# Lecture 2. Analysis on normed linear space (I)

S is open if $\forall x \in S$. $\exists \varepsilon > 0$. s.t. $B(x, \varepsilon) \subseteq S$.

definition
线性赋范空间

S is closed if complement open.

some examples.

$(0,1)$ open $[0,1]$ closed $(0,1]$ ?

Converge. $\{x_n\} \to x$. or $\lim_{n\to\infty} x_n = x$. if

$\lim_{n\to\infty} |x - x_n| = 0$ ← can be changed to any norm. proof?

in $\mathbb{R}$. if $x_n \to x$ in one norm. then $x_n \to x$ in any norm

may not true for infinite dimension.

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

$x_n \to x \Rightarrow x \in S$.

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

S is bounded if $\exists M < \infty$. s.t. $\forall x \in S$. $|x| < M$

S is compact if S is closed and bounded.

Another definition: S is compact if any open covers has a finite subcover.

continuous: a function f is continuous at x if

$\forall \varepsilon > 0$. $\exists \delta > 0$. s.t.

$\forall y \in X \cap B(x, \delta)$. $|f(y) - f(x)| < \varepsilon$.

Extreme value theorem (Weierstrass).

If $S$ is compact and $f: S \to \mathbb{R}$ continuous.

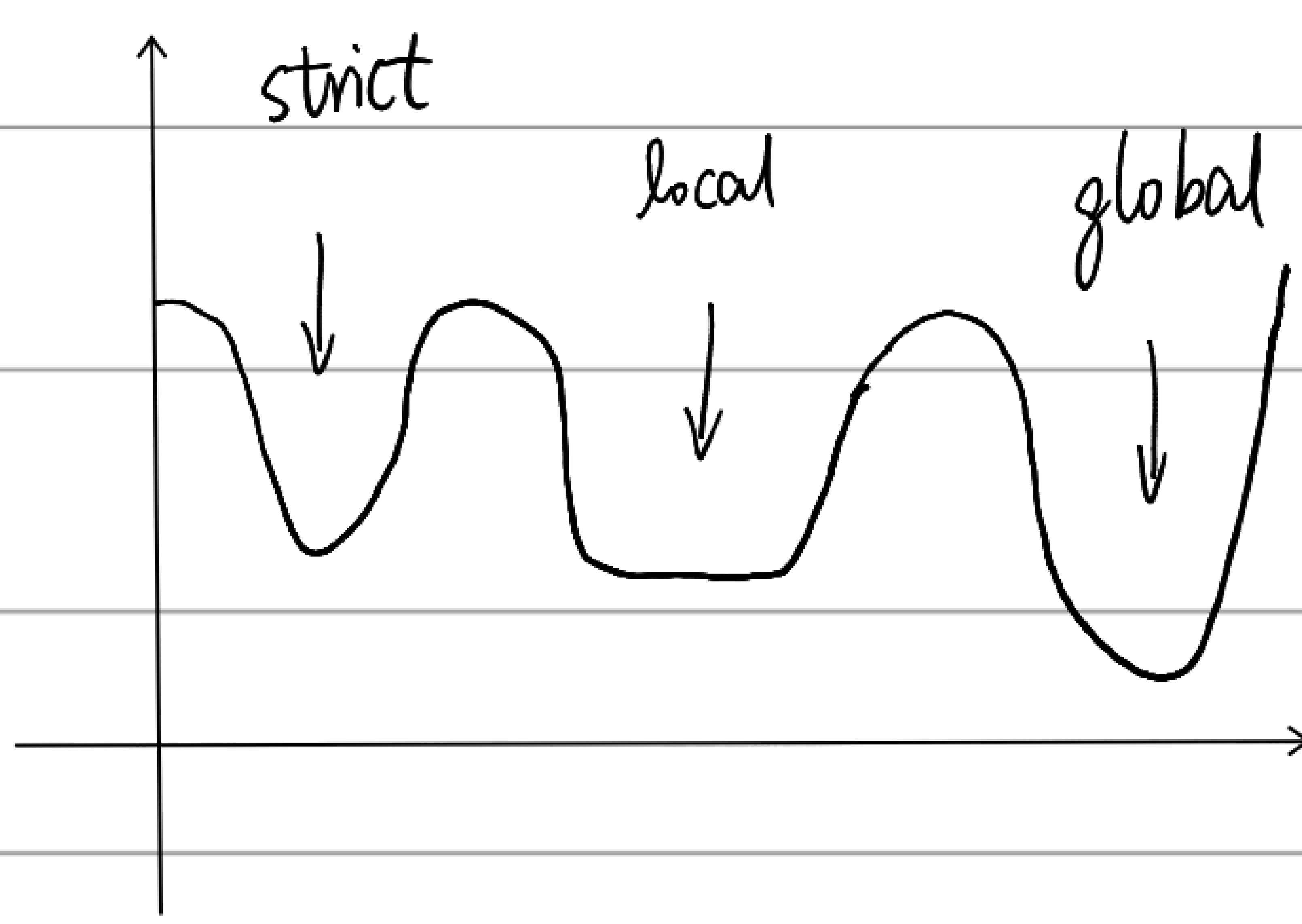
then $f$ is bounded and exists extreme value.

sufficient but not necessary. see $\sin x$ / $\cos x$.

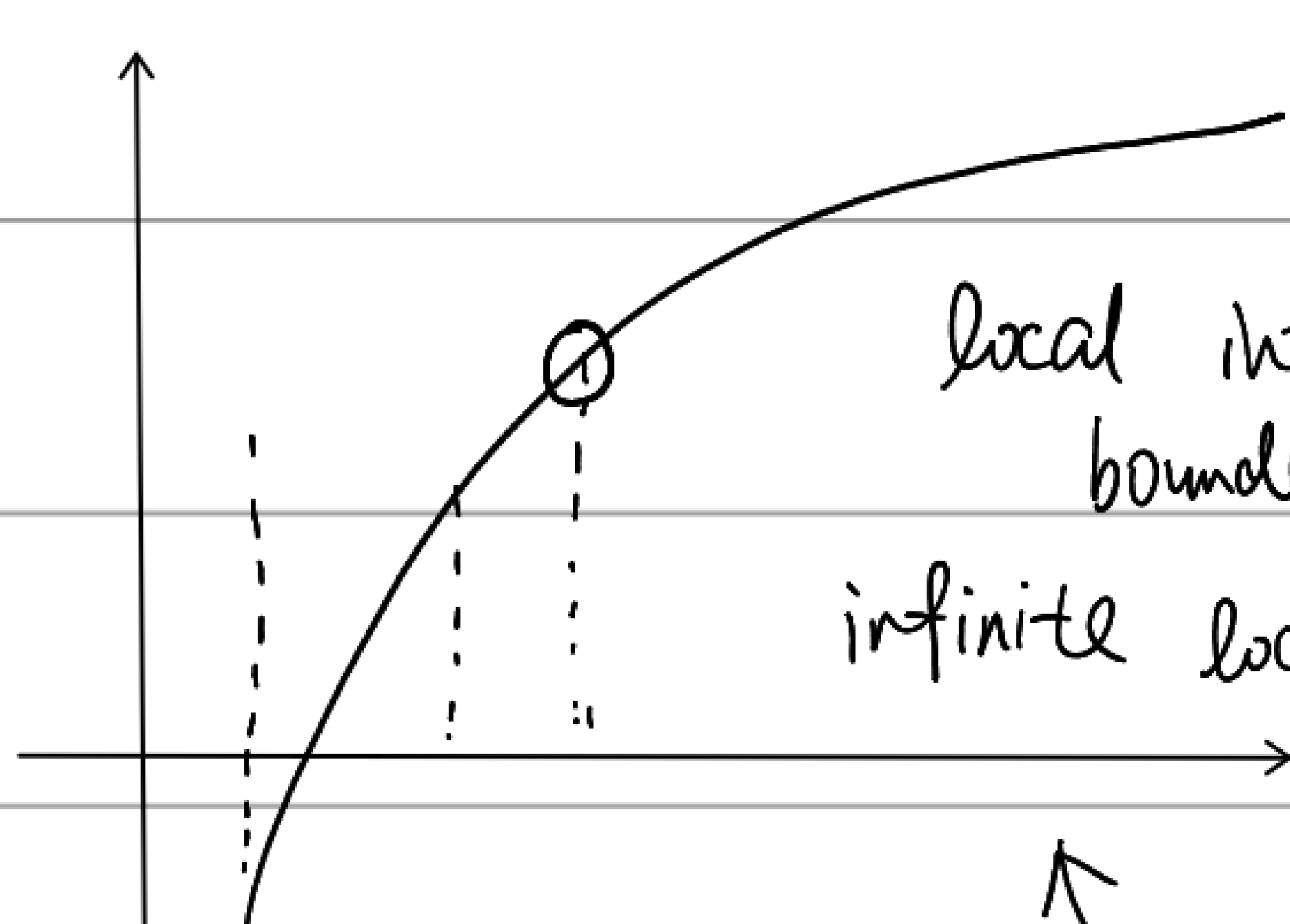
Local minimum

$x^*$ is a local minimum of $f$. if

$\exists\, \varepsilon > 0$. s.t. $\forall\, x \in X \cap B(x^*, \varepsilon)$. $f(x) \geq f(x^*)$.




why does not a function have min?



local information. bounded

infinite local info.

open coves have finite subcovers

$f(x): [0,1] \to \mathbb{R}$.

$f(\frac{p}{q}) = q$.

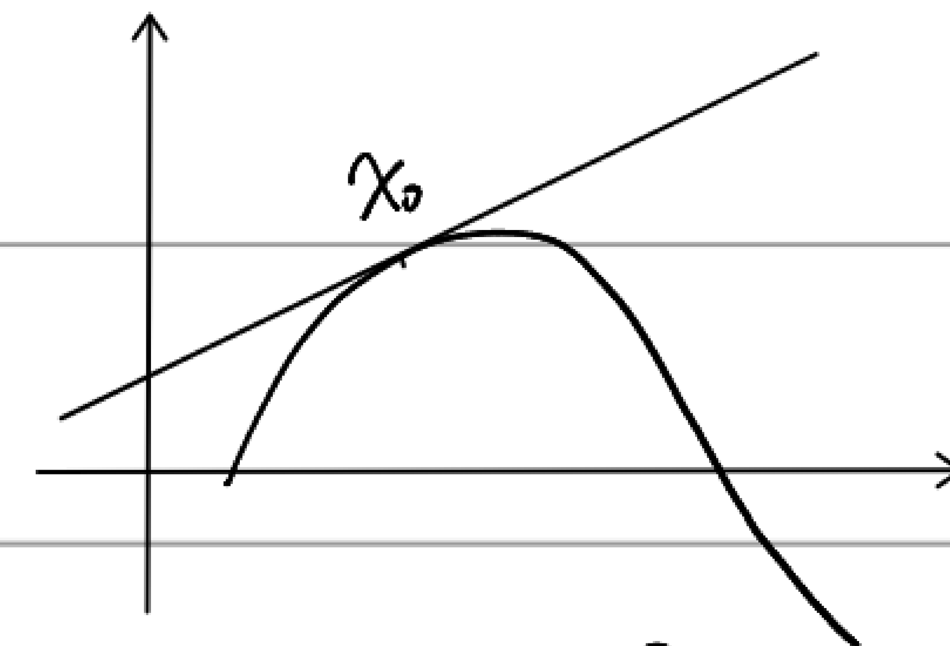
local information unbounded.

continuous

if $f: \mathbb{R} \to \mathbb{R}$ ? $\mathbb{R}$ is not compact.

if $f(\pm\infty) = \infty$. $\{x : f(x) < M\}$ compact.

Differential




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

Differential is a linear approximation of a function.

single value: $y = kx + b$.

general linear space ?

linear $f(a \cdot u + v) = a f(u) + f(v)$.

affine 仿射 线性：旋转 + 扭曲

仿射：线性 + 平移

any linear operator is a matrix $A$.

affine operator: $Ax + b$. $\begin{cases} x \in \mathbb{R}^n \\ A \in \mathbb{R}^{m \times n} \\ b \in \mathbb{R}^m \end{cases}$

$f(x) \approx A(x - x_0) + f(x_0)$

Def. differentiable: $f: \overset{X \subseteq}{\mathbb{R}^n} \to \mathbb{R}^m$: if $\exists$ matrix $\underset{\text{linear operator}}{A} \in \mathbb{R}^{m \times n}$

$$\lim_{\substack{x \to x_0 \\ x \in X}} \frac{\| f(x) - (A(x - x_0) + f(x_0)) \|}{\| x - x_0 \|} = 0.$$ $A$ is differential of $f$.

denoted by $Df(x_0) = A$ or $f'(x_0) = A$.

Given $f(x) = \left(f_1(x), f_2(x), \cdots, f_m(x)\right)^T$

$Df(x_0)_{ij} = \frac{\partial f_i}{\partial x_j}(x_0)$. remark: $x_0 \in \mathbb{R}^n$ is a vector.

called

Jacobian matrix

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{bmatrix}$$

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

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

Example.

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

$=2A$ if $A$ symmetric.

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

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

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

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

$= \sum_i A_{ik} x_i + \sum_j A_{kj} x_j$.

Chain rule.

If $f: X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ differentiable at $x_0 \in X$. $g: Y \subseteq \mathbb{R}^m \to \mathbb{R}^p$ differentiable at $y_0 = f(x_0)$

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

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

First-order necessary condition.

$f: \mathbb{R} \to \mathbb{R}$. 求最值. $f'(x_0) = 0$.

$\lim_{x \to x_0} \frac{|f(x) - f(x_0)|}{|x - x_0|} = 0$ 两个方向 $\lim_{x \uparrow x_0}$ $\lim_{x \downarrow x_0}$ $\geq 0$.

方向导数. $x_0 + tv$.

$g(t) = f(x_0 + tv)$ $g'(t) = Df(x_0+tv)^T v$

$= \nabla f(x_0 + tv)^T v = v^T \nabla f(x_0+tv)$.

in particular $t=0$. $\langle \nabla f(x_0), v \rangle$ $= \langle \nabla f(x_0+tv), v \rangle$.

Intuitively we hope $\forall v$. $\langle \nabla f(x_0), v \rangle \geq 0$.

Thm. 若 $f(x^*)$ local minimum $f$ is differentiable at $x^*$.

then for any direction $v$. $v^T \nabla f(x_0) \geq 0$.

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

$\forall t > 0$. $\frac{g(t) - g(0)}{t} \geq 0 \Rightarrow g'(0) = \lim_{t \downarrow 0} \frac{g(t) - g(0)}{t} \geq 0$.

Cor. $\forall v \in \mathbb{R}^n$. $v^T \nabla f(x_0) = 0$. setting $v = \nabla f(x_0)$

Proof: $v' = -v$. $v^T \nabla f(x_0) \geq 0$. $v'^T \nabla f(x_0) \geq 0$.

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