Lecture 3. Analysis on normed linear space (II

saddle point. 鞍点.

interior point. 内点. boundary point 边界

$\exists\, \varepsilon > 0.\quad B(x, \varepsilon) \triangleq \{ y : \|x - y\| < \varepsilon \} \subseteq X.$

First-order necessary condition. 有偏导不够. $f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} \\ 0. \end{cases}$

if $f(x^*)$ local minimum. f differentiable at x^*.

for any feasible direction v. $(\exists\, \varepsilon > 0 \mid \forall\, 0 < \delta < \varepsilon.\ x^* + \delta v \in X)$.

$v^T \nabla f(x^*) \geqslant 0.$

有方向导数不够

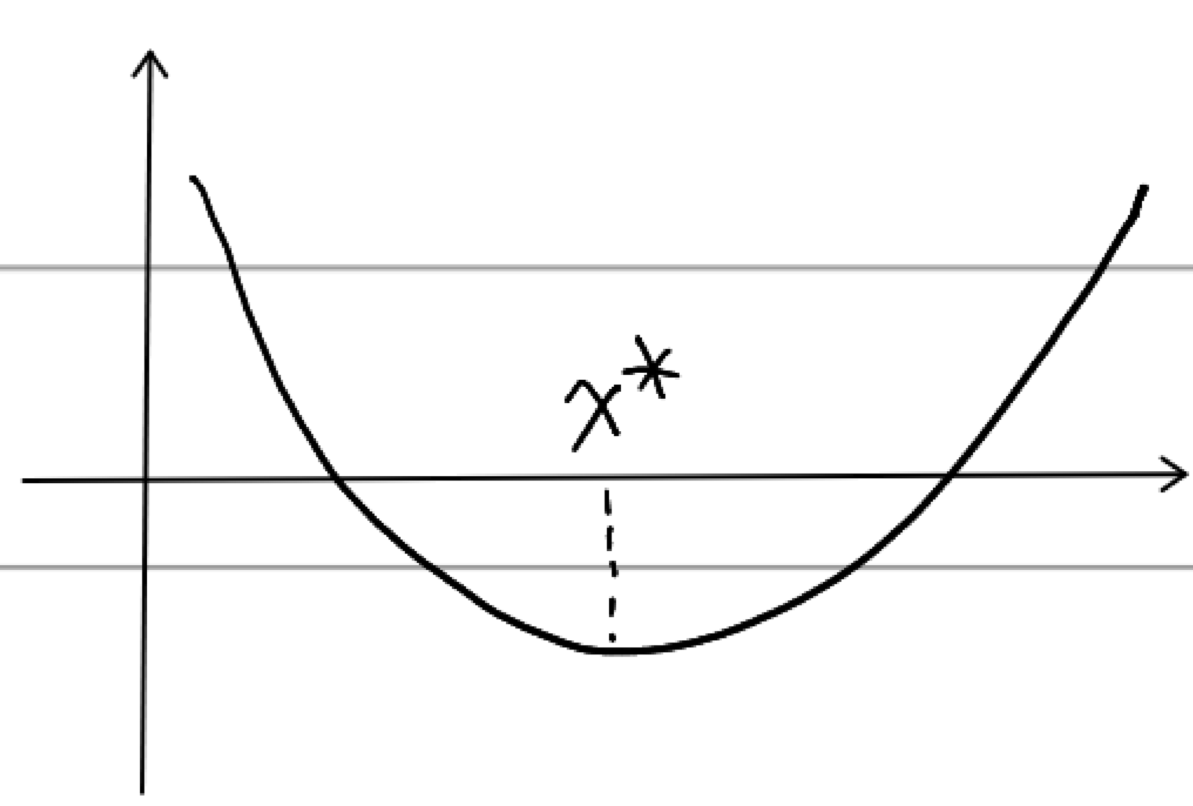
$f(x,y) = \begin{cases} \frac{y^2}{x} & x \neq 0 \\ 0 \end{cases}$

Cor. if x^* interior. $\nabla f(x^*) = 0$

偏导连续.

However. this condition does not suffice.

$f(x) = x^3.\qquad f(x_1, x_2) = x_1^2 - x_2^2.\quad f(x_1, x_2) = x_1^3 - x_2^3.$



$x \uparrow x^*$: $f(x) \downarrow$ $f'(x) < 0$.

$x \downarrow x^*$: $f(x) \downarrow$ $f'(x) > 0$.

$f''(x) > 0$?

特化 $f: \mathbb{R}^n \to \mathbb{R}$. that's why we introduce Jacobian.

Second order partial derivative. $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$

Hessian matrix. denoted by $\nabla^2 f(x_0)$.

Hessian is given by $[\nabla^2 f(x_0)]_{ij} = \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j}$.

if $\frac{\partial f}{\partial x_i \partial x_j}$ and $\frac{\partial f}{\partial x_j \partial x_i}$ exists in $B(x, \varepsilon)$.

and continuous at x_0. then $\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i}$.

Cor. Hessian is a symmetric matrix.

Taylor expansion (second order).

$$f(x_0+\delta) = f(x_0) + \delta f'(x_0) + \frac{\delta^2}{2} f''(x_0) + o(\delta^2).$$

$$f(\vec{x}_0+\vec{\delta}) = f(\vec{x}_0) + \nabla f(\vec{x}_0)^T \vec{\delta} + o(\|\delta\|).$$

$$f(\vec{x}_0+\vec{\delta}) = f(\vec{x}_0) + \nabla f(\vec{x}_0)^T \vec{\delta} + \frac{1}{2} \vec{\delta}^T \nabla^2 f(\vec{x}_0) \vec{\delta} + o(\|\delta\|^2)$$

or expansion.

$$f(\vec{x}_0+\vec{\delta}) = f(\vec{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x}_0)\, \delta_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0)\, \delta_i \delta_j + o(\ ^2)$$

Example. $f(x) = w^T x + b$. $\nabla f(x) = w$ $\nabla^2 f(x) = 0$.

$f(x) = x^T A x$ $\nabla f(x) = (A + A^T) x$ $\nabla^2 f(x) = A + A^T$

in particular. if A symmetric. $\nabla f(x) = 2Ax$. $\nabla^2 f(x) = 2A$.

verify it: $f(x_0+\delta) = (x_0+\delta)^T A (x_0+\delta) = x_0^T A x_0 + \delta^T A x_0 + x_0^T A \delta + \delta^T A \delta$

$= f(x_0) + (x_0^T A + x_0^T A^T)\delta + \delta^T A \delta$

Taylor: $f(x_0+\delta) = f(x_0) + (Ax_0 + A^T x_0)^T \delta + \frac{1}{2}\delta^T (A + A^T)\delta + o(\|\delta\|^2)$.

Chain rule for Hessian.

$h(x) = g(f(x))$ $\quad Dh(x) = Dg(f(x)) \cdot Df(x)$

if $h: \mathbb{R}^n \to \mathbb{R}^m$. $\quad Dh: \mathbb{R}^n \to \mathbb{R}^{m\times n}$ $\quad$ what is D^2h?

$h(x): \mathbb{R}^n \to \mathbb{R}$. $\quad g(x): \mathbb{R}^n \to \mathbb{R}$. $\quad f(x) = Ax+b$

$\nabla h(x) = \nabla g(f(x)) \cdot A$

$=0$.

$\nabla^2 h(x) = \nabla g(f(x)) \cdot DA + A^T D(\nabla g(f(x)))$

$= A^T \nabla^2 g(f(x)) A$.

In particular. $g(t) = f(x_0+tv)$ $\quad \nabla^2 g(t) = v^T \nabla^2 f(x_0+tv) v^T$.

if $f(x_0)$ is a local minimum. we know $\nabla f(x_0)=0$

$f(x_0+\delta) \geq f(x_0)$ $\quad \Longrightarrow \delta^T \nabla^2 f(x_0+\delta)\delta \geq 0$.

忽略小o项

Definite matrix.

Second-order necessary condition

Semidefinite. positive semidefinite if

$A^T = A$ (symmetric) $\quad \forall x.$ $\quad x^TAx \geq 0$

positive definite if symmetric and $\forall x \neq 0$ $\quad x^TAx > 0$

negative if ≤ 0 (semidefinite) $\quad < 0$ (definite)

indefinite if $x_1^TAx_1 < 0 < x_2^TAx_2$

Remark. x^TAx quadratic form 二次型

$$x^TAx = x^TA^Tx = x^T\left(\frac{1}{2}(A+A^T)\right)x$$

Properties of definiteness.

- $A \geqslant 0$ (semidefinite) iff all eigenvalues $\lambda \geqslant 0$.
- $A > 0$ (definite) iff all eigenvalues $\lambda > 0$

$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ $x = (a, b)^T$. $a^2+b^2+(a-b)^2$

$$x^TAx = (2a-b, -a+2b)\begin{pmatrix} a \\ b \end{pmatrix} = 2a^2 - ab - ab + 2b^2 \geqslant 0.$$

$$\det(\lambda I - A) = (\lambda-2)^2 - 1 = 0 \Rightarrow \lambda = 1, 3.$$

why? $A = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i v_i v_i^T$

$U = (v_1, v_2, \cdots v_n)$ $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots \lambda_n)$.

$$x^TAx = (U^Tx)^T\Lambda(U^Tx) = \sum_{i=1}^{n} \lambda_i \|v_i x\|^2.$$

$$x = \sum_{i=1}^{n} y_i v_i \qquad x^TAx = \sum_{i=1}^{n} \lambda_i \|y_i v_i\|^2.$$

Proof of $\delta^T \nabla^2 f(x_0)\delta \geqslant 0$:

otherwise $\exists \lambda < 0$. let v be the eigenvector with respect to λ.

$$f(x_0+tv) = f(x_0) + \nabla f(x_0)^T(tv) + \frac{1}{2}(tv)^T\nabla^2 f(x_0)(tv) + o(\|tv\|^2)$$

$$= f(x_0) + \frac{\lambda}{2}t^2\|v\|^2 + o(t^2\|v\|^2).$$