Lecture 16

§1 Some additional terminologies

1. **Definition:** Stationary points (驻点)

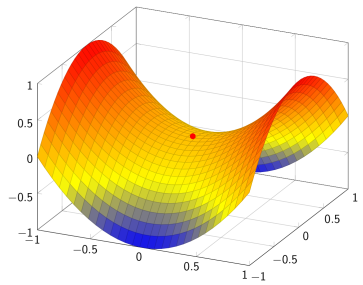
   若 x 满足 $Df(x)=0$，则 x 为 $f(\cdot)$ 的一个 stationary point

   注：FONC characterizes 所有的 stationary points

2. **Definition:** Saddle points (鞍点)

   若 x 为一个 stationary point ($Df(x)=0$)，但它并不是 local maximizer 或 local minimizer，则 x 为一个 saddle point

Consider $f(\mathbf{x}) = x_1^2 - x_2^2$

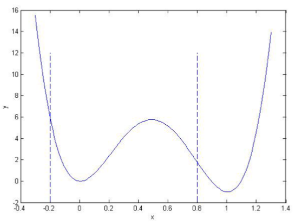


The gradient is $\nabla f(\mathbf{x}) = (2x_1, -2x_2)^\top$ and $\mathbf{x}^* = (0,0)^\top$ is the single stationary point of f.

§2 Constrained problems 与其 FONC

1. Example: constrained 与 unconstrained 问题的区别

Consider the example $f(x) = 100x^2(1-x)^2 - x$ with constraint $-0.2 \le x \le 0.8$.



In addition to the original local minimizer ($x_1 = 0.013$), there is one more local minimizer on the boundary ($x = 0.8$).

At the boundary ($x^* = 0.8$), the FONC is not satisfied

$$f'(0.8) < 0$$

However, at this point, in order to stay feasible, we can only go leftward. That is, in the Taylor expansion

$$f(x^*+d) = f(x^*) + df'(x^*) + o(d)$$

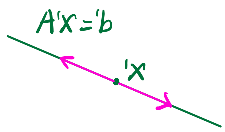
we can only take d to be negative (otherwise it won't be feasible).

Thus $f(x^*+d) > f(x^*)$ in a small neighborhood of x^* in the feasible region. Thus x^* is a local minimizer.

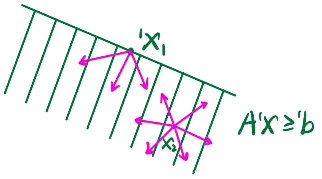
2. **Definition:** feasible direction (可行方向)

   给定 $x \in F$，称 d 为 x 处的 feasible direction 若 $\exists\, \bar{\alpha} > 0$ s.t. $x + \alpha d \in F$ for all $0 \le \alpha \le \bar{\alpha}$

   e.g. 若 $F = \{x \mid Ax = b\}$，则 x 处的 feasible direction 为 $\{d \mid Ad = 0\}$



若 $F=\{\mathbf{x} \mid A\mathbf{x} \geq \mathbf{b}\}$，则 $\mathbf{x}$ 处的 feasible direction 为 $\{\mathbf{d} \mid \mathbf{a}_i^T\mathbf{d} \geq 0 \text{ if } \mathbf{a}_i^T\mathbf{x}=b_i\}$



3. **Theorem:** FDNC for constrained problem

$\mathbf{x}^*$ 为 local minimum $\Rightarrow$ 对 $\forall$ feasible $\mathbf{d}$ at $\mathbf{x}^*$，有 $\nabla f(\mathbf{x}^*)^T\mathbf{d} \geq 0$

注：对于 unconstrained problem，由于所有方向均 feasible，则有 $\nabla f(\mathbf{x}^*)$ 必须为 0

（若 $\nabla f(\mathbf{x}^*)^T\mathbf{d} > 0$，则 $\nabla f(\mathbf{x}^*)^T(-\mathbf{d}) < 0$）

§3 An alternative view for FDNC

1. **Definition:** descent direction（下降方向）

令 f 为 continuously differentiable，则 $\mathbf{d}$ 为 descent direction $\Longleftrightarrow$ $\nabla f(\mathbf{x})^T\mathbf{d} < 0$

注：另一种定义：

$\mathbf{d}$ 为 descent direction $\Longleftrightarrow$ $\exists\, \bar{\gamma} > 0$ s.t. $f(\mathbf{x}+\gamma\mathbf{d}) < f(\mathbf{x})$ for all $0 < \gamma \leq \bar{\gamma}$

2. FDNC 的等价形式

若定义 set of feasible directions at $\mathbf{x}$ 为 $S_F(\mathbf{x})$，

set of descent directions at $\mathbf{x}$ 为 $S_D(\mathbf{x})$.

则 FDNC 等价于

$$S_F(\mathbf{x}^*) \cap S_D(\mathbf{x}^*) = \emptyset$$

即没有 feasible descent directions.

3. Nonlinear optimization with equality constraints

对于 equality constraints：

$$\min_{\mathbf{x}} \quad f(\mathbf{x})$$
$$\text{s.t.} \quad A\mathbf{x} = \mathbf{b}$$

The feasible direction set S_F 为 $\{\mathbf{d} \mid A\mathbf{d} = 0\}$

The descent direction set S_D 为 $\{\mathbf{d} \mid \nabla f(\mathbf{x})^T\mathbf{d} < 0\}$

Theorem: alternative system

$A\mathbf{d} = 0$ & $\nabla f(\mathbf{x})^T\mathbf{d} < 0$ 无解 (FDNC) $\Longleftrightarrow$ $\exists\, \mathbf{y}$ s.t. $A^T\mathbf{y} = \nabla f(\mathbf{x})$

因此，

$\mathbf{x}^*$ 为 local minimum $\Rightarrow$ $\exists\, \mathbf{y}$ s.t. $A^T\mathbf{y} = \nabla f(\mathbf{x}^*)$

证明：

① 先证明 $\exists\, \mathbf{y}$ s.t. $A^T\mathbf{y} = \nabla f(\mathbf{x}) \Rightarrow A\mathbf{d} = 0$ & $\nabla f(\mathbf{x})^T\mathbf{d} < 0$ 无解

$$A^T y = \nabla f(x)$$
$$\Rightarrow y^T A = \nabla f(x)^T$$
$$\Rightarrow y^T A d = \nabla f(x)^T d$$

若 $Ad=0$，则 $\nabla f(x)^T d = 0$ (contradiction)

② 再证明 $Ad=0$ & $\nabla f(x)^T d<0$ 无解 $\Rightarrow \exists y$ s.t. $A^T y = \nabla f(x)$

考虑 LP:

$$\min_d \quad \nabla f(x)^T d$$
$$\text{s.t.} \quad Ad = 0$$

其 dual problem 为

$$\max_y \quad 0$$
$$\text{s.t.} \quad A^T y = \nabla f(x)$$

若不存在 d 满足 $Ad=0$ 且 $\nabla f(x)^T d<0$，则 primal 的 optimal value 一定为 0

由 Strong duality theorem，dual problem 一定 feasible (optimal value 为 0)，

即 $A^T y = \nabla f(x)$

例: Consider the problem:

$$\text{minimize} \quad (x_1-1)^2+(x_2-1)^2$$
$$\text{s.t.} \quad x_1+x_2=1$$

由 FONC，$x=(x_1,x_2)$ 为 local minimizer，若 $\exists y$ 满足

$$A^T y = \nabla f(x)$$

此处 $A=(1,1)$，$\nabla f(x) = (2x_1-2, 2x_2-2)^T$

即
$$\begin{cases} 2x_1-2 = y \\ 2x_2-2=y \end{cases}$$

Also combined with the constraint $x_1+x_2=1$，有

$$x_1=x_2=\frac{1}{2}$$

为 the only candidate for local minimum，且 it is indeed a local minimizer.

例: Consider the problem:

$$\text{minimize}_\beta \quad \|X\beta - y\|_2^2$$
$$\text{s.t.} \quad W\beta = z$$

The gradient 为 $\frac{\partial}{\partial\beta}(X\beta-y)^T(X\beta-y) = 2X^T(X\beta-y)$

因此 FONC 为 $\exists z$, s.t.

$$W^T z = 2X^T(X\beta-y)$$

因此，一个 optimal β 必须满足:

$$\begin{cases} W\beta = z \\ X^T X\beta = \frac{1}{2}W^T z + X^T y \end{cases}$$

即 $\begin{bmatrix} W & 0 \\ X^TX & -\frac{1}{2}W^T \end{bmatrix} \begin{bmatrix} \beta \\ z \end{bmatrix} = \begin{bmatrix} \xi \\ X^Ty \end{bmatrix}$

由于 X 为 $m\times n$，W 为 $d\times n$，因此共 $n+d$ 个方程，$n+d$ 个未知数

4. Nonlinear optimization with inequality constraints

对于 inequality constraints：

$$\min_{x} \ f(x)$$
$$\text{s.t.} \quad Ax \geq b$$

The feasible direction set S_F 为 $\{d \mid a_i^T d \geq 0 \text{ if } a_i^T x = b_i\}$

The descent direction set S_D 为 $\{d \mid \nabla f(x)^T d < 0\}$

Theorem: alternative system

$$x^* \text{ 为 local minimum} \Rightarrow \exists y \geq 0 \text{ s.t. } \begin{cases} \nabla f(x^*) = A^T y \\ y_i \cdot (a_i^T x^* - b_i) = 0, \ \forall i \end{cases}$$

其中 a_i^T 为 A 的第 i 行

证明：

FONC 为不存在 d，使得 ① $\nabla f(x)^T d < 0$ ② $a_i^T d \geq 0$ for $i \in A(x)$

等价于 $\exists y \geq 0$ s.t. $\nabla f(x) = \sum_{i\in A(x)} a_i y_i$

等价于 $\exists y \geq 0$ s.t.

$$\begin{cases} \nabla f(x^*) = A^T y \\ y_i \cdot (a_i^T x^* - b_i) = 0, \ \forall i \end{cases}$$