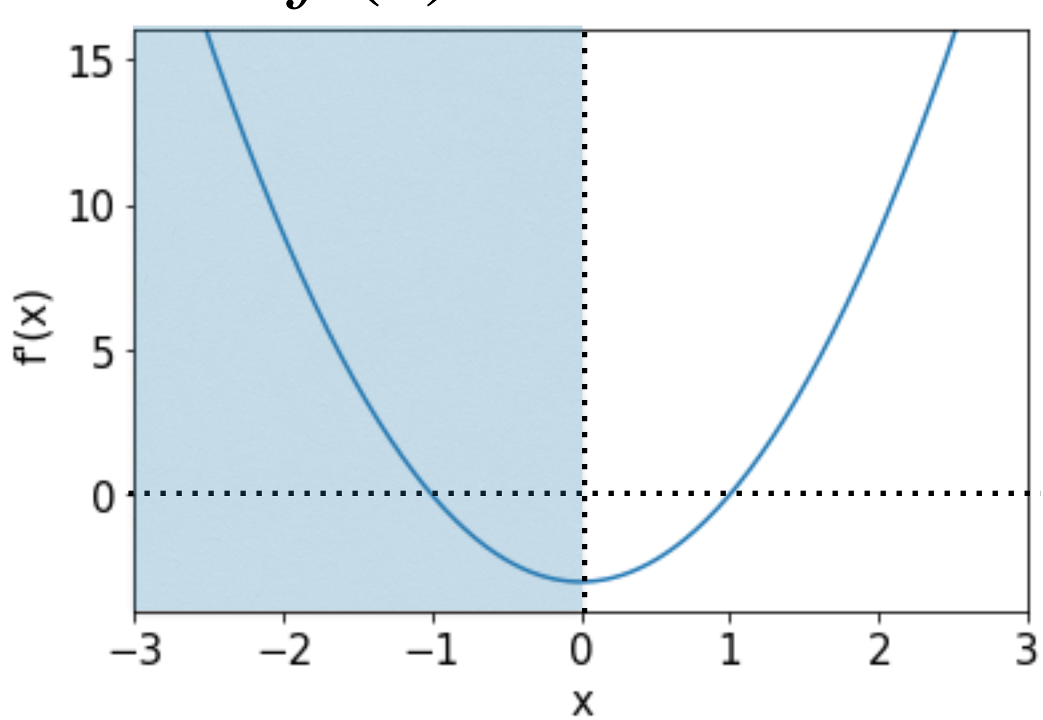


concave region

$f''(x) = 6x$



$f'(x) = 3x^2 - 3$



decreasing

If $f'(x) = 0$ and $f''(x) < 0$, then x is a (local) maximum (sufficient but not necessary condition).

Univariate Case

The function is convex *iif* $f''(x) \geq 0$ for all x.

If $f''(x) > 0$ for all x, a function is strictly convex
(sufficient but not necessary condition).

If $f'(x) = 0$ and $f''(x) > 0$, then x is a (local) minimum
(sufficient but not necessary condition).

Second Derivative Characterisation of Convexity:

Multivariate Case

Multivariate Case

The function is convex *iif* $f''(x) \geq 0$ for all x.

The function is convex *iif* $H_f(\mathbf{x}) \succcurlyeq 0$ (positive semidefinite) for all $\mathbf{x}$.

- Univariate case: second order derivative captures the curvatures.
- Multivariate case: Hessian “as a whole” can be seen as “instructions” on how the function is curved, and not its individual second order partial derivatives in isolation.

Positive Semidefinite Matrix

A $d \times d$ symmetric matrix A is positive semidefinite *iif* for any non-zero vector $\mathbf{z} \in \mathbb{R}^d$, the following is true:

$$\mathbf{z}^T A \mathbf{z} \geq 0$$

$$\mathbf{z}^T A \mathbf{z} = (z_1, z_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = (z_1, z_2) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1^2 + z_2^2$$

Satisfying the above with > defines a "positive definite" matrix.

Positive Semidefinite Matrix

A $d \times d$ symmetric matrix A is positive semidefinite *iif* for any non-zero vector $\mathbf{z} \in \mathbb{R}^d$, the following is true:

$$\mathbf{z}^T A \mathbf{z} \geq 0$$

$$\mathbf{z}^T A \mathbf{z} = (z_1, z_2) \begin{pmatrix} 3x_1^2 & 0 \\ 0 & 3x_2^2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = (3z_1x_1^2, 3z_2x_2^2) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 3z_1^2x_1^2 + 3z_2^2x_2^2$$

We could plug in certain values of x_1 and x_2 into A to determine whether it is positive semidefinite for those values.

Satisfying the above with > defines a "positive definite" matrix.

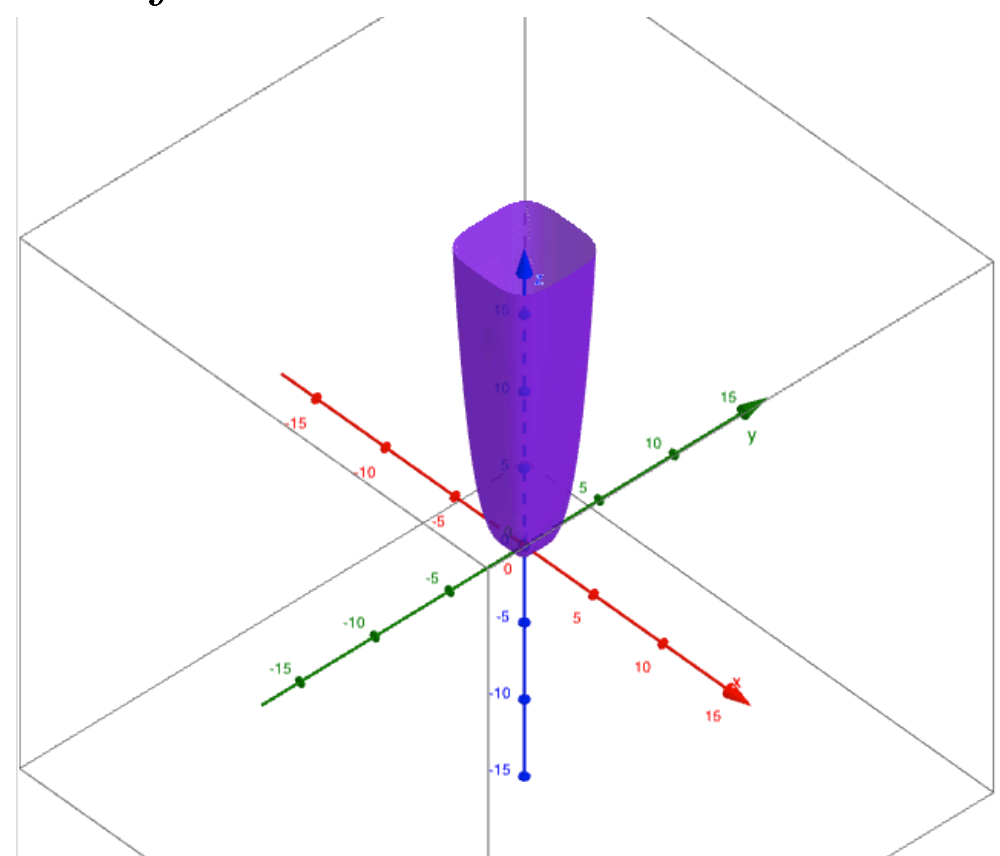
Second-Derivative Characterisation of Convexity

A twice differentiable function $f(\mathbf{x})$ is convex *iif*:

- its domain C is a convex set and
- its Hessian $H_f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in C$.

For any $\mathbf{z}, \mathbf{x}$, we have $\mathbf{z}^T H_f(\mathbf{x}) \mathbf{z} \geq 0$.

$$H_f(\mathbf{x}) = \begin{pmatrix} 3x_1^2 & 0 \\ 0 & 3x_2^2 \end{pmatrix}$$



Second-Derivative Characterisation of Convexity

A twice differentiable function $f(\mathbf{x})$ is convex *iif*:

- its domain C is a convex set and
- its Hessian $H_f(\mathbf{x})$ is positive semidefinite for all $\mathbf{x} \in C$.

If a twice differentiable function $f(\mathbf{x})$:

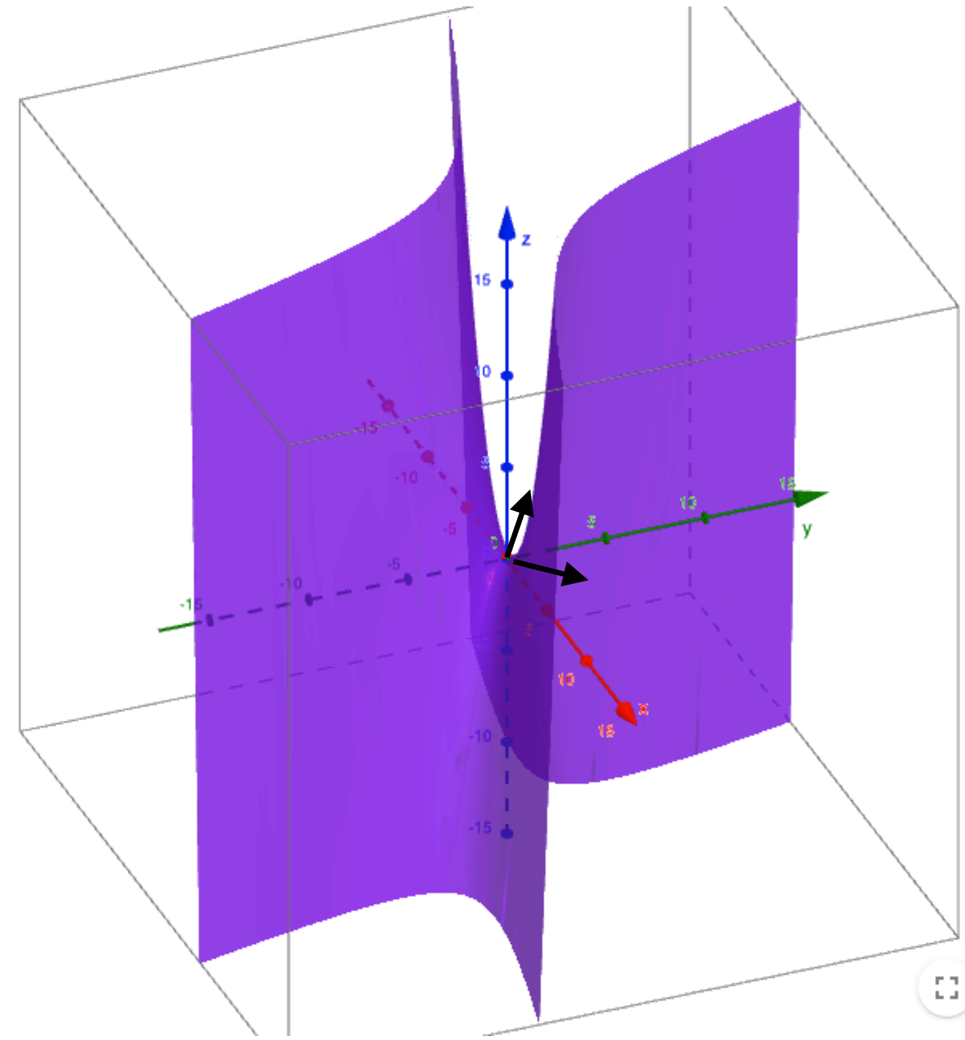
- has a convex set C as its domain and
- its Hessian $H_f(\mathbf{x})$ is positive definite for all $\mathbf{x} \in C$

it is a strictly convex function. (sufficient but not necessary condition)

Watch Out!

A Hessian with only positive entries may not be positive semidefinite, e.g.:

$$H_f(\mathbf{x}) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$



$$(z_1, z_2)\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = (z_1 + 2z_2, 2z_1 + z_2)\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = z_1^2 + 4z_1z_2 + z_2^2$$

If $z_1 = 1$ and $z_2 = -1$, we get -2

Source of matrix: https://en.wikipedia.org/wiki/Definite_matrix

Eigenvalues and Eigenvectors

- The eigenvalues of H capture the direction of the principal curvatures of the function $f(\mathbf{x})$, where the curvature is most pronounced.

- The eigenvalues of H capture the curvature itself.
- If all eigenvalues are ≥ 0, the curvature is always positive, "upwards".
- The eigenvalues are ≥ 0 *iif* $H_f(\mathbf{x}) \succcurlyeq 0$.

Watch Out!

A Hessian with negative entries may still be positive semidefinite, e.g.:

$$H_f(\mathbf{x}) = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

$$(z_1, z_2, z_3)\begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = z_1^2 + (z_1 - z_2)^2 + (z_2 - z_3)^2 + z_3^2$$

Source of matrix: https://en.wikipedia.org/wiki/Definite_matrix

Multivariate Case

The function is convex *iif* $f''(x) \geq 0$ for all x.

The function is convex *iif* $H_f(\mathbf{x}) \succcurlyeq 0$ (positive semidefinite) for all $\mathbf{x}$.

If $f''(x) > 0$ for all x, a function is strictly convex (sufficient but not necessary condition).

If $H_f(\mathbf{x}) \succ 0$ (positive definite) for all $\mathbf{x}$, a function is strictly convex (sufficient but not necessary condition).

If $f'(x) = 0$ and $f''(x) > 0$, then x is a (local) minimum (sufficient but not necessary condition).

If $\nabla f(\mathbf{x}) = 0$ and $H_f(\mathbf{x}) \succ 0$, then x is a (local) minimum (sufficient but not necessary condition).

Further Reading

- Recommended:
  - Charu C. Aggarwal's Linear Algebra and Optimization for Machine Learning. Sections 3.3.8 (Positive Semidefinite Matrices), 4.2.1(Univariate Optimization), 4.2.3 (Multivariate Optimization), 4.3 (Convex Functions).

- Optional:
  - Stephen Boyd and Lieven Vandenberghe's Convex Optimization, Cambridge University Press, 2004. Sections 2.1.4 (Convex Sets) and 3.1 (Basic Properties and Examples) until Section 3.1.5 (inclusive). Available at: https://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf