Lecture 2. Analysis in vector spaces (I)

Consider the minimum of $f(x)$ as examples.

Global minimum and local minimum

$\min_{x \in \Omega} f(x)$. $\quad x^* = \arg\min_{x \in \Omega} f(x)$. usually $\Omega = \mathbb{R}^n$.

x^* is a global minimum if $f(x^*) \le f(x) \quad \forall x \in \Omega$

x^* is a local minimum if $\exists \varepsilon > 0$ such that

$$f(x^*) \le f(x) \quad \forall x \in \Omega \cap B(x^*, \varepsilon).$$

where $B(x^*, \varepsilon) \triangleq \{x : \|x - x^*\| < \varepsilon\}$ is an open ball.

Usually global minimum may not exist.

- $f(x) = x$. $\Omega = \mathbb{R}$. $\quad \inf f(x) = -\infty \quad \min f(x)$ ×
- $f(x) = \frac{1}{x}$. $\Omega = \mathbb{R}_{>0}$. $\quad \inf f(x) = 0$. $\quad \min f(x)$ ×

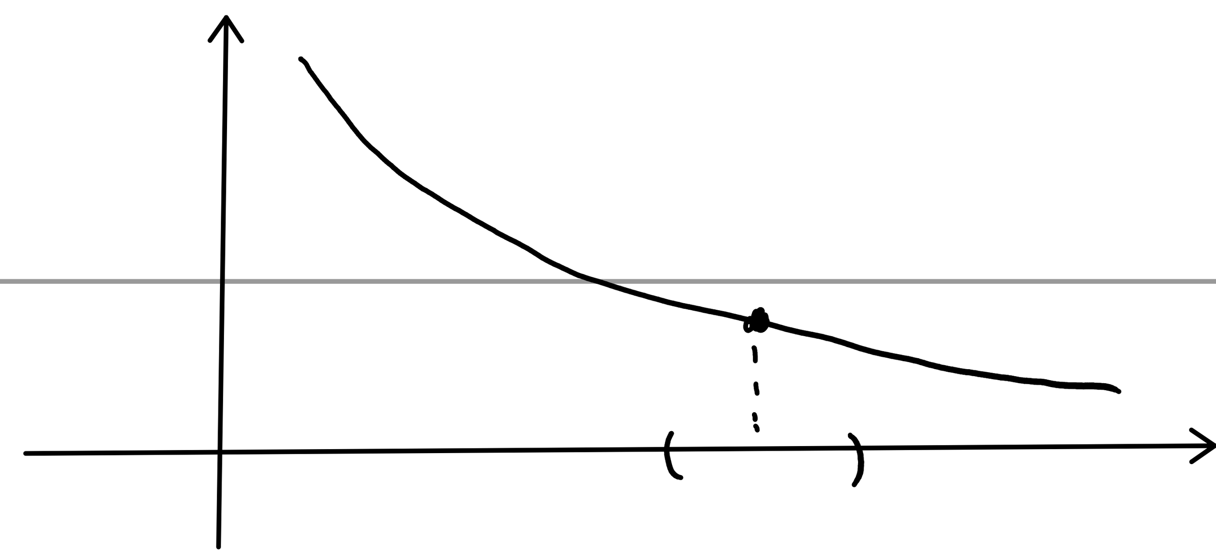
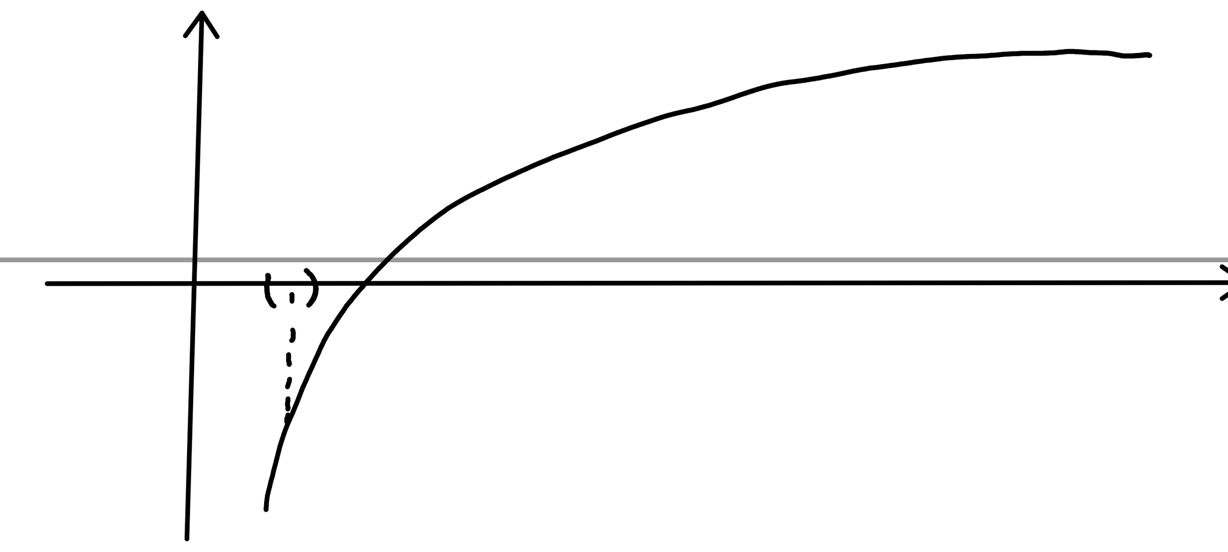
What kind of $f(x)$ (and Ω) has global minimum?

continuous? bounded? not a proper characterization

Definition (continuity). $f: \underset{\uparrow \text{dom}(f)}{D} \to \mathbb{R}$ is continuous at $\alpha \in D$ if

$$\forall \varepsilon > 0.\ \exists \delta.\ \forall x \in D,\ \|x - \alpha\| < \delta \overset{x \in D \cap B(\alpha, \delta)}{\Rightarrow} |f(x) - f(\alpha)| < \varepsilon.$$

(Topological) The inverse image of an open set is open.

Two reasons: unbounded / δ too small. $\Rightarrow$ finite observation ✓

Definition: A set is compact if any open covers has a finite sub.

Theorem (Heine–Borel). For $S \subseteq \mathbb{R}^n$, compact iff bounded and closed.

Definition: $B(x, \varepsilon) : \{x' : \|x - x'\| < \varepsilon\}$. open ball. $(-1, 1)$

what is a closed set? complement is open. $[-1, 1]$

Theorem: S is closed iff $\forall$ sequence $\{x_n\} \subseteq S$. $\lim x_n \in S$.

Definition (convergence). $\{x_n\} \to x$, or $\lim x_n = x$. if

$\lim_{n\to\infty} |x_n - x| = 0$ can be replaced by any norm.

Remark. In $\mathbb{R}^n$ if $x_n \to x$ with some norm, it hold for any norm.

Remark. Above is not true for infinite dimensional space.

$f_k(x) = \begin{cases} k & [0, 1/k^2] \\ 0 & \text{otherwise.} \end{cases}$ $\|f\|_1 \to 0$. $\|f\|_2 \to 1$. $\|f\|_3 \to \infty$

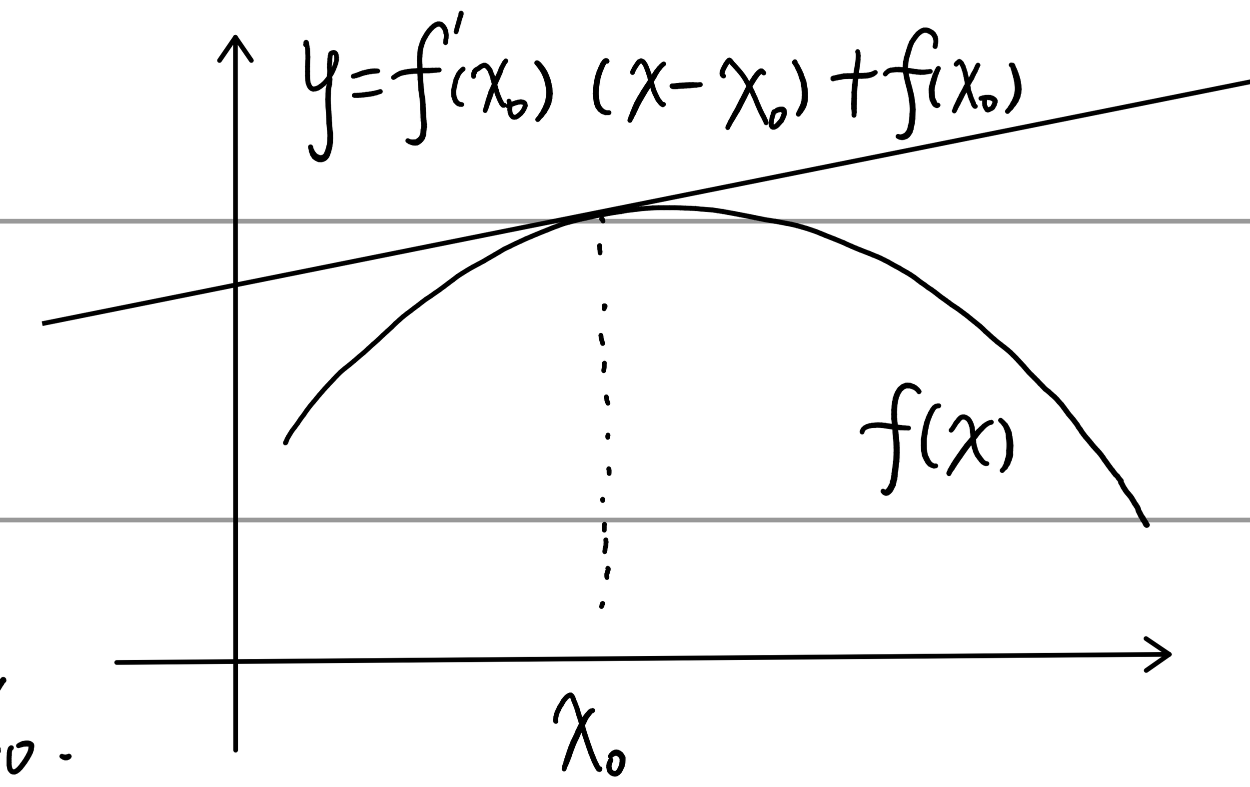
Theorem (Extreme value theorem, Weierstrass).

If S is a compact set, $f: S \to \mathbb{R}$ is continuous. then

f is bounded and has extreme values (both min / max)

If $f: \mathbb{R} \to \mathbb{R}$? If $f(\pm\infty) = \infty$. $\{x : f(x) \leq M\}$ is compact.

Differentiability.



$f(x): \mathbb{R} \to \mathbb{R}$. $f(x_0 + \delta) \approx k\delta + f(x_0)$

Differential is a linear approximation at x_0.

How about general case? in $\mathbb{R}^n$. any linear operator is a matrix.

linear: $f(ru + v) = r f(u) + f(v)$. $y = kx + b$? affine! 仿射.

线性：旋转 + 伸缩. 仿射：线性 + 平移

Definition (Differential). Suppose $f: \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $x_0 \in \Omega$. f is differentiable at x_0 if $\exists$ a matrix $J: \mathbb{R}^n \to \mathbb{R}^m$ ($J \in \mathbb{R}^{m \times n}$) s.t.

$$\lim_{\Omega \ni x \to x_0} \frac{\| f(x) - f(x_0) - J(x - x_0) \|_{\mathbb{R}^m}}{\| x - x_0 \|_{\mathbb{R}^n}} = 0.$$

(J is given by the Jacobian).

J is the differential of f at x_0. denoted by $Df(x_0) = J$.

partial derivative $\frac{\partial}{\partial x_i} f(y) = \lim_{h \to 0} \frac{1}{h} \left(f(y_1, \dots y_{i-1}, y_i + h, y_{i+1}, \dots y_n) - f(y) \right)$

directional derivative $\nabla_v f(x) = \partial_v f(x) = \lim_{h \to 0} \frac{1}{h} \left(f(x + hv) - f(x) \right)$.

Remark: $\exists$ partial derivatives $\not\Rightarrow$ differentiable. $f(x,y) = \begin{cases} x & y \neq x^2 \\ 0 & y = x^2 \end{cases}$

is not differentiable at $(0,0)$. even $\forall$ direction v, $\exists$ directional derivatives

$\not\Rightarrow$ differentiable. see $f(x,y) = \begin{cases} y^3/(x^2 + y^2) & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$

However, if partial derivatives exist in $B(x_0, \varepsilon)$ and continuous at x_0

then $f(x)$ is differentiable at x_0.

On the other hand, if f differentiable at $\vec{x}_0$,

$Df(x_0)$ is given by the Jacobian matrix $J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_1}(\vec{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\vec{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\vec{x}_0) \end{bmatrix}$

In particular, if $f: \mathbb{R}^n \to \mathbb{R}$. $Df \in \mathbb{R}^{1\times n}$

gradient (梯度). $\nabla f(x) = (Df)^T = \left(\frac{\partial f}{\partial x_1}(x), \cdots \frac{\partial f}{\partial x_n}(x)\right)^T$

Examples. $\begin{cases} f(x) = kx+b & f'(x) = k. \\ F(x) = Ax+b & DF(x) = A. \end{cases}$

$= 2(Ax)^T$ if A symmetric.

$\begin{cases} f(x) = ax^2 & f'(x) = 2ax \\ F(x) = x^TAx & DF(x) = ?\ \ x^T(A+A^T) \end{cases}$

$F(x) = \sum_{ij} A_{ij}\, x_i x_j$. $\quad \frac{\partial F}{\partial x_k} = \sum_{ij} A_{ij}\left(x_i \frac{\partial x_j}{\partial x_k} + x_j \frac{\partial x_i}{\partial x_k}\right)$.

multiplication: $f, g: \mathbb{R}^n \to \mathbb{R}^m$. $h \triangleq f^T g$.

$Dh(x) = f(x)^T Dg(x) + g(x)^T Df(x)$.

Chain rule: if $f: X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ differentiable at x_0, $g: Y \subseteq \mathbb{R}^m \to \mathbb{R}^l$

diffentiable at $f(x_0)$ then $h \triangleq g(f(x))$ differentiable at x_0.

$Dh(x_0) = Dg(y_0 = f(x_0))\, Df(x_0)$ (Dg. Df matrix. order !!!)

First-order necessary condition for optimality

Consider $f: \mathbb{R} \to \mathbb{R}$. $x_0 = \operatorname{argmin} f(x)$. $\Rightarrow f'(x_0) = 0$.

$f'(x_0) = \lim\limits_{x \to x_0} \dfrac{f(x) - f(x_0)}{x - x_0}$ 两个方向 $\lim\limits_{x \uparrow x_0}$ $\lim\limits_{x \downarrow x_0}$ both ≥ 0.

Now $f: \mathbb{R}^n \to \mathbb{R}$. consider directional derivative $\nabla_v f = \langle \nabla f, v \rangle$.

why? $g(t) = f(x_0 + tv)$. $g'(t) = Df(x_0 + tv)\, v = \nabla f(x_0 + tv)^T v$.

$f(x_0 + tv) \approx f(x_0) + \nabla_v f(x_0)\, v$. intuitively, hope $\forall v$. $\nabla_v f(x_0) \geq 0$.

Theorem. Suppose $f(x^*)$ is a local minimum and f is differentiable at x^*. Then for any feasible direction v. $\nabla f(x_0)^T v \geq 0$.

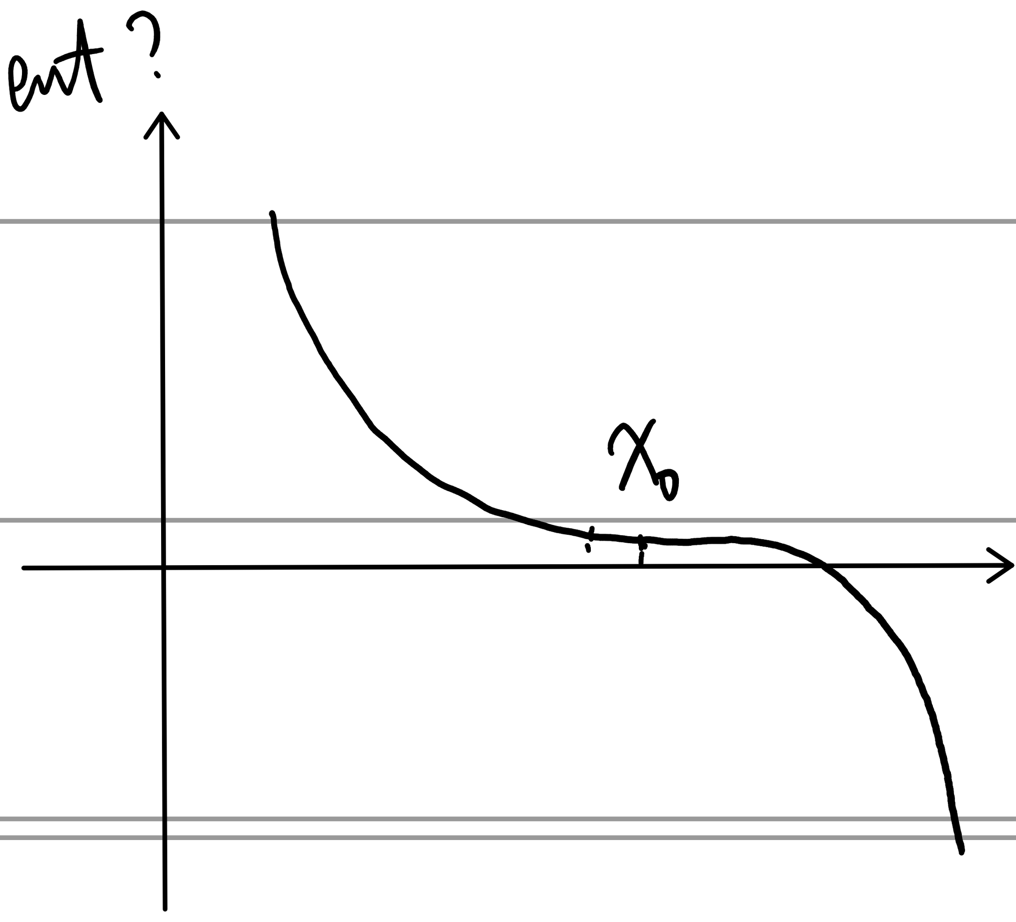
Proof. Fix $v \in \mathbb{R}^n$. $g(t) \triangleq f(x^* + tv)$. where $x^* + tv \in B(x^*, \varepsilon)$.

$\forall t > 0$, $\frac{1}{t}(g(t) - g(0)) \geq 0 \Rightarrow g'(0) = \lim\limits_{t \downarrow 0} \frac{1}{t}(g(t) - g(0)) \geq 0$ □

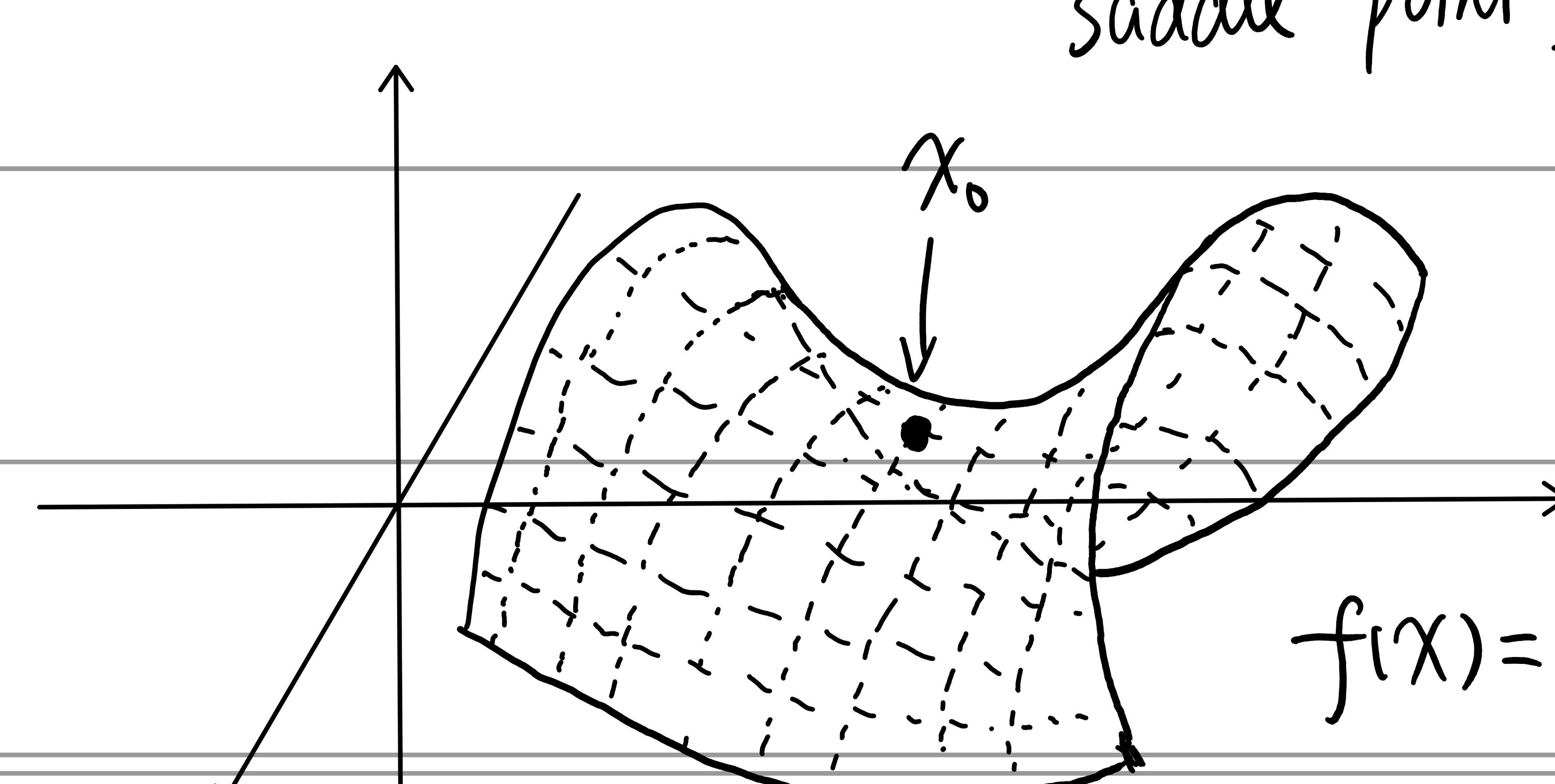
feasible direction v: $\exists t_0 > 0$ s.t. $\forall t \in (0, t_0)$, $x^* + tv \in \operatorname{dom} f$.

Corollary: If x^* interior ($\exists \varepsilon > 0$. s.t. $B(x^*, \varepsilon) \subseteq \operatorname{dom} f$) then $\forall v \in \mathbb{R}^n$. $\nabla f(x^*)^T v = 0$. so $\nabla f(x^*)^T = 0$ by setting $v = \nabla f(x^*)$.

sufficient?



saddle point 鞍点.



$f(x) = x_1^2 - x_2^2$