Basics of Convex Optimization

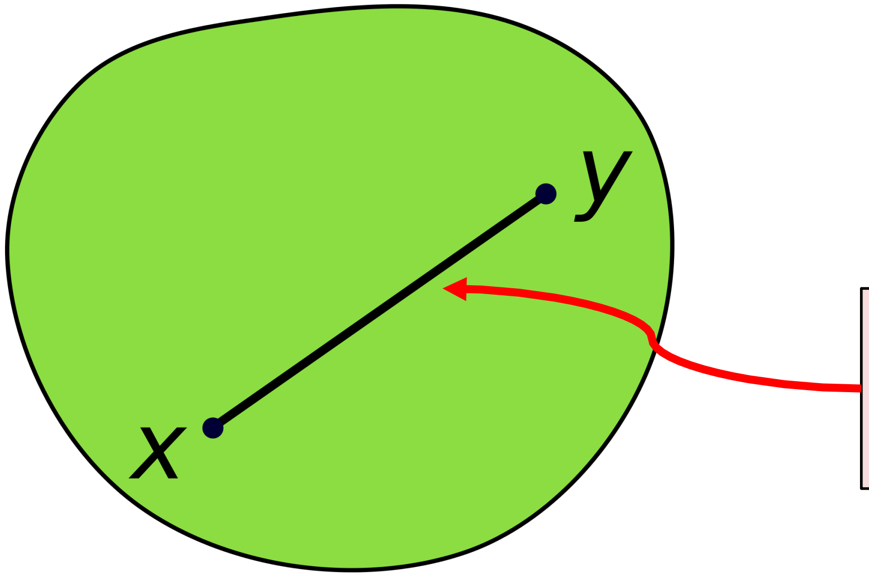
Shusen Wang

Convex Sets

Convex Set

Definition (Convex Set).

A set $\mathcal{C}$ is convex if and only if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and any $\eta \in (0,1)$, the point $\eta\mathbf{x} + (1-\eta)\mathbf{y}$ is also in $\mathcal{C}$.



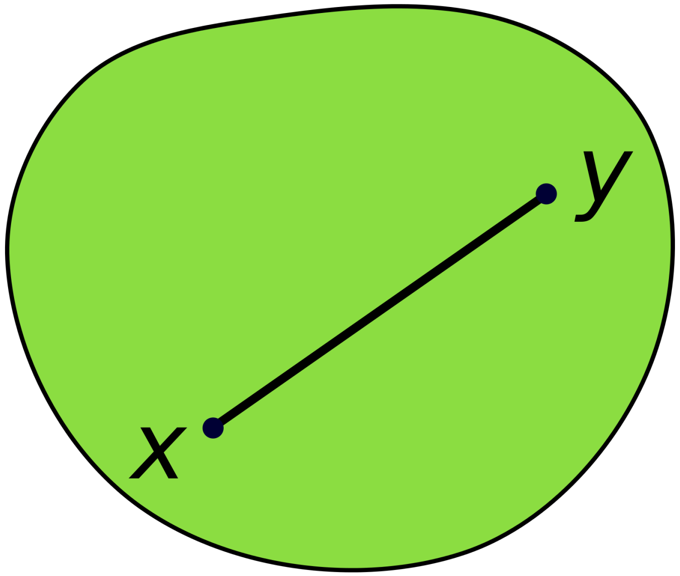
By definition, the line segment between $\mathbf{x}$ and $\mathbf{y}$ is in $\mathcal{C}$.

A convex set $\mathcal{C}$.

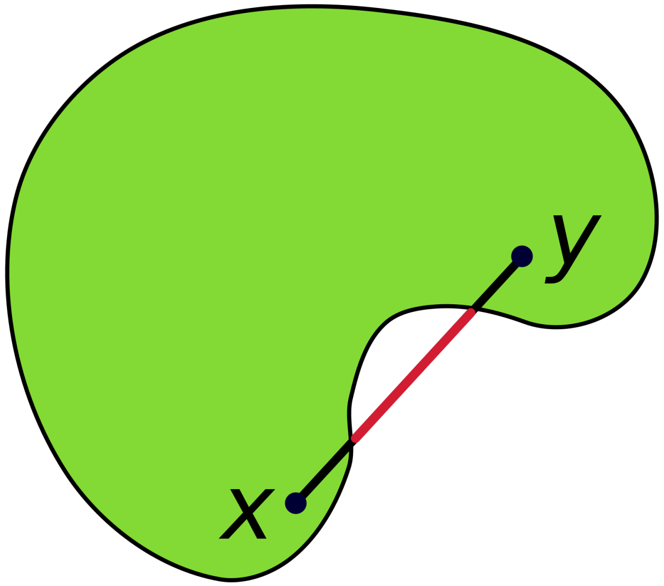
Convex Set

Definition (Convex Set).

A set $\mathcal{C}$ is convex if and only if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and any $\eta \in (0, 1)$, the point $\eta\mathbf{x} + (1 - \eta)\mathbf{y}$ is also in $\mathcal{C}$.



A convex set $\mathcal{C}$.

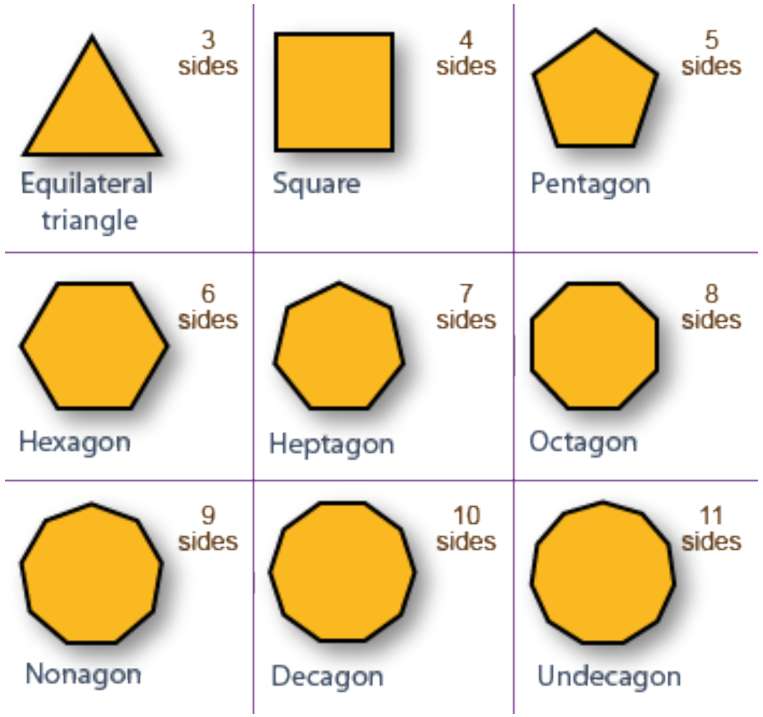


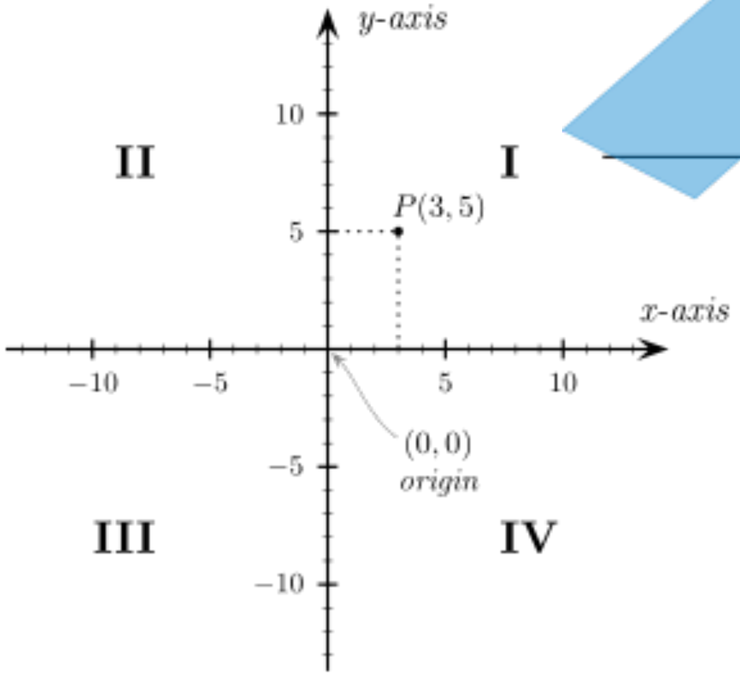
A non-convex set.

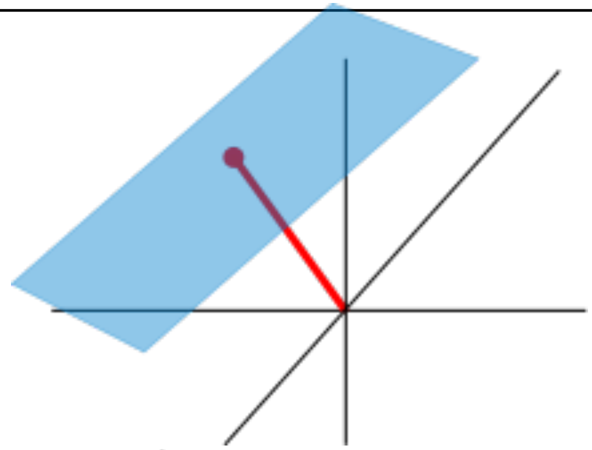
Convex Set: Examples

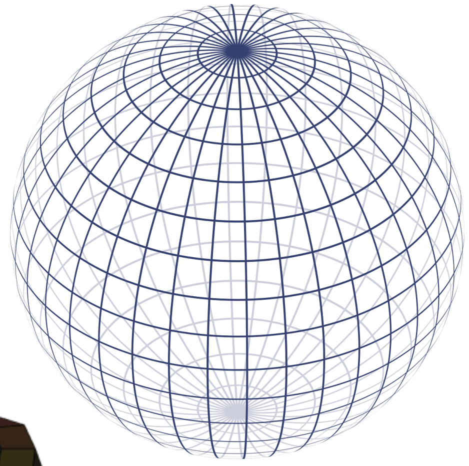
Definition (Convex Set).

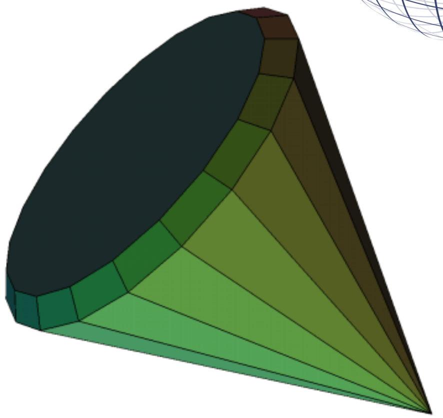
A set $\mathcal{C}$ is convex if and only if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and any $\eta \in (0, 1)$, the point $\eta\mathbf{x} + (1 - \eta)\mathbf{y}$ is also in $\mathcal{C}$.





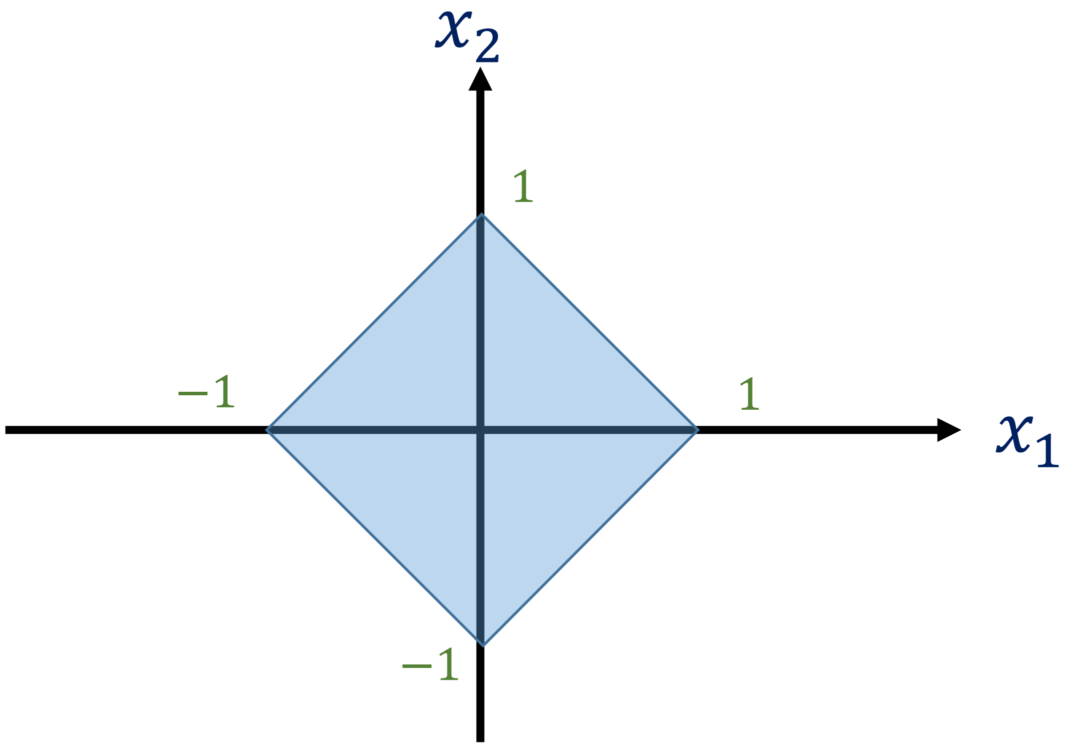






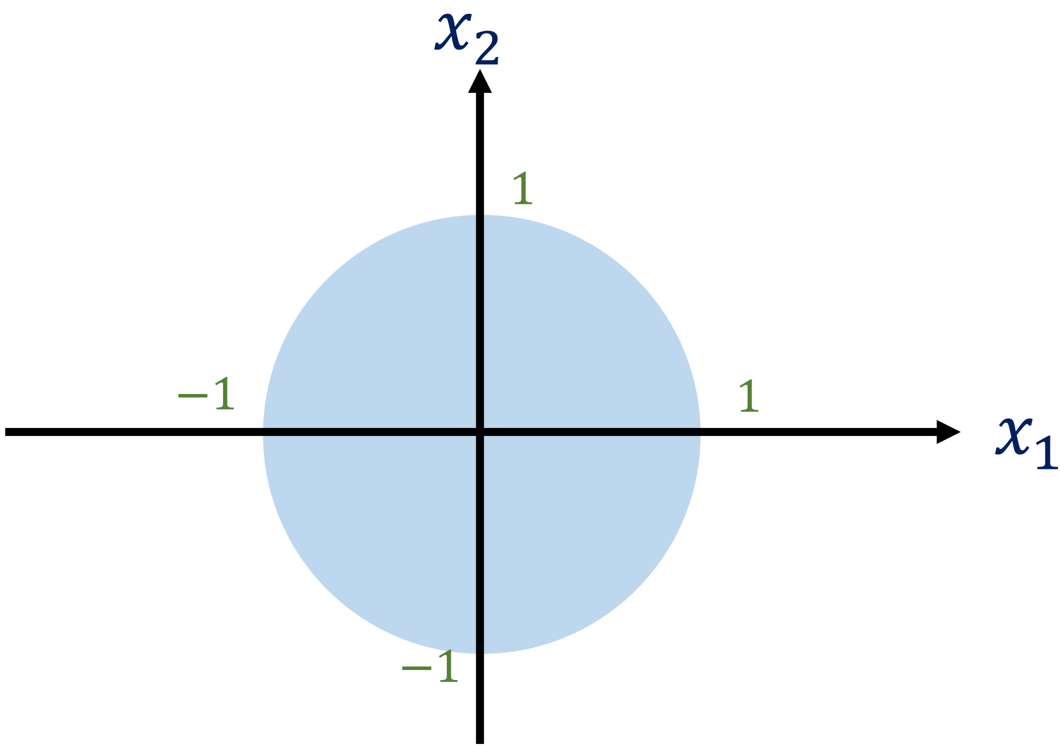
Convex Set: Examples

Example: The ℓ_1-norm ball $\left\{\mathbf{x}: \;\; ||\mathbf{x}||_1 \leq 1\right\}$.



Convex Set: Examples

Example: The ℓ_2-norm ball $\left\{\mathbf{x}: \ ||\mathbf{x}||_2 \leq 1\right\}$.



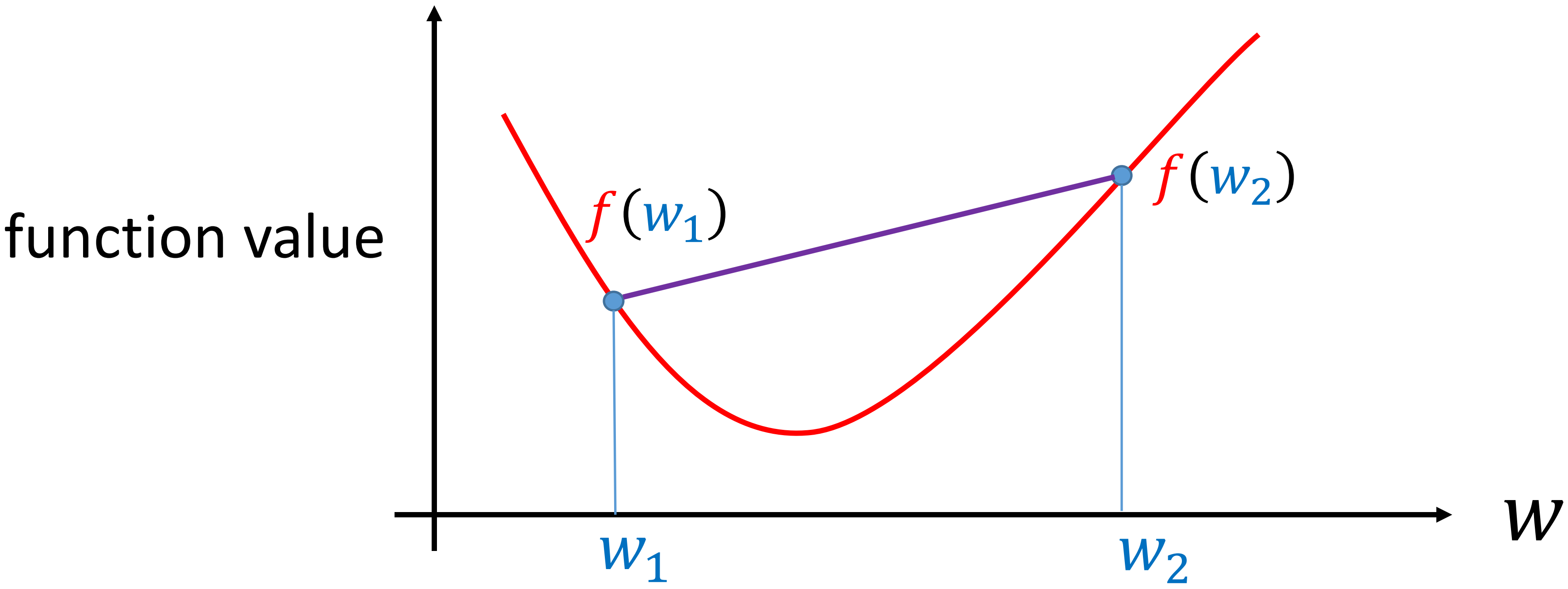
Convex Functions

Convex Function

Definition (Convex Function).

- Let $\mathcal{C}$ be a convex set and $f: \mathcal{C} \mapsto \mathbb{R}$ be a function.
- f is convex if for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{C}$ and any $\eta \in (0, 1)$,

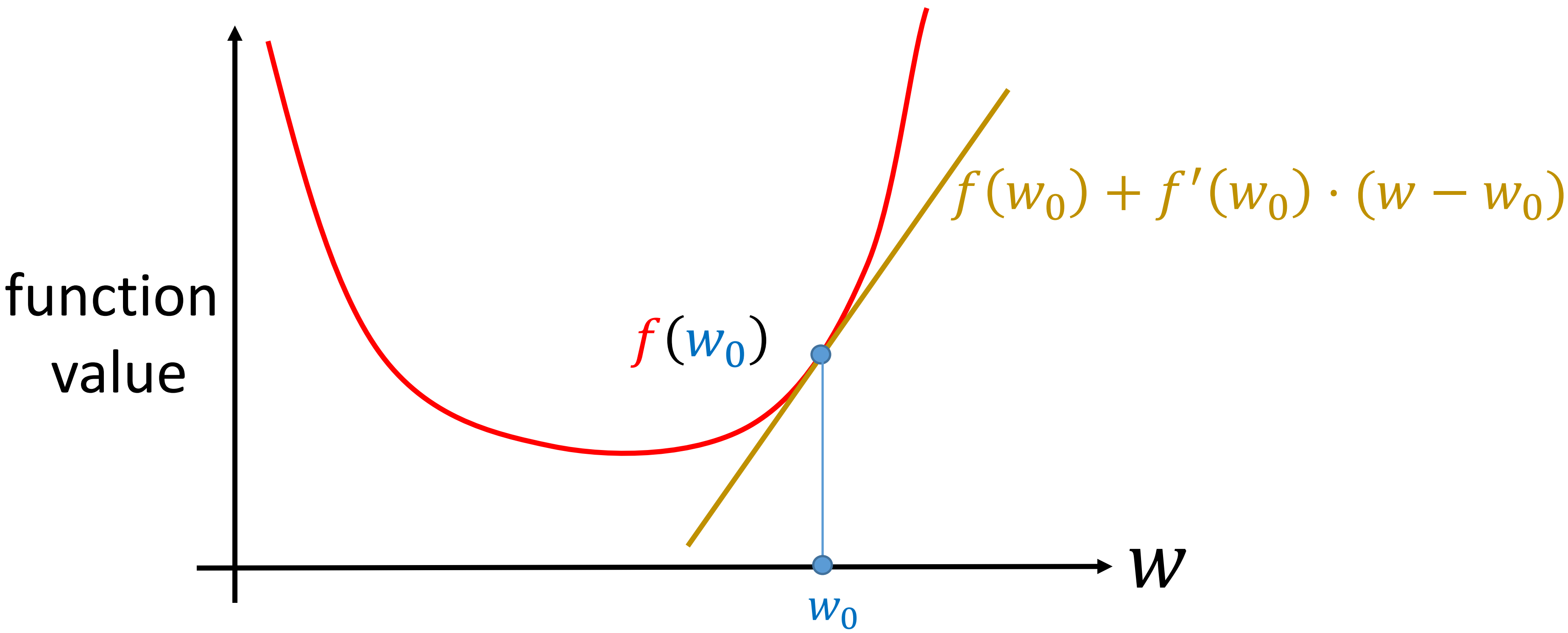
$$f(\eta \mathbf{w}_1 + (1-\eta)\mathbf{w}_2) \ \le \ \eta f(\mathbf{w}_1) + (1-\eta) f(\mathbf{w}_2).$$



Convex Function: Properties

Properties of convex function:

1. $f(\mathbf{w}_0) + \nabla f(\mathbf{w}_0)^T(\mathbf{w} - \mathbf{w}_0) \leq f(\mathbf{w})$. (Assume f is differentiable).



Convex Function: Properties

Properties of convex function:

1. $f(\mathbf{w}_0) + \nabla f(\mathbf{w}_0)^T(\mathbf{w} - \mathbf{w}_0) \leq f(\mathbf{w})$. (Assume f is differentiable).
2. The Hessian matrix is everywhere positive semi-definite: $\nabla^2 f(\mathbf{w}) \succcurlyeq \mathbf{0}$.

- Assume f is twice differentiable.
- $\mathbf{H} \in \mathbb{R}^{d \times d}$ is positive semi-definite $\Longleftrightarrow$ for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^T\mathbf{H}\mathbf{x} \geq 0$.

Convex Functions

Question: Are they convex functions?

- $f(w) = w^2 + w - 1$, for $w \in \mathbb{R}$.
- $f(w) = w^4$, for $w \in \mathbb{R}$.
- $f(w) = \log_e w$, for $w > 0$.
- $f(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||_2^2$, for $\mathbf{w} \in \mathbb{R}^d$.
- $f(\mathbf{w}) = \frac{1}{2}||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2$, for $\mathbf{w} \in \mathbb{R}^d$.

Convex Function: Properties

Property: Combination of convex functions is convex function.

- Let g and h be convex functions.
- Then $h\big(g(\mathbf{w})\big)$ is convex function.

Example:

- Linear function $\mathbf{g}(\mathbf{w}) = \mathbf{X}^T\mathbf{w} - \mathbf{y}$ is convex.
- Vector norm $h(\mathbf{w}) = \|\mathbf{w}\|_2^2$ is convex.
- ➔ $h\big(\mathbf{g}(\mathbf{w})\big) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$ is convex function.

Convex Function: Properties

Property: Combination of convex functions is convex function.

- Let $f_1, \cdots, f_k$ be convex functions.
- Then $f(\mathbf{w}) = \lambda_1 f_1(\mathbf{w}) + \cdots + \lambda_k f_k(\mathbf{w})$ is convex function.

Example:

- $f_1(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$ is convex function.
- $f_2(\mathbf{w}) = \|\mathbf{w}\|_2^2$ is convex function.
- ➔ $f_1(\mathbf{w}) + \lambda f_2(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{w}\|_2^2$ is convex function.

Convex Optimization

Convex Optimization

Definition (Convex Optimization).

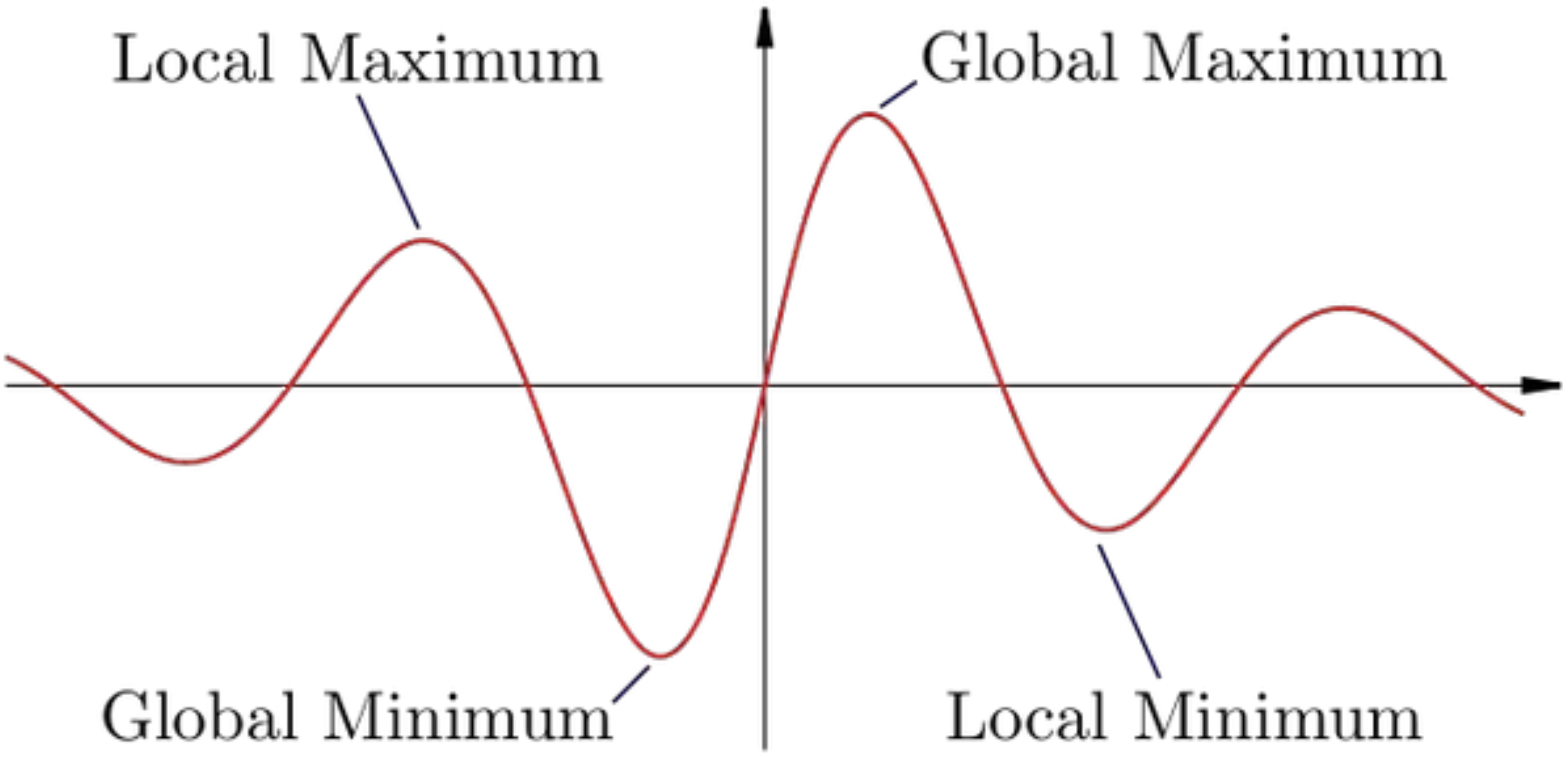
- Optimization: $\min_{\mathbf{w}} f(\mathbf{w})\,; \quad \text{s.t.}\ \mathbf{w} \in \mathcal{C}.$
- It is convex optimization if it has two properties:

    1. $\mathcal{C}$ (feasible set) is convex set,
    2. f (objective function) is convex function.

Convex Optimization: Examples

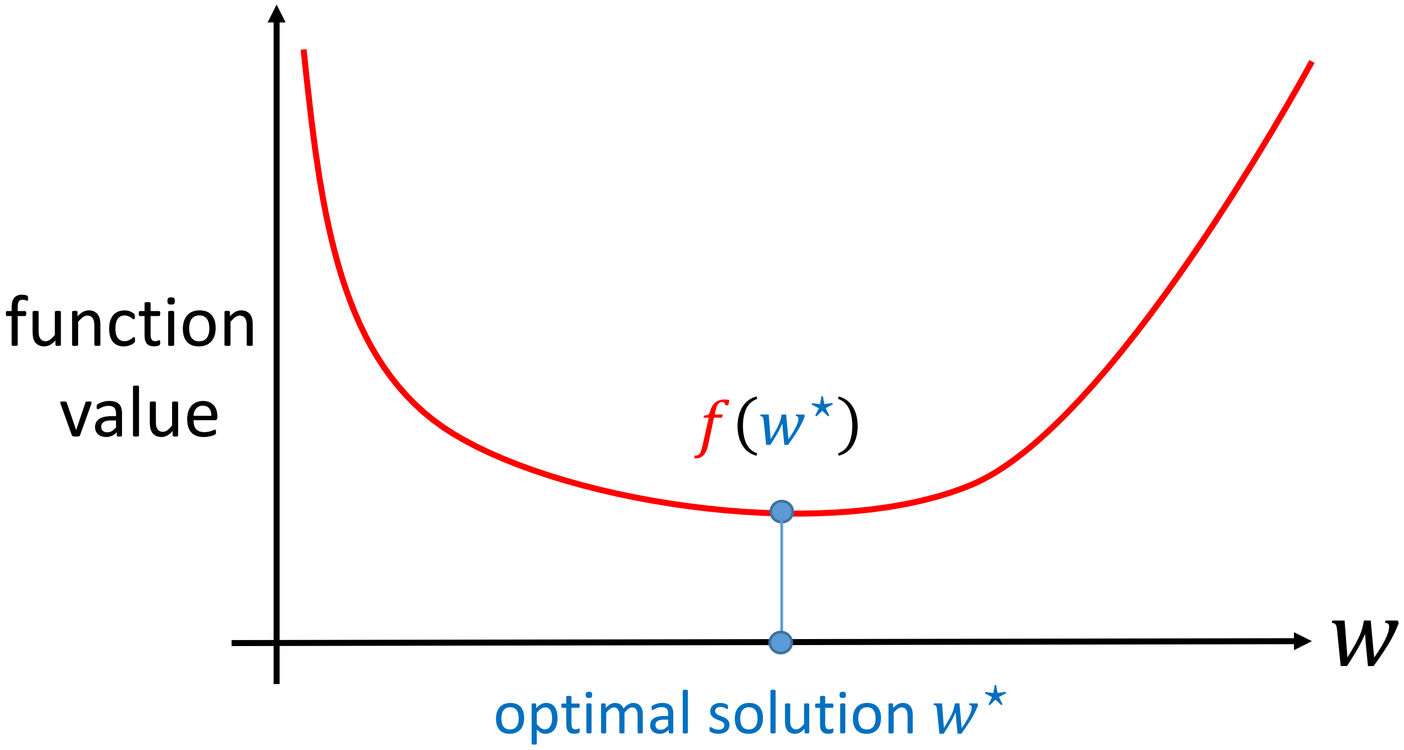
- Least squares regression: $\min_{\mathbf{w}} \; ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2$.
- Logistic regression: $\min_{\mathbf{w}} \sum_j \log\left(1 + \exp\left(-y_j \mathbf{w}^T \mathbf{x}_j\right)\right)$.
- SVM: $\min_{\mathbf{w},b} ||\mathbf{w}||_2^2 \; + \; \lambda \sum_j \left[1 - y_j\left(\mathbf{w}^T \mathbf{x}_j + b\right)\right]_+$.
- LASSO: $\min_{\mathbf{w}} \; ||\mathbf{X}\mathbf{w} - \mathbf{y}||_2^2 ; \quad s.t. \; ||\mathbf{w}||_1 \leq t$.

Local and Global Optima



Convex Optimization: Properties

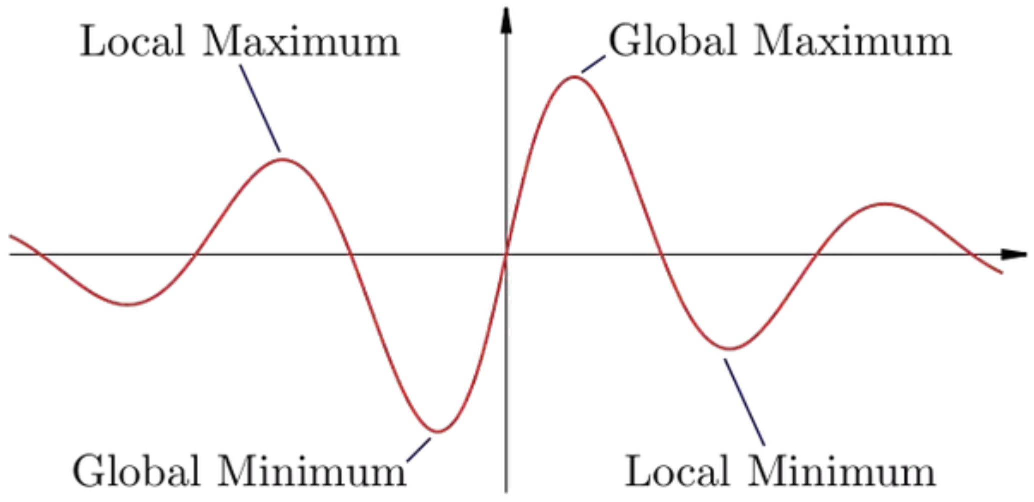
Property: For convex optimization, every local minimum is global minimum.



Convex Optimization: Properties

First-order optimality condition (necessary condition):

- Consider the unconstrained optimization: $\min_{\mathbf{w}} f(\mathbf{w})$.
- If $\mathbf{w}^\star$ is local minimum, then the gradient $\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}}$ at $\mathbf{w}^\star$ is zero.



Convex Optimization: Properties

First-order optimality condition (necessary condition):

- Consider the unconstrained optimization: $\min_{\mathbf{w}} f(\mathbf{w})$.
- If $\mathbf{w}^\star$ is local minimum, then the gradient $\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}}$ at $\mathbf{w}^\star$ is zero.

Property of convex optimization (sufficient condition):

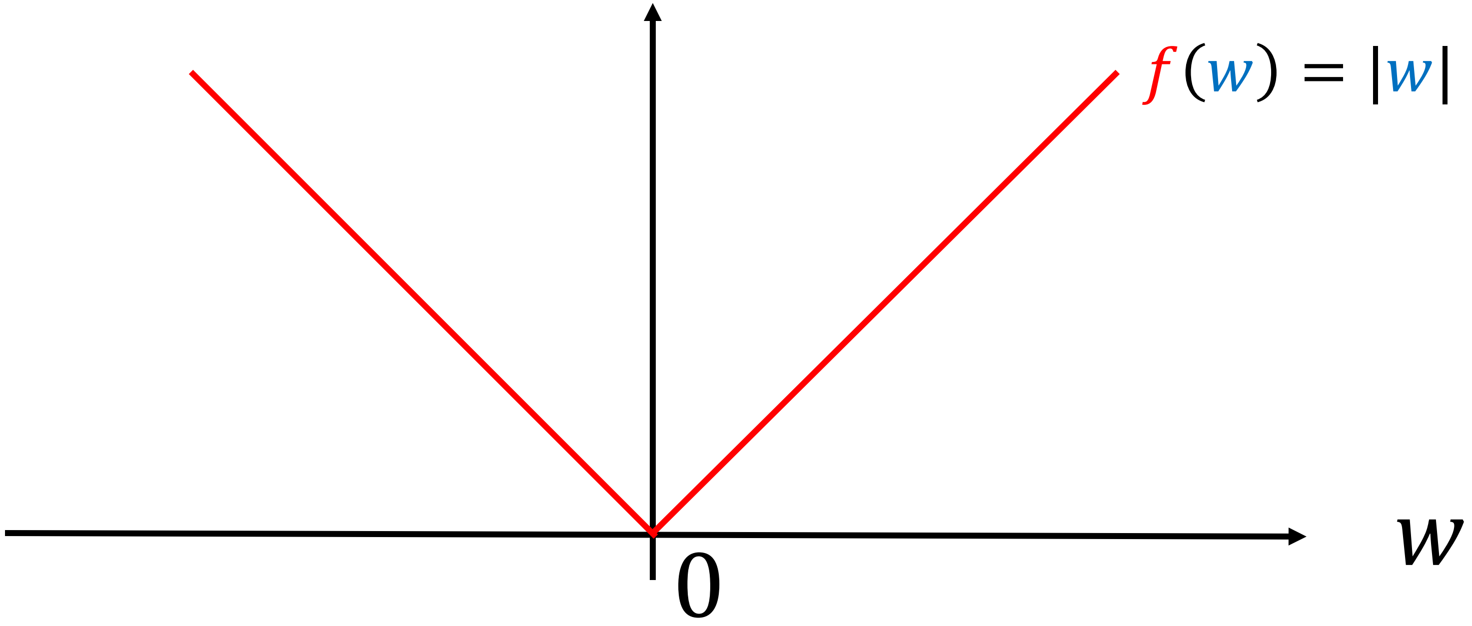
- Let $\min_{\mathbf{w}} f(\mathbf{w})$ be convex optimization.
- If $\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}}$ at $\mathbf{w}^\star$ is zero, then $\mathbf{w}^\star$ is global minimum.

Subgradient and Subdifferential

Non-Differentiable Functions

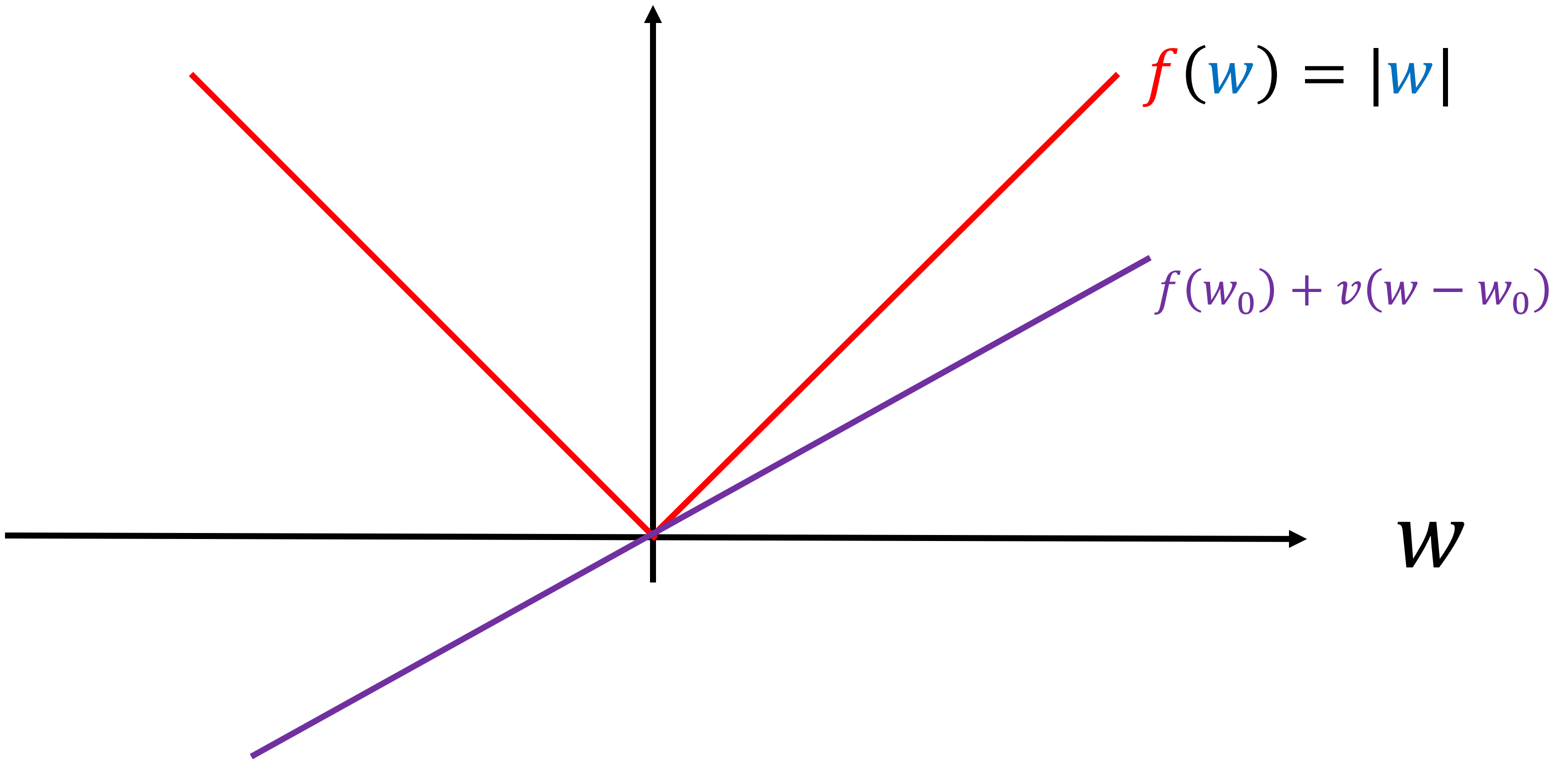
- Example of non-differentiable functions: $f(w) = |w|$

$$\frac{\partial f}{\partial w} = \begin{cases} +1, & \text{if } w > 0; \\ \text{undefined}, & \text{if } w = 0; \\ -1, & \text{if } w < 0. \end{cases}$$



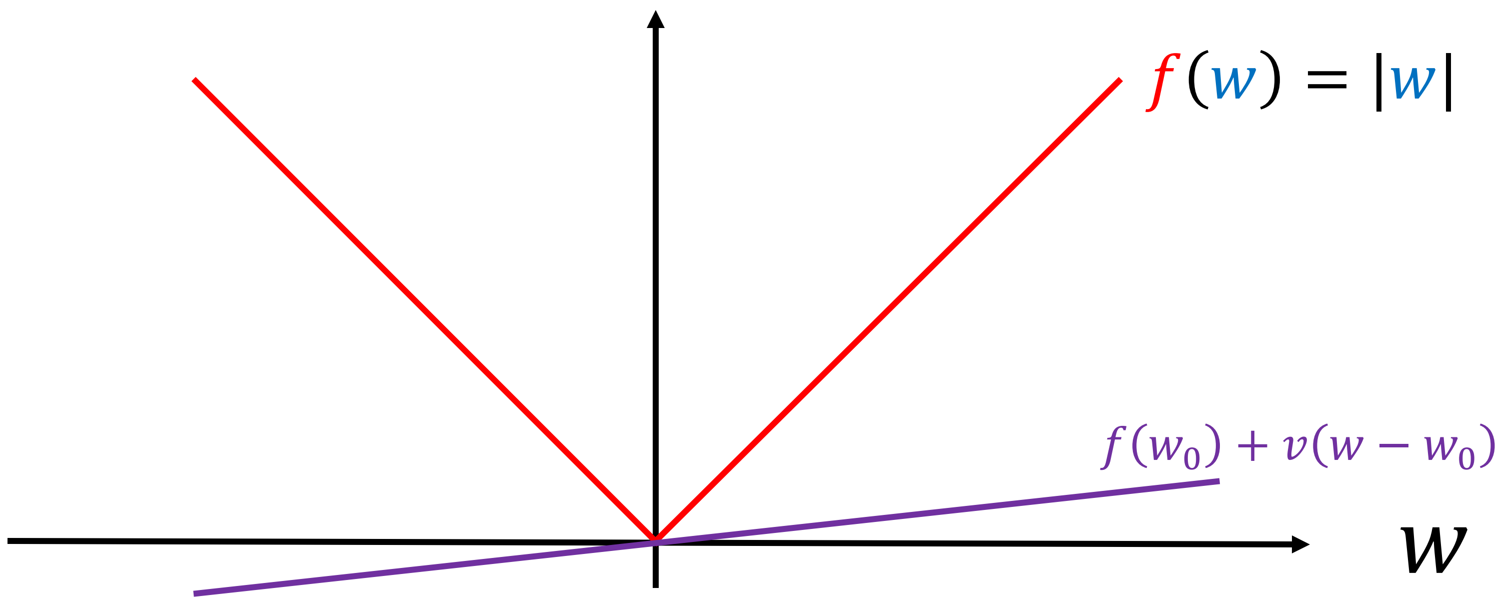
Subgradient of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.



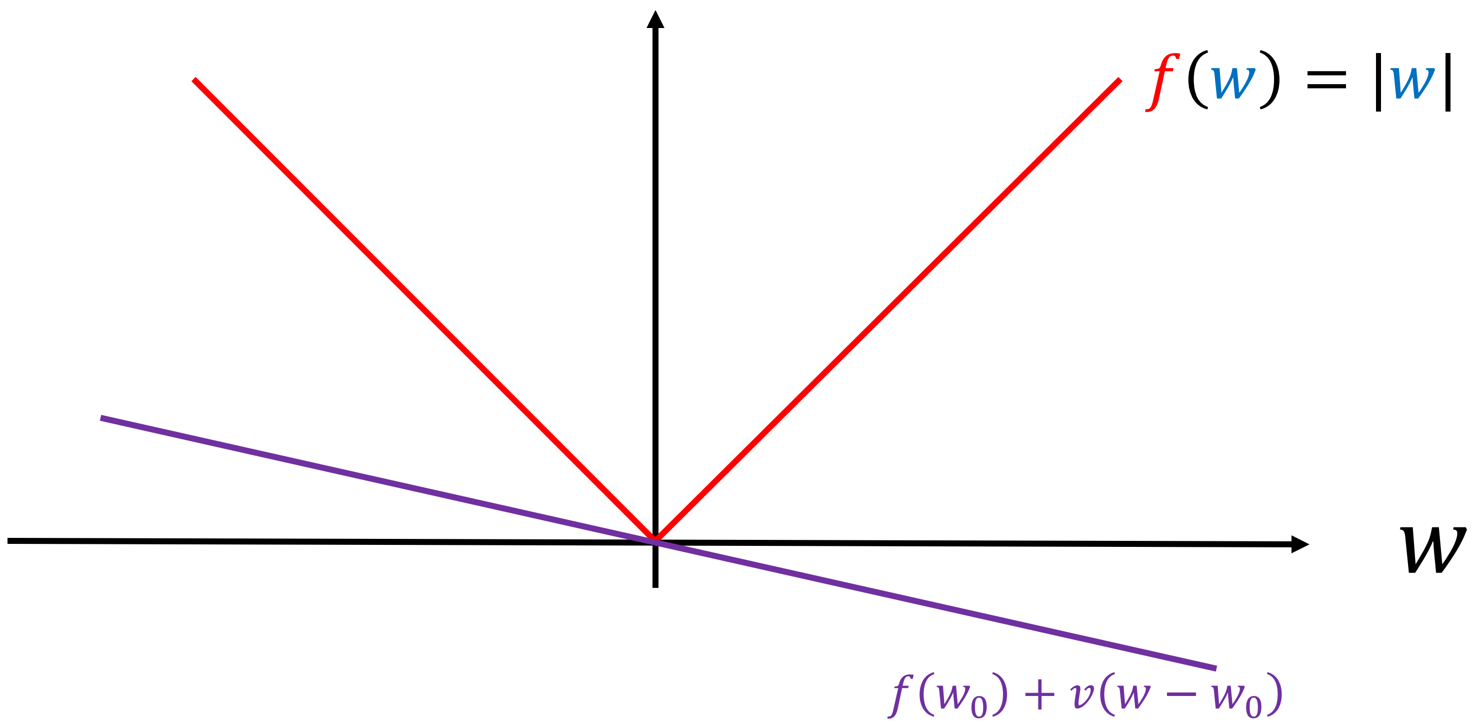
Subgradient of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.



Subgradient of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.



Subdifferential of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.

Definition (Subdifferential). The set containing all the subgradients of f at $\mathbf{w}_0$ is called the subdifferential. Denote the set by $\partial f(\mathbf{w}_0)$.

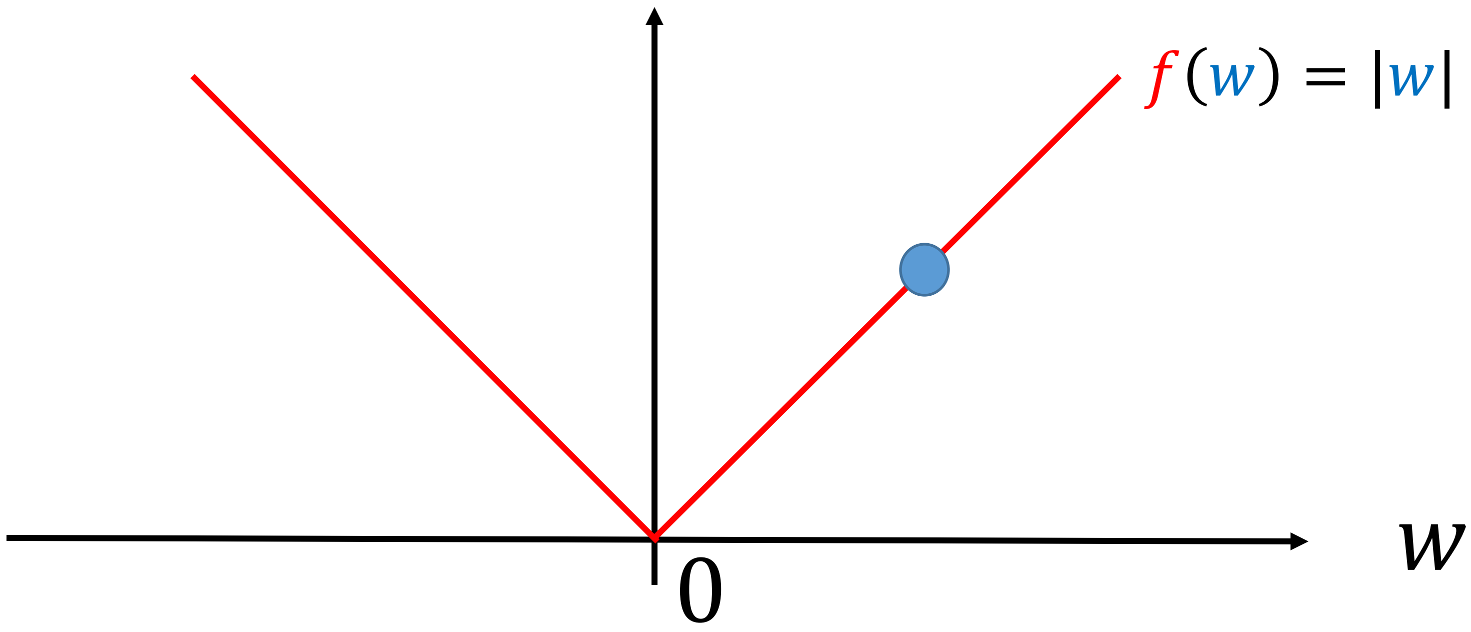
Subdifferential of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.

Definition (Subdifferential). The set containing all the subgradients of f at $\mathbf{w}_0$ is called the subdifferential. Denote the set by $\partial f(\mathbf{w}_0)$.

Example: $f(w) = |w|$

- $\partial f(3) = \{1\}$.



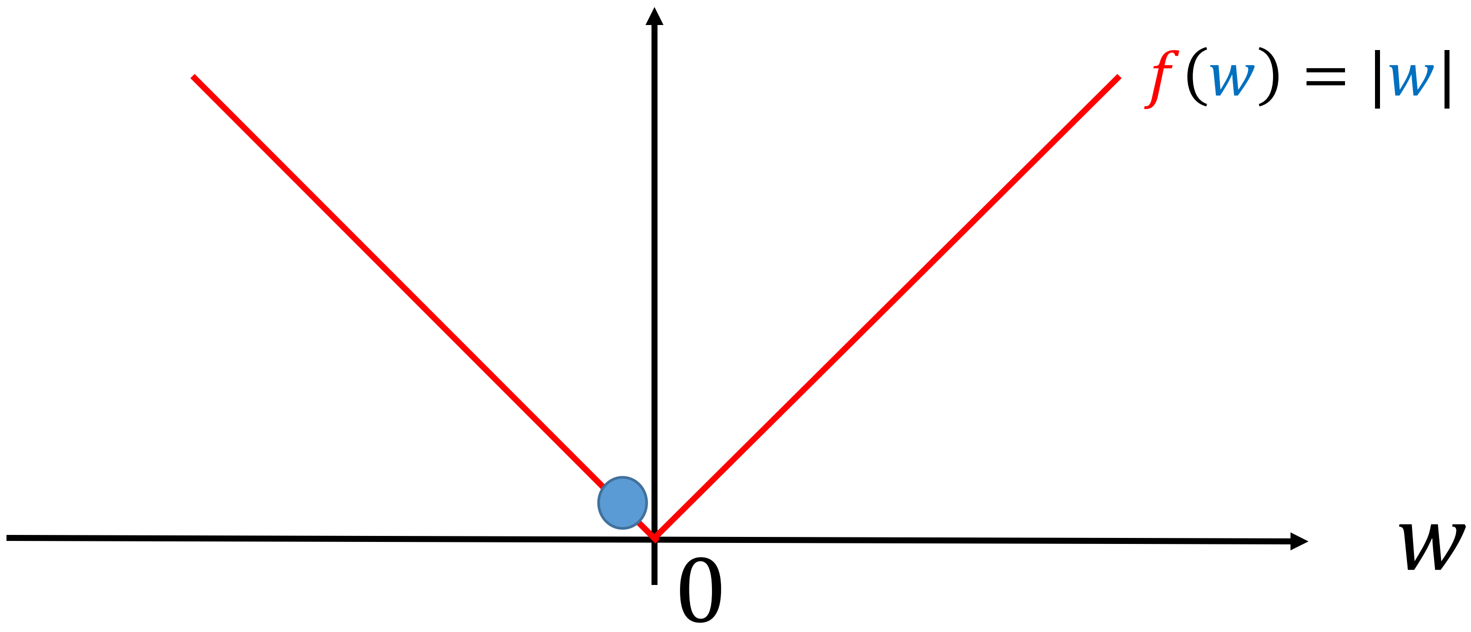
Subdifferential of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.

Definition (Subdifferential). The set containing all the subgradients of f at $\mathbf{w}_0$ is called the subdifferential. Denote the set by $\partial f(\mathbf{w}_0)$.

Example: $f(w) = |w|$

- $\partial f(3) = \{1\}$.
- $\partial f(-0.1) = \{-1\}$.



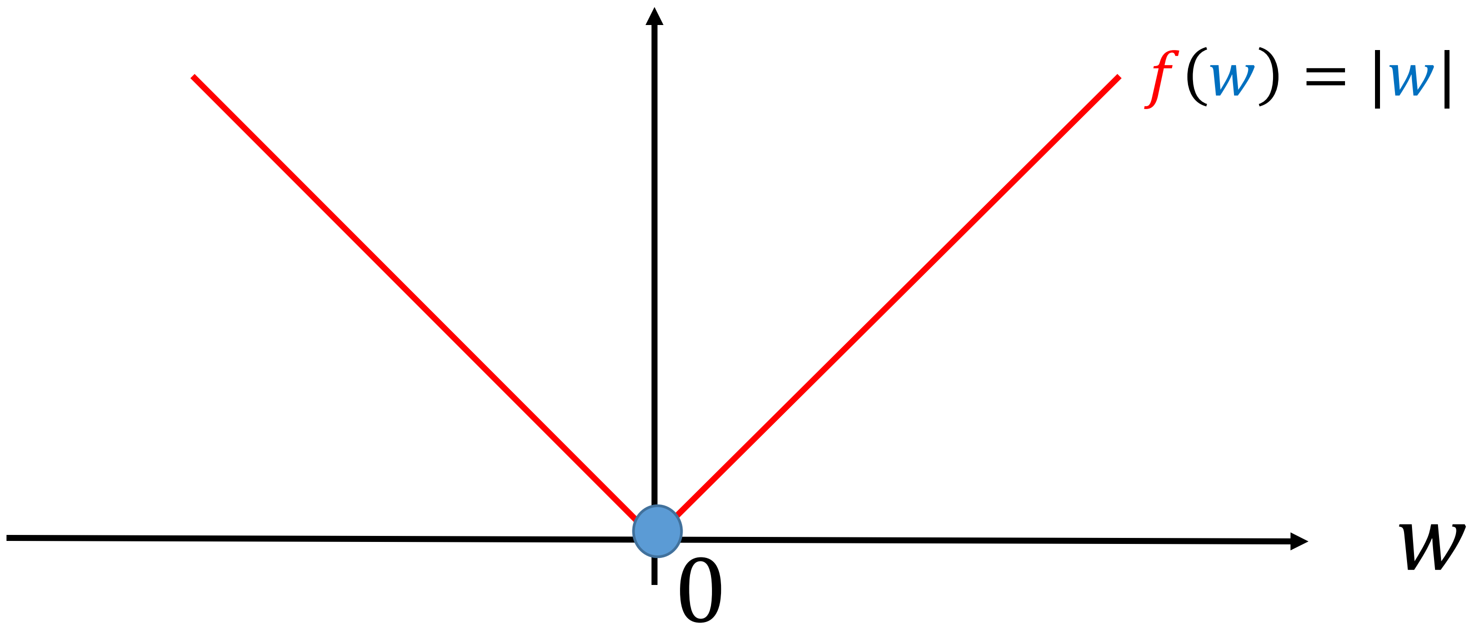
Subdifferential of Convex Function

Definition (Subgradient). A vector $\mathbf{v}$ is called a subgradient of f at $\mathbf{w}_0$ if for any $\mathbf{w}$, $f(\mathbf{w}) \geq f(\mathbf{w}_0) + \mathbf{v}^T(\mathbf{w} - \mathbf{w}_0)$.

Definition (Subdifferential). The set containing all the subgradients of f at $\mathbf{w}_0$ is called the subdifferential. Denote the set by $\partial f(\mathbf{w}_0)$.

Example: $f(w) = |w|$

- $\partial f(3) = \{1\}$.
- $\partial f(-0.1) = \{-1\}$.
- $\partial f(0) = [-1, 1]$.



A Property of Convex Optimization

Let f be a convex function.

Property: $\mathbf{w}^\star = \min_{\mathbf{w}} f(\mathbf{w}) \iff 0 \in \partial f(\mathbf{w}^\star)$.

Example: $\min_{w} \{f(w) = |w + 5|\}$

- $\partial f(-5) = [-1, 1]$.
- Obviously $0 \in \partial f(-5)$.
- $w^\star = -5$ minimizes f .