Week 1, Lectures 2-3

Motivation and Applications again (§ 1.2)

<u>Example</u>: A chemical company has the following set-up:

- 2 factories F_1 and F_2
- 12 retail outlets $R_1, R_2, \ldots, R_{12}$

Furthermore,

- Each factory F_i can produce a_i tons of product per week (<u>capacity</u>)
- Each retail outlet has a demand b_j tons of product.

c_{ij} = cost of shipping 1 ton of product from factory i to outlet j.

x_{ij} = # tons shipped i to j.

Total cost:

$$f = \sum_{i=1}^{2} \sum_{j=1}^{12} x_{ij} c_{ij}$$

<u>Constraints</u>

<u>Capacity</u> $\displaystyle\sum_{j=1}^{12} x_{ij} \leq a_i$

Demand $\sum_{i=1}^{2} x_{ij} \begin{cases} = b_j \\ \geq b_j \end{cases}$

Positivity $x_{ij} \geq 0 \quad i = 1, 2; \ j = 1, \dots, 12.$

We can reshape the unknowns x_{ij} into a ~~matrix~~ vector $\underline{x} \in \mathbb{R}^{24}$. The problem is now linear in $\underline{x}$. **LINEAR PROGRAMMING.**

Continuous versus discrete optimization

Often, instead of $\underline{x} \in \mathbb{R}^n$ we can have $\underline{x} \in \mathbb{N}^n$ or $\underline{x} \in \{0,1\}^n$.

Examples:

- Power grid n different types of powerplant (wind, solar, nuclear, ...)

$$PP1, \ PP2, \ \dots \ PPn.$$

$$\underline{x} = (\dots, \underbrace{\dots}_{\in \mathbb{N}_0}, \dots, \dots)$$

$$\underline{x} \in \mathbb{N}_0^n.$$

- Putting a factory in city i:

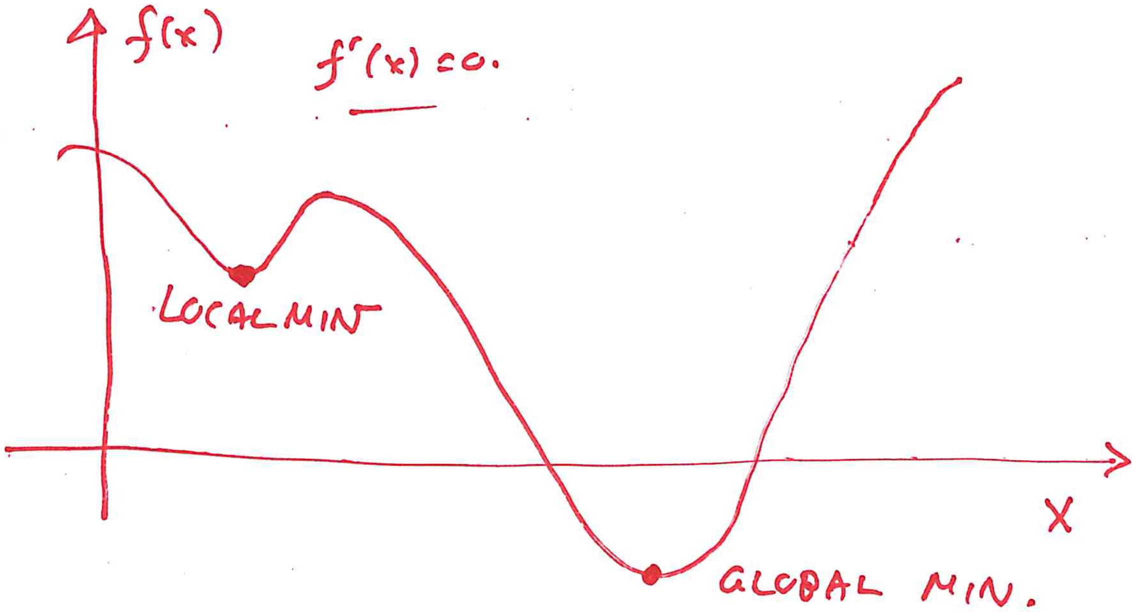
City 1, .., city n

Y/N $(\underset{0 \text{ or } 1}{\ } , \ldots , \underset{0 \text{ or } 1}{\ })$.

$\underline{x} \in \{0,1\}^n$.

We focus in this module on <u>continuous optimization</u>.

Global versus local optimization.

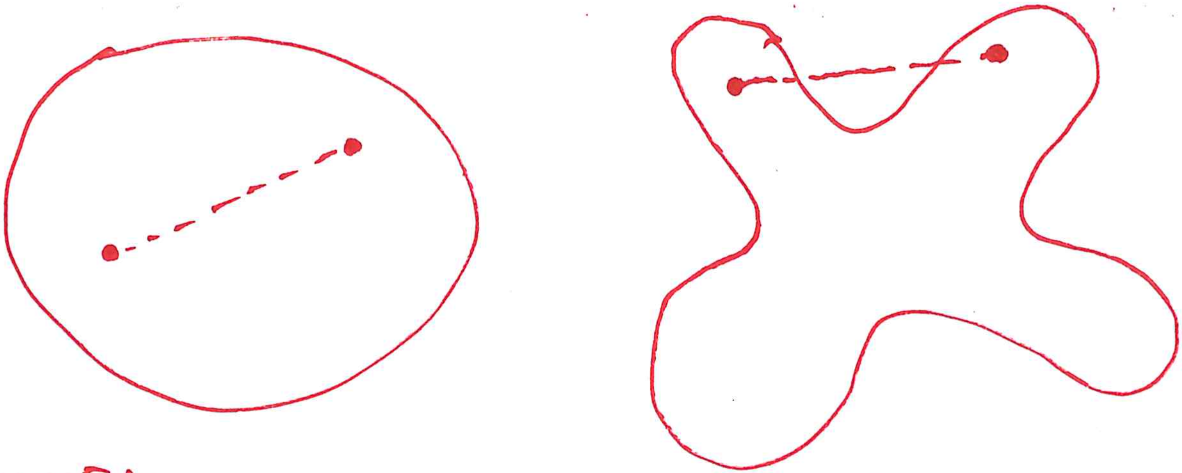


- Linesearch / Trust-Region / Newton will find a local min.
- Global optimizers (e.g. simulated annealing) will find the global min.

For <u>convex functions</u> there is a unique minimum.

§ 1.3 Convex sets

<u>Def$^{\underline{n}}$</u>: Let $S \subset \mathbb{R}^n$. S is called convex if ~~any~~ a straight-line segment joining any two points in S, is ~~entire~~ entirely contained in S.



CONVEX. NOT CONVEX.

S is convex, if, given any $\underline{x}$ and $\underline{y}$ in S, the line segment

$$\underline{x}(t) = \underline{x}t + (1-t)\underline{y} \qquad t \in [0,1]$$

is contained entirely in S.

Exam §1.3

<u>Examples</u>: $\boxed{\text{Unit Ball}}$

Dot-product notation:

$$\langle \underline{x}, \underline{y} \rangle = \sum_{i=1}^{n} x_i y_i$$

$$\| \underline{x} \|_2^2 = \langle \underline{x}, \underline{x} \rangle = \sum_{i=1}^{n} x_i^2$$

Theorem 1.1 The unit ball

$$B = \{ \underline{x} \in \mathbb{R}^n \mid \|\underline{x}\|_2^2 \le 1 \}$$

is a convex set.

<u>Proof:</u> $\cancel{\underline{x}(t)}$ Let $\underline{x}$ and $\underline{y} \in B$.

Take the line segment

$$\underline{x}(t) = \underline{x}t + (1-t)\underline{y}, \quad t \in [0,1]$$

Show: $\|\underline{x}(t)\|_2^2 \le 1 \quad \forall\, t \in [0,1]$.

$$\begin{aligned}
\|\underline{x}(t)\|_2^2 &= \langle \underline{x}(t), \underline{x}(t) \rangle \\
&= \langle \underline{x}t + (1-t)\underline{y}, \underline{x}t + (1-t)\underline{y} \rangle \\
&= t^2 \langle \underline{x}, \underline{x} \rangle + 2t(1-t)\langle \underline{x}, \underline{y} \rangle \\
&\quad + (1-t)^2 \langle \underline{y}, \underline{y} \rangle \\
&= t^2 \underbrace{\|\underline{x}\|_2^2}_{=1} + 2t(1-t)\langle \underline{x}, \underline{y} \rangle + (1-t)^2 \underbrace{\|\underline{y}\|_2^2}_{1} \\
&= t^2 + 2t(1-t)\langle \underline{x}, \underline{y} \rangle + (1-t)^2 \\
&\overset{C.S.}{\le} t^2 + 2t(1-t) \underbrace{\|\underline{x}\|_2}_{\le 1} \underbrace{\|\underline{y}\|_2}_{\le 1} + (1-t)^2 \\
&= t^2 + 2t(1-t) + (1-t)^2 \\
&= \cancel{t^2} + \cancel{2t} - \cancel{2t^2} + 1 - \cancel{2t} + \cancel{t^2}. \\
&= 1
\end{aligned}$$

$$\Longrightarrow \|\underline{x}(t)\|_2^2 \le 1 \quad \forall\, t \in [0,1].$$

Hence, $\underline{x}(t) \in B \quad \forall \ t \in [0,1]$.

So B is convex.

<u>Polyhedra</u> :

$$S = \{ \underline{x} \in \mathbb{R}^n \mid A\underline{x} = b, \ C\underline{x} \leq d \}$$

where A and B are matrices of appropriate dimension and b and $\underline{d}$ are vectors.

$C\underline{x} \leq d$ means: $[C\underline{x}]_i \leq d_i$.

S is a convex set .

<u>Convex Function</u> :

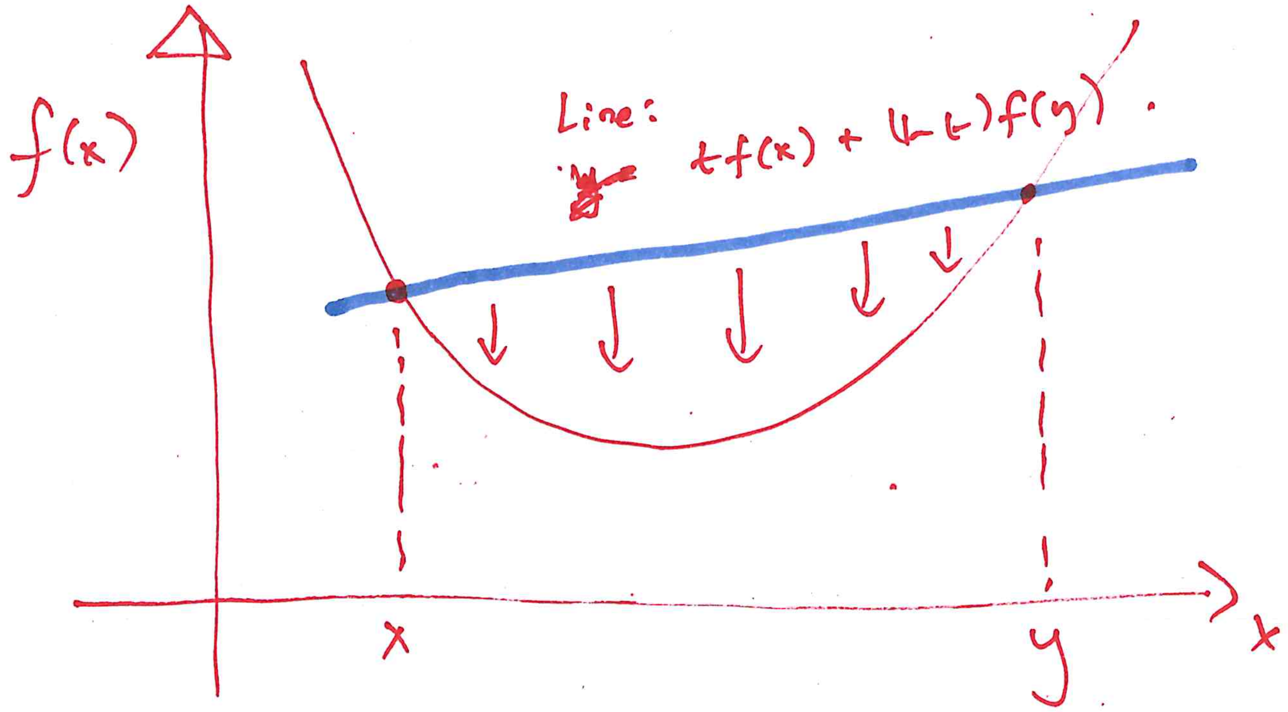
Let $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$.

The function f is convex if :

- S is a convex set
- The following relation holds :

$$f(t\underline{x} + (1-t)\underline{y}) \leq t\, f(\underline{x}) + (1-t) f(\underline{y})$$

for all $t \in [0,1]$ and all $\underline{x}, \underline{y} \in S$.



<u>Examples:</u>

- Let $S = \mathbb{R}^n$. Then the linear function
$$f(\underline{x}) = \langle \underline{b}, \underline{x} \rangle + \alpha$$
is convex. Here, $\underline{b}$ is a const. vector and α is a constant. CHECK!

- Let $S = \mathbb{R}^n$. Then the following quadratic function is convex:
$$f(\underline{x}) = \langle \underline{x}, H\underline{x} \rangle \qquad (1)$$
Here, H is symmetric semi-definite.

<u>Theorem</u>: The function in (1) is convex.

Proof: Let $\underline{x}$ and $\underline{y}$ be any two points in $\mathbb{R}^n$.

Look at:

$$f(t\underline{x} + (1-t)\underline{y})$$
$$= \langle t\underline{x} + (1-t)\underline{y}, H[t\underline{x} + (1-t)\underline{y}] \rangle$$
$$= t^2 \langle \underline{x}, H\underline{x} \rangle + t(1-t)\langle \underline{x}, H\underline{y} \rangle$$
$$+ (1-t)t \langle \underline{y}, H\underline{x} \rangle$$
$$+ (1-t)^2 \langle \underline{y}, H\underline{y} \rangle .$$

Look at: $\langle \underline{y}, H\underline{x} \rangle = \langle H\underline{x}, \underline{y} \rangle$
$$= \langle \underline{x}, H^T\underline{y} \rangle$$
$$\overset{\text{symmetric}}{=} \langle \underline{x}, H\underline{y} \rangle .$$

Hence:

$$f(t\underline{x} + (1-t)\underline{y})$$
$$= t^2 \langle \underline{x}, H\underline{x} \rangle + 2t(1-t)\langle \underline{x}, H\underline{y} \rangle$$
$$+ (1-t)^2 \langle \underline{y}, H\underline{y} \rangle \quad (*)$$

Also, look at:

$$t f(\underline{x}) + (1-t) f(\underline{y})$$
$$= t\langle \underline{x}, H\underline{x} \rangle + (1-t)\langle \underline{y}, H\underline{y} \rangle \quad (**)$$

Take $(*) - (**)$: $\quad t^2 - t = t(t-1)$

$$t^2 \langle \underline{x}, H\underline{x} \rangle + 2t(1-t) \langle \underline{x}, H\underline{y} \rangle + (1-t)^2 \langle \underline{y}, H\underline{y} \rangle - t \langle \underline{x}, H\underline{x} \rangle - (1-t) \langle \underline{y}, H\underline{y} \rangle$$

$$= (t^2 - t) \langle \underline{x}, H\underline{x} \rangle + 2t(1-t) \langle \underline{x}, H\underline{y} \rangle + \left[(1-t)^2 - (1-t) \right] \langle \underline{y}, H\underline{y} \rangle$$

$$= -t(1-t) \langle \underline{x}, H\underline{x} \rangle + 2t(1-t) \langle \underline{x}, H\underline{y} \rangle + (1-t) \left[\cancel{1} - t - \cancel{1} \right] \langle \underline{y}, H\underline{y} \rangle$$

$$= -t(1-t) \left[\langle \underline{x}, H\underline{x} \rangle - 2 \langle \underline{x}, H\underline{y} \rangle + \langle \underline{y}, H\underline{y} \rangle \right]$$

$$= \underbrace{-\underbrace{t(1-t)}_{\text{pos. or zero}} \left[\underbrace{\langle (\underline{x} - \underline{y}), H(\underline{x} - \underline{y}) \rangle}_{\text{pos. or zero.}} \right]}_{\text{neg. or zero}}$$

$$\leq 0.$$

Hence, $(*) - (**) \le 0$, or

$(*) \le (**)$

Hence,

$$f(t\underline{x} + (1-t)\underline{y}) \le t f(\underline{x}) + (1-t) f(\underline{y}) \quad \forall\, t \in [0,1].$$

Hence, f is convex. ∎

Tuesday lecture will be a recorded online lecture.

<u>OP</u>: $\min\limits_{\underline{x} \in S} f(\underline{x})$, $S \subset \mathbb{R}^n$.

Write the solution as $\underline{x}_*$:

$$\underline{x}_* = \arg\min_{\underline{x} \in S} f(\underline{x}). \quad (2)$$

<u>Definition</u>: $\underline{x}_*$ is a <u>global minimizer</u> if

$$f(\underline{x}_*) \le f(\underline{y}) \quad \forall\, \underline{y} \in S.$$

$\underline{x}_*$ is a <u>local minimizer</u> if there exists a neighbourhood $\mathcal{N} \subset S$ such that $\underline{x}_* \in \mathcal{N}$, and such that

$$f(\underline{x}_*) \leq f(\underline{y}) \quad \forall \underline{y} \in \mathcal{N}.$$

Here, a neighbourhood means a non-empty open set.

$\underline{x}_*$ is a strict local minimizer if there exists a ~~A~~ neighbourhood $\mathcal{N} \subset S$ such that $\underline{x}_* \in \mathcal{N}$, and such that

$$f(\underline{x}_*) < f(\underline{y}) \quad \text{for all } \underline{y} \neq \underline{x}_* \in \mathcal{N}.$$

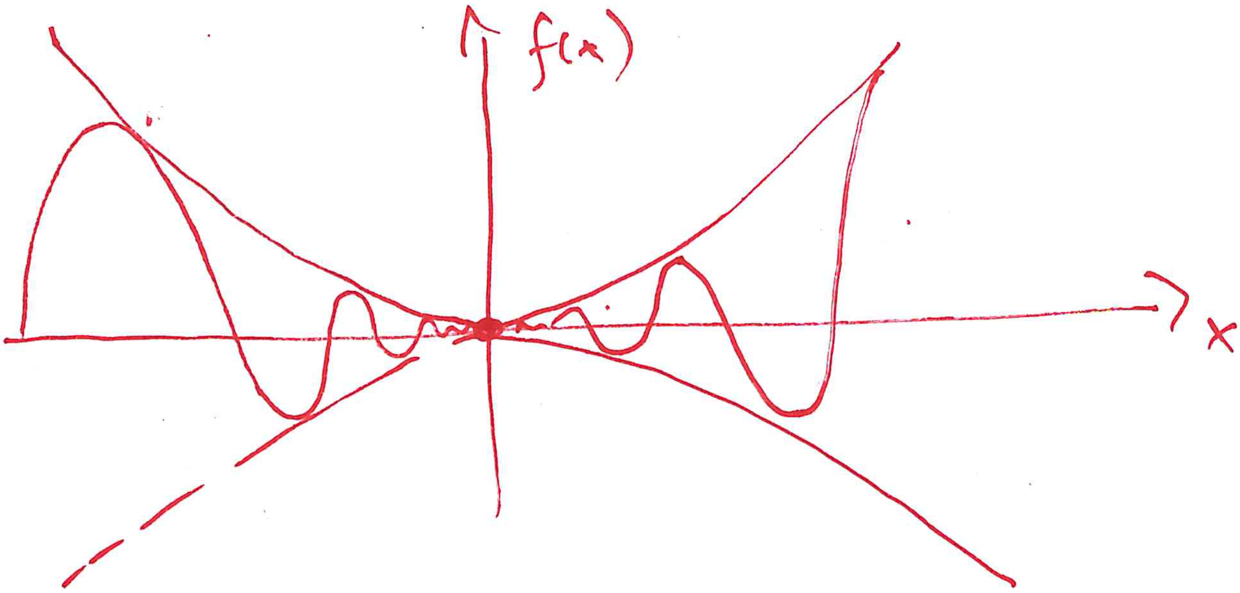
<u>Examples</u>:

$f(\underline{x}) = 1$, $\underline{x} \in \mathbb{R}^n$. Then every $\underline{x}$ is a local minimizer. But if I take $n = 1$, and $f(x) = (x-2)^4$, then $x_* = 2$ is a strict local minimizer.

$\underline{x}_*$ is an isolated local minimizer if there exists a neighbourhood $\mathcal{N}$ of $\underline{x}_*$ such that $\underline{x}_*$ is the only local minimizer in $\mathcal{N}$.

Pathological example:

$$f(x) = \begin{cases} x^4 \cos(1/x) + 2x^4, & x \neq 0. \\ 0 & x = 0. \end{cases}$$



There will always be a local min. ~~arbitra~~ arbitrarily close to the local min at $x=0$, so $x_* = 0$ is not isolated.

§ 2.2. Necessary conditions for optimality when f is differentiable ($n=1$).

Taylor's Remainder Theorem: Let $f: \mathbb{R} \to \mathbb{R}$ be continuous with continuous first derivative. Assume $h > 0$. Then there exists $\eta \in (0,h)$ such that

$$f(x+h) = f(x) + f'(x+\eta)h.$$

Furthermore, if f is twice continuously differentiable, then there exists $\eta \in (0, h)$ such that:

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2} f''(x+\eta) h^2 .$$

$$\boxed{f(x+h) \simeq f(x) + f'(x)h + \frac{1}{2} f''(x) h^2}$$

Theorem 2.2 (First-order necessary condition). Let x_* solve the OP (2). Then $f'(x_*) = 0$.

Proof by contradiction. Suppose that $f'(x_*) \neq 0$.

Look at $f'(x_*) < 0$. Then by continuity, there exists a $\delta > 0$ such that

$$f'(x) < 0 \text{ for all } x \in \underbrace{(x_* - \delta, x_* + \delta)}_{I} .$$

Look at $x = x_* + h$, where $|h| < \delta$.

Apply Taylor's Remainder Theorem:

$$f(\overset{x_*+h}{x}) = f(x_*) + f'(x_* + \eta) h ,$$

where $|\eta| < |h| < \delta$.

Since $x_* + \eta \in I$, $f'(x_* + \eta) < 0$.

Take $h = -k f'(x_* + \eta)$, $k > 0$.

Hence,

$$f(x) = f(x_*) - k\,|f'(x_*+\eta)|^2$$

Hence,

$$f(x) < f(x_*).$$

Contradiction, since x_* is the local minimizer.

The calculation for $f'(x_*) > 0$ is the same — a contradiction arises in both cases, hence:

$$f'(x_*) = 0.$$

Remark: The result for $n > 1$ is analogous:

$$\underline{\nabla} f(x_*) = 0.$$

Theorem: (Second-order necessary condition)

Let x_* solve the OP (2), with $n = 1$.

Then $f''(x_*) \geq 0$.

Proof by contradiction. Assume that $f''(x_*) < 0$.

By continuity, there exists a $\delta > 0$ such that

$$f''(x) < 0 \quad \text{for all } x \in \underbrace{(x_* - \delta, x_* + \delta)}_{I}.$$

Consider $x = x_* + h$, with $|h| < \delta$.

$$f(x) = f(x_*) + \cancel{f'(x_*)h} + \frac{1}{2} f''(x_*+\eta)h^2,$$

where $|\eta| < |h| < \delta$, by Taylor's Remainder Theorem.

$$\Rightarrow f(x) = f(x_*) + \frac{1}{2} \underbrace{f''(x_* + \eta) h^2}_{\text{neg.}}.$$

$$\Rightarrow f(x) < f(x_*) \quad \text{contradiction.}$$

Hence, $\boxed{f''(x_*) \geq 0}$ ∎

<u>Remark</u>: In n-dimensions, the second-order necessary condition is that the Hessian matrix,

$$H_{ij} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{x_*}$$

is positive-semi-definite.

<u>Reminder</u>: $\langle \underline{\xi}, H \underline{\xi} \rangle \geq 0$

for all $\underline{\xi} \in \mathbb{R}^n$.

<u>Theorem</u>: Second-order sufficient condition: Suppose that f is twice differentiable with continuous second derivative. If $f'(x_*) = 0$ and $f''(x_*) > 0$ (strict), then x_* is a strict local minimizer of f.

<u>Proof</u>: Idea the same as before (p.14).

Extension to n- dimensions:

<u>Theorem 2.7</u> Suppose that f is twice differentiable with continuous second derivative. If $\nabla f(x_*) = 0$ and if the Hessian matrix H is <u>positive-definite</u>, then x_* is a strict local minimizer.

<u>Reminder</u>: H is positive-definite if

$$\langle \underline{\xi}, H\underline{\xi} \rangle > 0 \quad \forall \ \underline{\xi} \neq 0 \in \mathbb{R}^n.$$

<u>Remark</u>: Insisting on H being ~~~~ strictly positive definite rules out saddle points and degenerate critical points ($f(x) = x^4$).

For convex functions the situation is much nicer:

<u>Theorem</u>: When f is convex, any local minimizer ~~~~ x_* is a global minimizer. If f is differentiable, then any stationary point ($\underline{\nabla f = 0}$) is a global minimizer.

EXAM

Part 1. Show that a local minimizer really is a global minimizer. Proof by contradiction. Suppose that $\underline{x}_*$ is a local minimizer, but that there exists $\underline{y} \in S$ such that

$$f(\underline{y}) < f(\underline{x}_*).$$

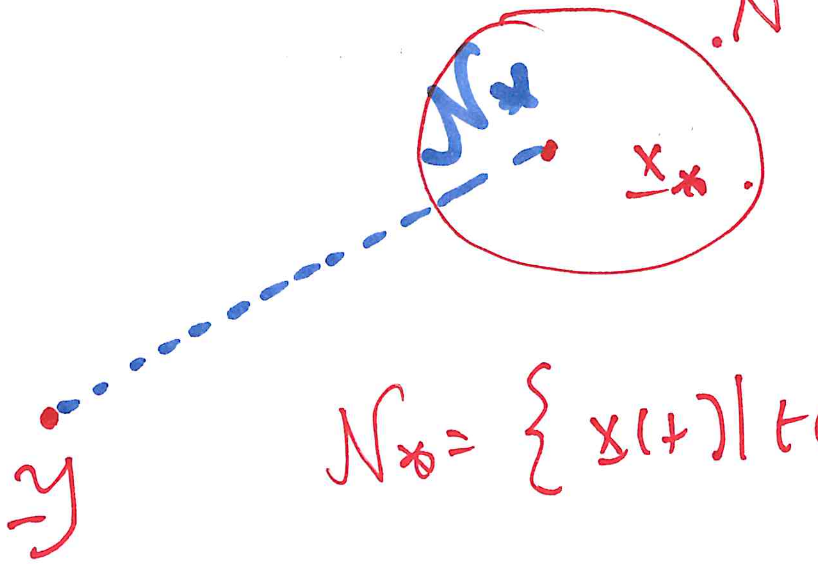
We construct the line segment

$$\underline{x}(t) = t\underline{y} + (1-t)\underline{x}_*, \quad t \in [0,1]$$

By convexity:

$$\begin{aligned} f(\underline{x}(t)) &\leq t f(\underline{y}) + (1-t) f(\underline{x}_*) \\ &\leq t f(\underline{x}_*) + (1-t) f(\underline{x}_*) \\ &= f(\underline{x}_*). \end{aligned}$$

$$\Longrightarrow f(\underline{x}(t)) \leq f(\underline{x}_*).$$



$$\mathcal{N}_* = \{\underline{x}(t) \mid t \in [0,1]\} \cap \mathcal{N}.$$

We have:

$$f(\underline{x}(t)) \leq f(\underline{x}_*) \quad \forall \; \underline{x}(t) \in N_*.$$

So no matter the size of N, there are points in N such that

$$f(\cdots) \leq f(\underline{x}_*).$$

This is a contradiction. Hence,

$$f(\underline{y}) \geq f(\underline{x}_*) \quad \forall \; \underline{y} \in S.$$

Hence, $\underline{x}_*$ is a global min.

<u>Part 2.</u> Suppose that $\underline{x}_*$ is a critical point, $\nabla f(\underline{x}_*) = 0$.

$\lim_{t \to 0}$.

$$0 = (\underline{y} - \underline{x}_*) \cdot \nabla f(\underline{x}_*)$$

$$= \frac{d}{dt} f(\underline{x}_* + t(\underline{y} - \underline{x}_*)) \Big|_{t=0}.$$

$$= \lim_{t \downarrow 0} \frac{f(\underline{x}_* + t(\underline{y} - \underline{x}_*)) - f(\underline{x}_*)}{t}$$

$$= \lim_{t \downarrow 0} \frac{f(t\underline{y} + (1-t)\underline{x}_*) - f(\underline{x}_*)}{t}$$

$$\leq \lim_{t \downarrow 0} \frac{t f(\underline{y}) + \cancel{(1-t)} f(\underline{x}_*) \cancel{- f(\underline{x}_*)}}{t}$$

$$\Rightarrow 0 \le \lim_{t \downarrow 0} \frac{\cancel{t} f(\underline{y}) - \cancel{t} f(\underline{x}_*)}{\cancel{t}}$$

$$\Rightarrow 0 \le f(\underline{y}) - f(\underline{x}_*)$$

$$\Rightarrow f(\underline{x}_*) \le f(\underline{y}) \quad \forall\, \underline{y} \in S.$$

Hence, $\underline{x}_*$ is a global min. $\blacksquare$