$$\boxed{\text{minimize } \{f(x) : x \in F\}, \text{ where } f: \mathbb{R}^n \to \mathbb{R}, \quad F \subseteq \mathbb{R}^n}$$

(1) when $F = \mathbb{R}^n$, we have an unconstrained optimization problem.

(2) $F = \{x \in \mathbb{R}^n : h(x) = 0,\ g(x) \leq 0\}$. where $h: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^n \to \mathbb{R}^k$

e.g.

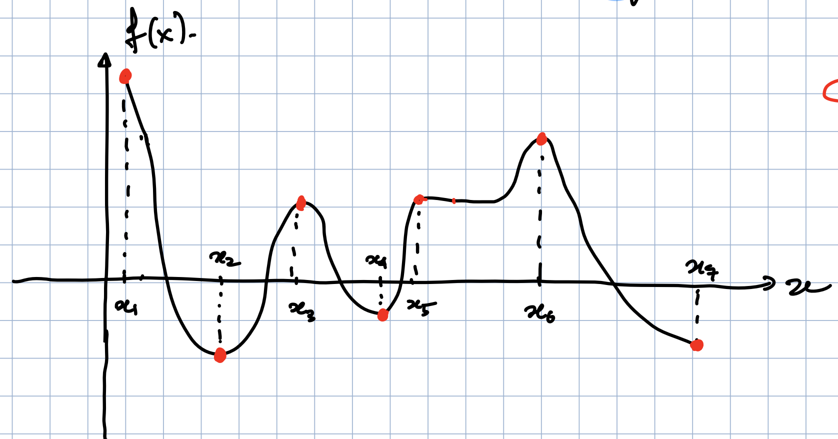
$$\min \quad 3x_1^2 + 5x_2^2$$

$$\text{s.t.} \quad \left.\begin{aligned} 3x_1 + 4x_2 &= 10. \\ x_1^3 + 3x_2 &\geq 1. \end{aligned}\right\} \text{ these constraints define } F.$$

→ what does it mean to minimize?

$\boxed{\text{Definition}}$ $x^* \in F$ is called **local minimum** of f on F if there exists $\epsilon > 0$ such that $f(x) \geq f(x^*)$. for each $x \in F$ such that $\|x - x^*\| \leq \epsilon$. if the inequality $f(x) > f(x^*)$ is strict for all x. $\|x - x^*\| \leq \epsilon$, $x \neq x^*$. then we call x **strict local minimum**.

$\boxed{\text{Definition}}$ $x^* \in F$ is a **global minimum** of f on F if $f(x) \geq f(x^*)$ $\underline{\forall x \in F}$. (not just for $\|x - x^*\| \leq \epsilon$). if $f(x) > f(x^*)$ $\forall x \in F$ $x \neq x^*$, then x^* is called strict **global minimum**.

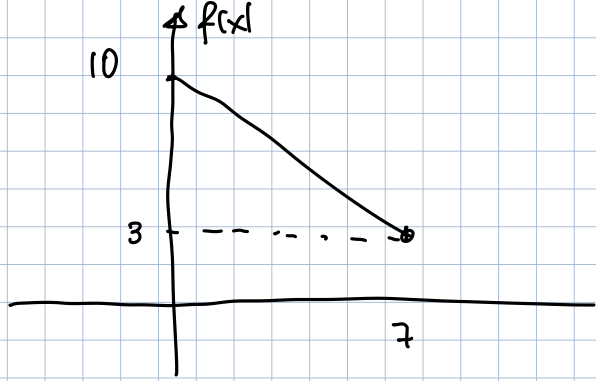


CHARACTERIZE the points

Remark $\quad \min f(x) = -\max(-f(x))$

The existence of a global minimum is not guaranteed.

$$\min \quad 10 - x \quad \text{where} \quad \underbrace{0 \leq x < 7}_{F}. \quad (7 \notin F)$$



Let's prove it by contradiction.
Let $x^* = 7 - \delta$ for some $\delta > 0$ be the global min.
Consider $x' = 7 - \frac{\delta}{2}$ $\left(\frac{\delta}{2} \text{ still} > 0\right)$.

$x' \in F$ and $10 - (7 - \delta/2) = 3 + \delta/2 < 3 + \delta = 10 - x^*$.

$\boxtimes$

Another example:

$$\min \overbrace{(10 - x)}^{f(x)}, \quad \text{where } x \geq 0.$$

The min doesn't exist, $f(x) \to -\infty$ as $x \to +\infty$

Weierstrass theorem if $f(x)$ is continuous on a non-empty feasible set S that is closed and bounded, then $f(x)$ has a global minimum in S.

this is an existential theorem. and gives no algorithmic way to find it!

SADDLE POINTS

0 is a point of inflection where f changes from concave to convex.

During the previous lecture, we saw that for $f(x) = x^3$ $(x \in \mathbb{R})$ $f'(0) = 0$ but $x = 0$ is neither max nor min.

The point $(x^*, y^*) \in \mathbb{R}^{n+m}$ is a saddle point of $f(x,y)$ where $f: \mathbb{R}^{n+m} \to \mathbb{R}$ if.

considering $f(x, y^*)$ as a function of x (fixed $y = y^*$) we get a local min at x^*.

Analog remark: local max at y^* for $f(x^*, y)$.

$$f(x^*, y) \leq f(x^*, y^*) \leq f(x, y^*).$$

$$\forall x: \|x - x^*\| \leq \varepsilon, \quad \forall \|y - y^*\| \leq \varepsilon$$

Necessary conditions for local minimum (unconstrained case $F = \mathbb{R}^n$)

1st order condition: Assume f is continuous, differentiable. x^* is a local minimum of f. Then $\boxed{\nabla f(x^*) = 0}$

Let's do the proof to see one of the many ways Taylor polynomials are useful.

(→ informal)

Proof for the sake of contradiction $\nabla f(x^*) \neq 0$.

Dfn $x(a) = x^* - a \nabla f(x^*)$ $(a \geq 0)$.

→ $1 \times n$ in our convention.

By Taylor: $f(x(a)) = f(x^* - a\nabla f(x^*)) = f(x^*) + \nabla f(x^*)\,(x(a) - x^*) + \text{error } o(a)$

$= f(x^*) - a\,(\nabla f(x^*))\,(\nabla f(x^*))^T$

$(z_1, \ldots, z_n)$

Thus. $f(x(a)) - f(x^*) = -a \sum z_i^2 + o(a)$ error.

For small a, the first term dominates and thus we conclude.

$f(x(a)) < f(x^*)$.

Exercise. Repeat the above proof to show that $\nabla f(x^*)$ for x^* local maximum. Instead of moving opposite from the gradient, move along $\nabla f(x^*)$.

Definition $\{x^* : \nabla f(x^*) = 0\}$ is the set of stationary points of x.

2nd order condition. Assume f is continuous, twice diff.

$\nabla f(x^*) = 0$, $z^T \frac{\partial^2 f}{\partial x^2} z = z^T H_f(x^*) z \geq 0$. $\forall z \in \mathbb{R}^n$.

Equivalently the Hessian H_f evaluated at x^* is positive semidef. (All eigenvalues are ≥ 0).

The proof also based on Taylor approximation, but as expected we will involve 2^{nd} order approximation !.

$\boxed{\text{Proof}}$ As we proved before $\nabla f(x^*) = 0$ is the 1^{st} order necessary cond.

Let's assume that $\exists y \in \mathbb{R}^n$ such that $y^T \frac{\partial^2 f(x^*)}{\partial x^2} y < 0$.

We will use y as the direction to move along.

Dfn $x(a) = x^* + a y \quad , a \geq 0$.

By Taylor's theorem:

for $\hat{a}$ sufficiently small.

$$f(x(a)) - f(x^*) = a \left(\cancel{\nabla f(x^*)} \right)^{\to 0} y + \frac{1}{2} a^2 y^T \frac{\partial^2 f(x^*)}{\partial x^2} y + o(a^2) < 0$$

Thus, $f(x(a)) - f(x^*) < 0 \quad \forall a \in (0, \hat{a})$.

Contradiction ! ☒

Example. $\min\limits_{x_1, x_2} f(x_1, x_2) = 3(x_1 - x_2)^2 + (x_1 + x_2)^3$.

$$\nabla f = [\, 6(x_1 - x_2) + 3(x_1 + x_2)^2 \,, \; -6(x_1 - x_2) + 3(x_1 + x_2)^2 \,] = 0.$$

$$\Rightarrow \quad x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$H_f = \frac{\partial^2 f}{\partial x^2} = \begin{bmatrix} 6 + 6(x_1 + x_2) & -6 + 6(x_1 + x_2) \\ -6 + 6(x_1 + x_2) & 6 + 6(x_1 + x_2) \end{bmatrix}$$

When we evaluate $H_f(0,0) = \begin{bmatrix} +6 & -6 \\ -6 & +6 \end{bmatrix}$ which is positive semidef.

(verify in 2 ways!)

THIS DOES NOT MEAN IT'S LOCAL MIN

Consider the eigenvector $\begin{bmatrix}1\\1\end{bmatrix} = z$ of $H_f(x^*)$ with eigenvalue 0.

$f(x^* + az) = 8a^3$ (so for $a < 0$ we get a value $< f(x^*)$).

This is why we call these conditions necessary!

↳ (but not sufficient).