Convex Optimization

(1) Unconstrained opt.

$$\min_{x \in \mathbb{R}^n} f(x)$$

(CZ Ch6)

1st der. test: $Df(x) = 0$

$$Df(x) = (f_{x_1} \cdots f_{x_n})$$

$$\nabla f = \begin{pmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{pmatrix}$$

(2) Equality constrained opt.

(CZ. ch.20)

$$\min f(x)$$
$$\text{s.t. } \vec{h}(x) = 0$$

$$\vec{h}(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_p(x) \end{pmatrix}$$

Lagrange mult.

$$L(x, \lambda) = f(x) + \lambda^T \vec{h}(x)$$

$$\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix}$$

1st der. test: $Df(x) + \lambda^T D\vec{h} = 0$

$$D\vec{h} = \begin{pmatrix} Dh_1 \\ Dh_2 \\ \vdots \\ Dh_p \end{pmatrix}^{p \times n}$$

(3) Inequality constrained opt. (CZ ch 21)

$$\min \ f(x)$$
$$\text{s.t.} \quad \vec{h}(x) = 0 \qquad \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}$$
$$\vec{g}(x) \leq 0 \qquad \begin{pmatrix} g_1 \\ \vdots \\ g_q \end{pmatrix}$$

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x)$$

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_q \end{pmatrix} \geq 0$$

Find $x, \lambda, \mu \geq 0$ s.t.

$$\begin{cases} Df(x) + \lambda^T Dh(x) + \mu^T D\vec{g}(x) = \\ h(x) = 0 \\ g(x) \leq 0 \\ \mu^T g(x) = 0 \\ \mu \geq 0 \end{cases}$$

Karush-Kuhn-Tucker

KKT.

For (local) min, need 2^{nd} der. test:

$$H(x) = [D^2 f(x)] \geq 0 \quad \text{(unconstrained opt.)}$$

$H(x) = H^T(x)$

$(f_{x_i x_j} = f_{x_j x_i})$

(symmetric)

$$= \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{bmatrix}$$

$$H > 0 \iff \langle HV, V\rangle > 0 \quad \forall V \neq 0$$

$$\iff \text{eigenvalues of } H > 0$$

positive definite

$$H \geq 0 \iff \langle HV, V\rangle \geq 0 \quad \forall V$$

$$\iff \text{eigenvalues of } H \geq 0$$

semi-positive definite

$$\langle HV, V\rangle = V^T H V = \Big(\sum_{i,j}\Big) H_{ij} V_i V_j$$

Taylor series expansion

1D:

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

nD:

$$f(y) = f(x) + Df(x)(y-x) + \frac{1}{2}\langle D^2f(x)(y-x), y-x\rangle + \cdots$$

$$\sum_i f_{x_i}(x)(y_i - x_i) \qquad \frac{1}{2}\sum_{i,j} f_{x_i x_j}(x)(y_i - x_i)(y_j - x_j)$$

(Unconstrained Opt. CZ. Ch.6)

Theorem 6.2 ***Second-Order Necessary Condition (SONC).*** *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\mathbf{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\mathbf{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then*

$$\mathbf{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\mathbf{d} \geq 0,$$

where $\boldsymbol{F}$ is the Hessian of f. □

Corollary 6.2 ***Interior Case.*** *Let $\boldsymbol{x}^*$ be an interior point of $\Omega \subset \mathbb{R}^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, then*

$$\nabla f(\boldsymbol{x}^*) = \mathbf{0},$$

and $\boldsymbol{F}(\boldsymbol{x}^*)$ is positive semidefinite ($\boldsymbol{F}(\boldsymbol{x}^*) \geq 0$); that is, for all $\mathbf{d} \in \mathbb{R}^n$,

$$\mathbf{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\mathbf{d} \geq 0.$$

Theorem 6.3 ***Second-Order Sufficient Condition (SOSC), Interior Case.*** *Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that*

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f. □

Equality Constrained Opt. (CZ. Ch.20)

Theorem 20.4 ***Second-Order Necessary Conditions.*** *Let $\boldsymbol{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \le n$, and $f, \boldsymbol{h} \in \mathcal{C}^2$. Suppose that $\boldsymbol{x}^*$ is regular. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:*

1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.
2. *For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} \ge 0$.* □

Theorem 20.5 ***Second-Order Sufficient Conditions.*** *Suppose that $f, \boldsymbol{h} \in \mathcal{C}^2$ and there exists a point $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:*

1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.
2. *For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \ne \boldsymbol{0}$, we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$.*

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$. □

Inequality Constrained Opt. (CZ. Ch.21)

Theorem 21.2 ***Second-Order Necessary Conditions.*** *Let $\boldsymbol{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \le n$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$, and $f, \boldsymbol{h}, \boldsymbol{g} \in \mathcal{C}^2$. Suppose that $\boldsymbol{x}^*$ is regular. Then, there exist $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that:*

1. $\boldsymbol{\mu}^* \ge \boldsymbol{0}$, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$, $\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*) = 0$.
2. *For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} \ge 0$.* □

Theorem 21.3 ***Second-Order Sufficient Conditions.*** *Suppose that $f, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{C}^2$ and there exist a feasible point $\boldsymbol{x}^* \in \mathbb{R}^n$ and vectors $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that:*

1. $\boldsymbol{\mu}^* \ge \boldsymbol{0}$, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$, $\boldsymbol{\mu}^{*\top}\boldsymbol{g}(\boldsymbol{x}^*) = 0$.
2. *For all $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, $\boldsymbol{y} \ne \boldsymbol{0}$, we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{y} > 0$.*

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$. □

There is an important case the 2^{nd} der. test is not needed.

Convex Optimization

[CZ, ch. 22]

$$\min_{x \in \Omega} f(x) \leftarrow \text{convex function}$$

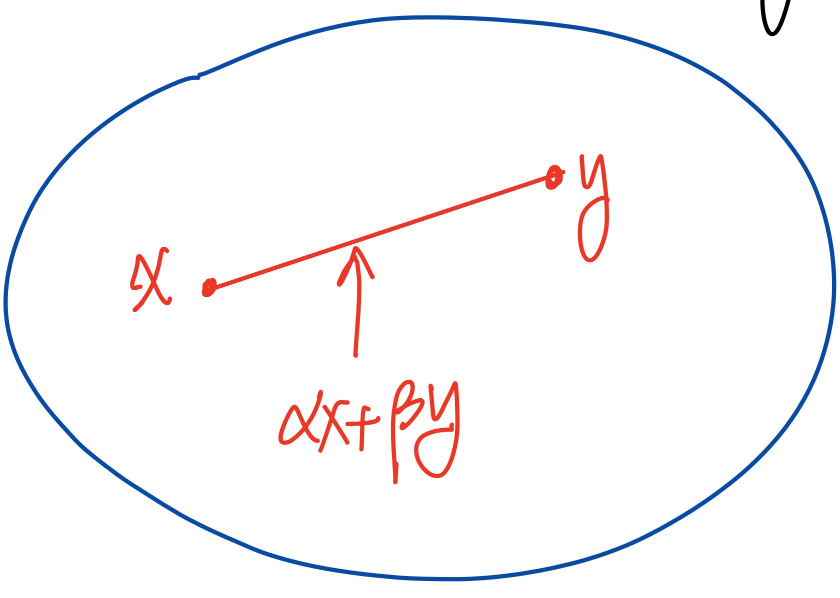
$x \in \Omega \leftarrow$ convex set.

(1) $\Omega \subseteq \mathbb{R}^n$ is convex if

$$\forall\, x, y \in \Omega \implies \alpha x + \beta y \in \Omega, \quad \alpha, \beta \geq 0, \ \alpha + \beta = 1$$

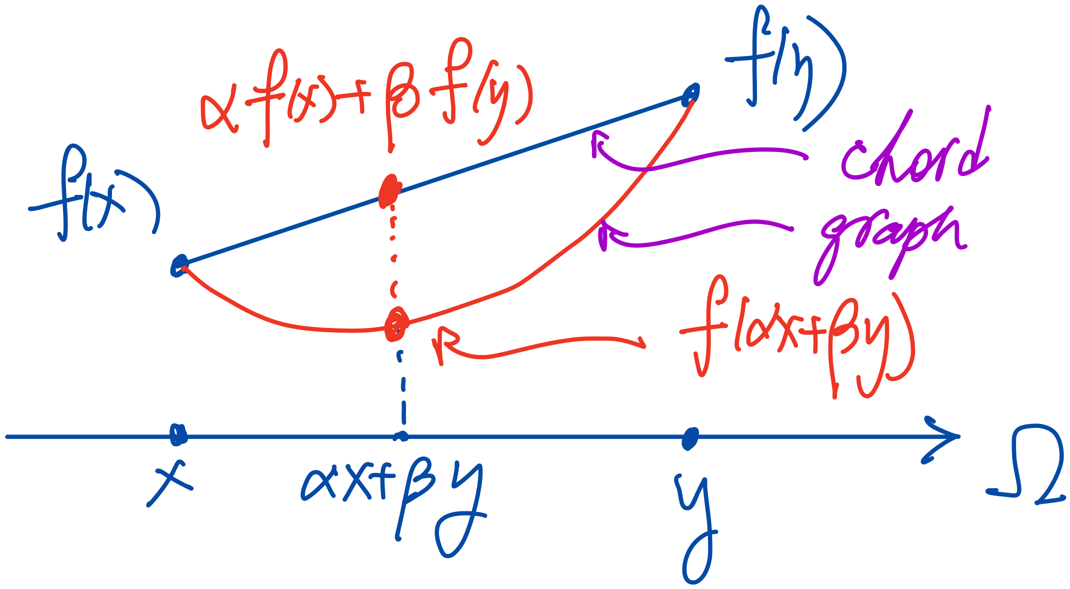
for all

$$(\lambda x + (1-\lambda) y)$$



(2) $f : \Omega \longrightarrow \mathbb{R}$ is convex if

$\forall x, y \in \Omega$, $f(\alpha x + \beta y) \overset{(*)}{\leq} \alpha f(x) + \beta f(y)$



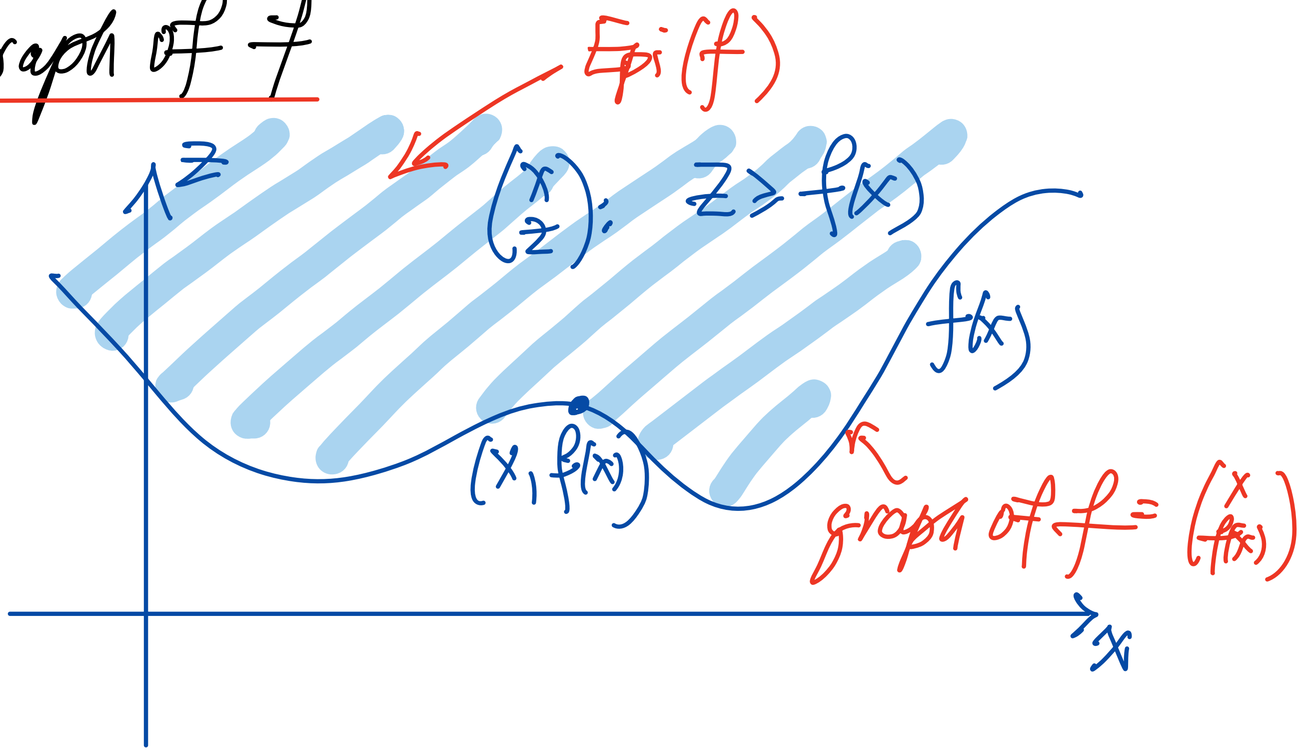
(The graph of f lies below the chord)

f is strictly convex if "=" holds in (*) only when $\beta = 0$ (at x) or $\alpha = 0$ (at y)

Various Equivalent Def. of Convexity of f

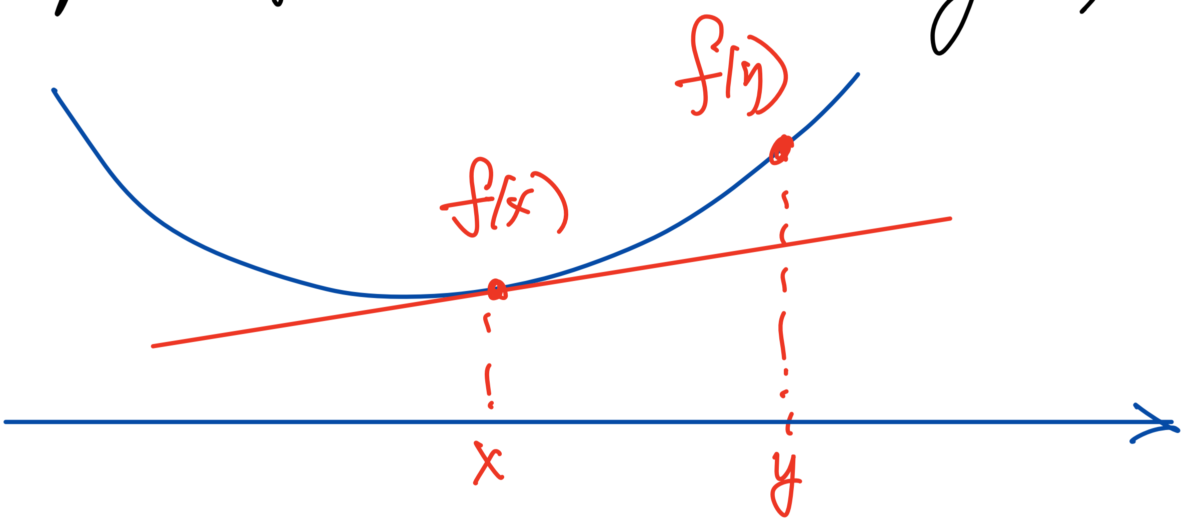
(1) $$\mathrm{Epi}(f) := \left\{ \binom{x}{z} : \quad z \geq f(x) \right\}$$

Epigraph of f



f is a convex function $\Leftrightarrow$ Epi(f) is a convex set

(2) Graph of f lies above its tangent plane



$$\forall x, y \in \Omega: \quad \boxed{f(y) \geq f(x) + Df(x)(y-x)}$$

(3) $\left[D^2 f(x)\right] \geq 0 \quad \forall x \in \Omega$

Hessian H positive semi-definite

$$H \geq 0 \iff \langle HV, V \rangle \geq 0 \quad \forall V$$

<u>Quadratic function:</u>

$$\boxed{f(x) = \frac{1}{2} x^T A x + B^T x}$$

$A = A^T$, symmetric

$$Df(x) = x^T A + B^T$$

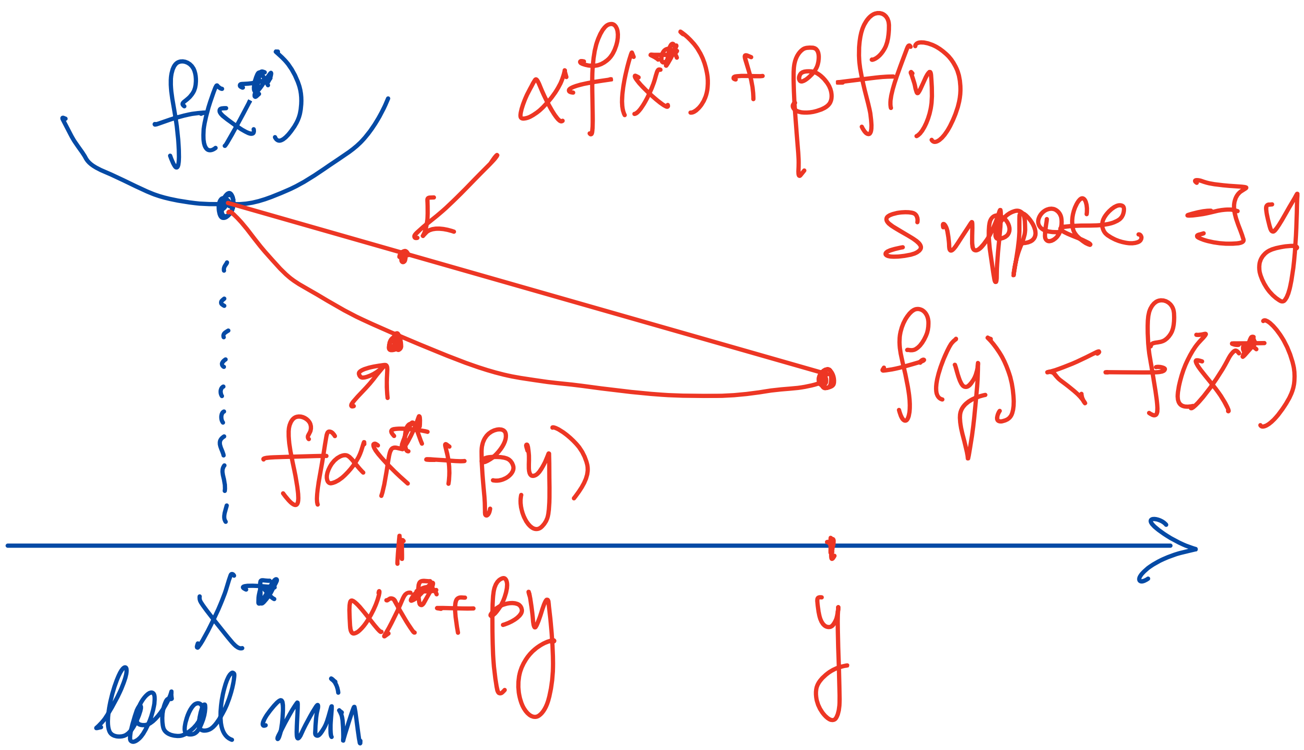
$$D^2 f(x) = A$$

Hence <u>f is convex $\iff A \geq 0$</u>

[CZ]

Theorem 22.6 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on a convex set $\Omega \subset \mathbb{R}^n$. Then, a point is a global minimizer of f over Ω if and only if it is a local minimizer of f.* $\square$

In particular, local min $\Rightarrow$ global min.



$$f(\alpha x^* + \beta y) \leq \alpha f(x^*) + \beta f(y) < f(x^*)$$

let $\beta \to 0$ (i.e. $\alpha \to 1$)

[CZ]

Lemma 22.2 *Let $f : \Omega \to \mathbb{R}$ be a convex function defined on the convex set $\Omega \subset \mathbb{R}^n$ and $f \in \mathcal{C}^1$ on an open convex set containing Ω. Suppose that the point $\boldsymbol{x}^* \in \Omega$ is such that for all $\boldsymbol{x} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^*$, we have*

$$Df(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Pf

$$f(y) \geq f(x^*) + \underbrace{Df(x^*)(x - x^*)}_{\geq 0}$$

$$\geq f(x^*)$$

Corollary 22.2 *Let $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function defined on the convex set $\Omega \subset \mathbb{R}^n$. Suppose that the point $\boldsymbol{x}^* \in \Omega$ is such that*

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Constrained Optimization

Theorem 22.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points*

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\},$$ equality constraint

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top.$$

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □

Theorem 22.9 *Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points*

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$, $\boldsymbol{h}, \boldsymbol{g} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, and $\boldsymbol{\mu}^* \in \mathbb{R}^p$, such that

1. $\boldsymbol{\mu}^* \ge \boldsymbol{0}$.
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*\top} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.
3. $\boldsymbol{\mu}^{*\top} \boldsymbol{g}(\boldsymbol{x}^*) = 0$.

Then, $\boldsymbol{x}^*$ is a global minimizer of f over Ω. □