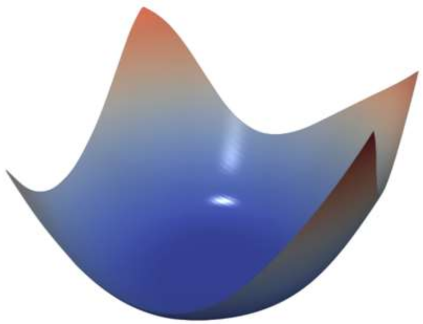
第六章　非线性规划

- 非线性规划问题及其数学模型
- 非线性规划解析解法
- 非线性规划数值解法
- 非线性规划与人工智能



非线性规划的解析解法

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束 （**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

非线性规划的数学模型

$$\min f(\boldsymbol{x})$$

s.t.

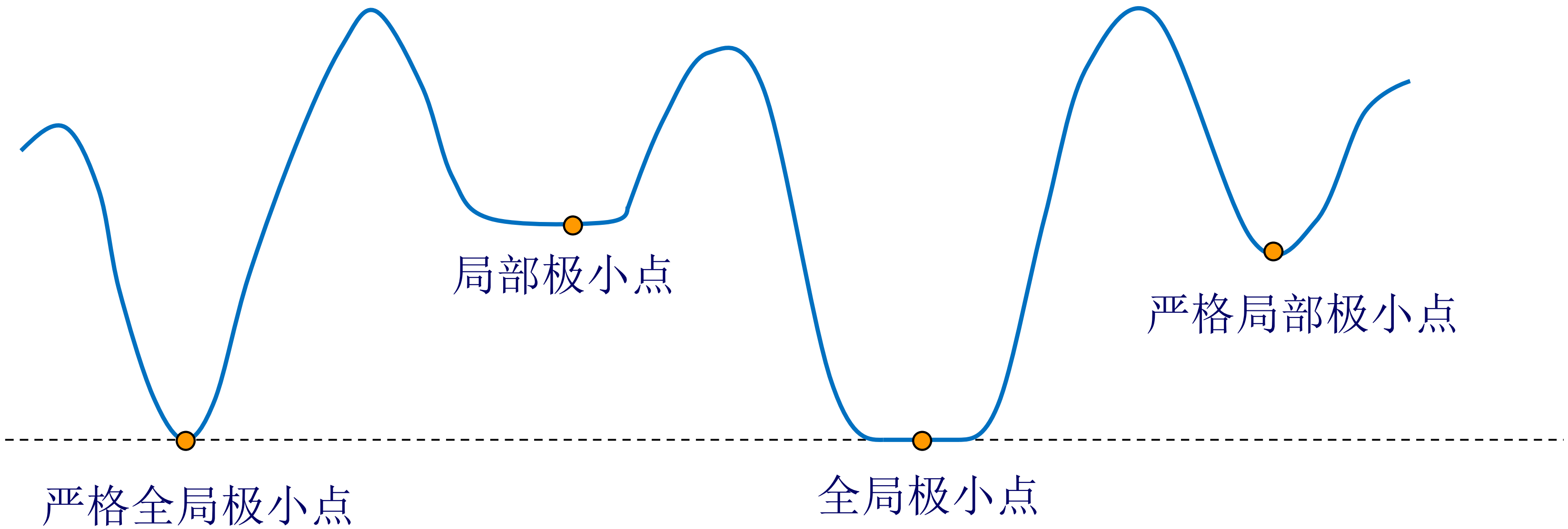
$$h_i(\boldsymbol{x}) = 0 \qquad i = 1, 2, \ldots, m$$

$$g_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, l$$

$$\boldsymbol{x} \in R^n$$

以求最小值为标准问题

非线性规划解的类型



- *局部极小点*：　∃ε>0，满足 $f(x^*)\le f(x)$，　$x\in N(x^*,\varepsilon)\cap$ 可行域
- *全局极小点*（最小点）：$f(x^*)\le f(x)$，$x\in$ 可行域

严格极小点：　“<”

问题：规划的目标是找哪一种极小点？

非线性规划的解析解

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束 （**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

无约束极值问题

$$\min f(\boldsymbol{x}) \qquad \boldsymbol{x} \in R^n$$

- 一阶极值条件（必要条件）
- 二阶极值条件（必要条件、充分条件）

一阶极值条件（必要条件）

驻点条件

$$\nabla f(\boldsymbol{x}^*) = \left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T \Bigg|_{\boldsymbol{x}=\boldsymbol{x}^*} = \mathbf{0}$$

梯度

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$$f(\boldsymbol{x}+\Delta\boldsymbol{x}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \Delta\boldsymbol{x} + \mathrm{O}(\|\Delta\boldsymbol{x}\|)$$

梯度的几何性质

- $\nabla f(x)$ 为目标函数$f(x)$等值面在x的法向量
- $\nabla f(x)$是目标函数值$f(x)$在x点增长最快的方向

证明： $f(\boldsymbol{x}+\lambda\boldsymbol{p}) = f(\boldsymbol{x}) + \lambda\nabla f(\boldsymbol{x})^T \boldsymbol{p} + \mathrm{O}(\lambda)$ $\quad \|\boldsymbol{p}\|_2 = 1$

■ 方向导数（沿p方向的导数）：

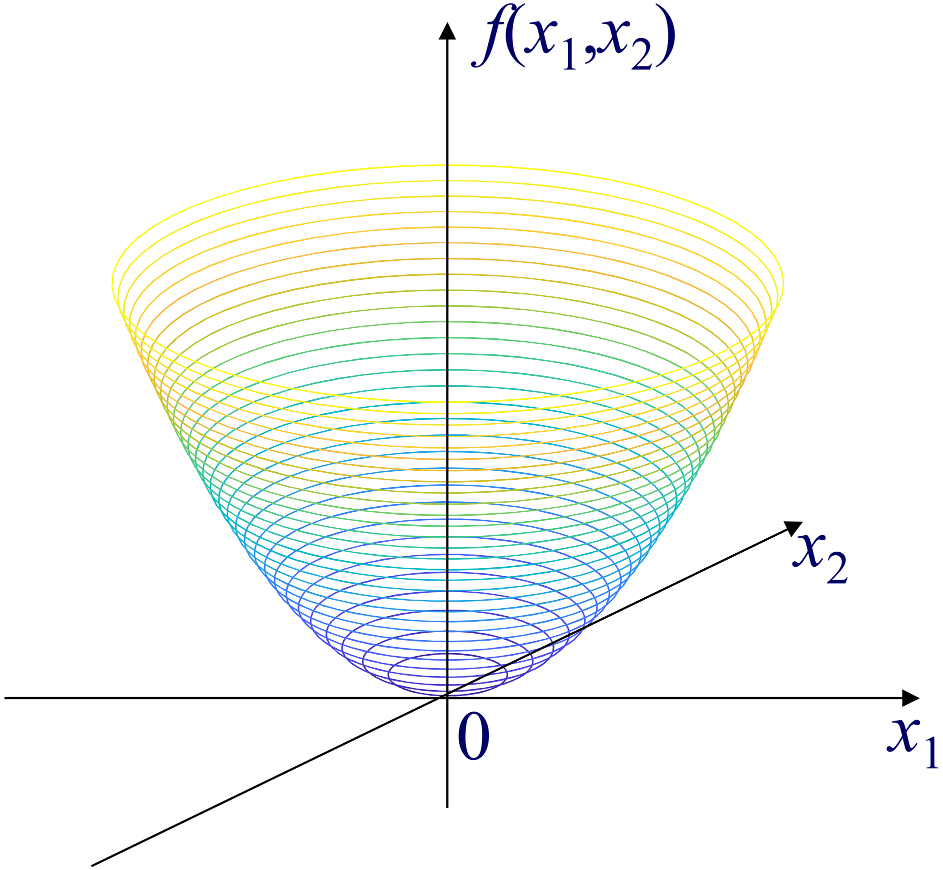
$$D_{\boldsymbol{p}}(\boldsymbol{x}) = \lim_{\lambda\to 0}\frac{f(\boldsymbol{x}+\lambda\boldsymbol{p}) - f(\boldsymbol{x})}{\lambda} = \nabla f(\boldsymbol{x})^T \boldsymbol{p} = \|\nabla f(\boldsymbol{x})\|_2 \cos(\theta)$$

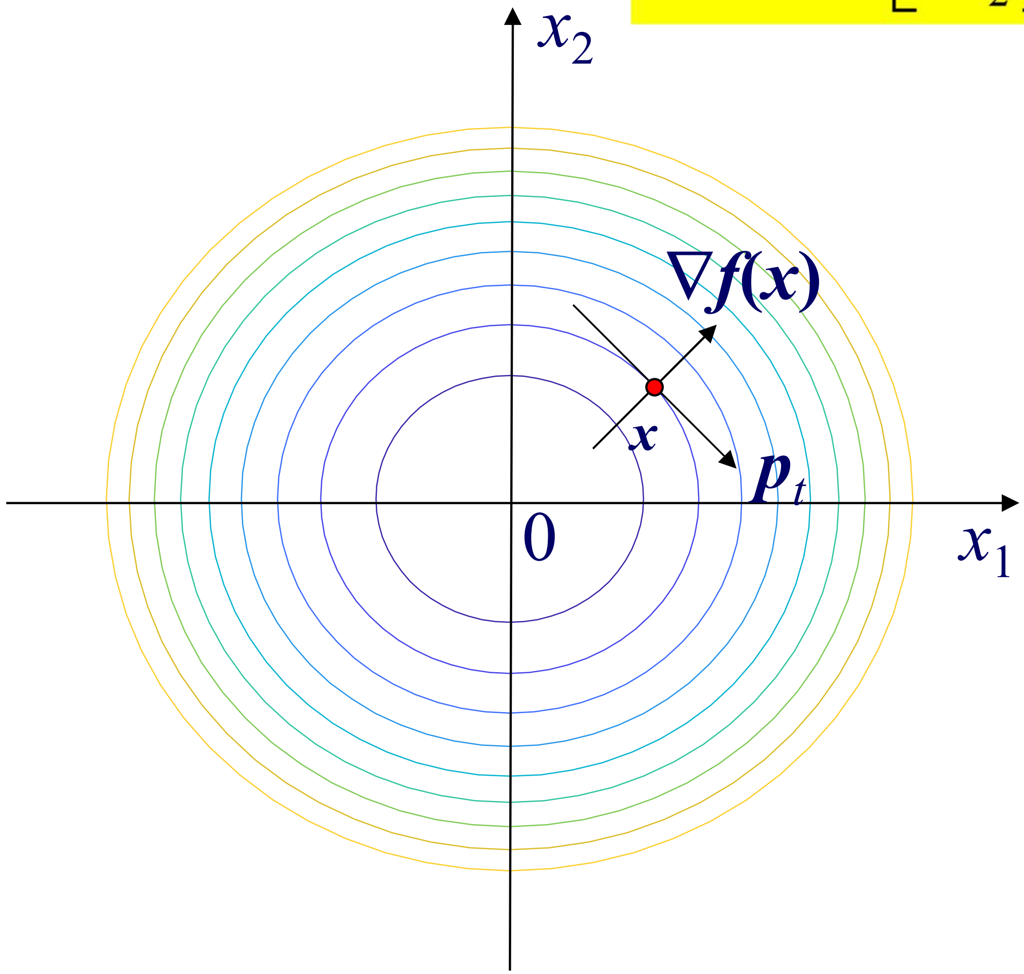
- p取$f(x)$等值面切方向时， $\nabla f(\boldsymbol{x})^T \boldsymbol{p} = 0$
- p取$\nabla f(x)$时，方向导数最大。

梯度图示

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$





Taylor公式的二阶展开

$$f(\boldsymbol{x}+\Delta\boldsymbol{x}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \Delta\boldsymbol{x} + \frac{1}{2}\Delta\boldsymbol{x}^T \nabla^2 f(\boldsymbol{x})\Delta\boldsymbol{x} + \mathrm{O}(\|\Delta\boldsymbol{x}\|^2)$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_2} \cdots\cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2^2} \cdots\cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_2 \partial x_n} \\ \vdots & \vdots \quad \cdots\cdots & \vdots \\ \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_2} \cdots\cdots & \frac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}$$

Hessian Matrix

二阶导数矩阵

矩阵的正定与负定

二次型： $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j$　$\boldsymbol{A}$为对称矩阵

$\forall \boldsymbol{x}\neq 0$　$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}>0 \Longrightarrow \boldsymbol{A}>0$ 正定矩阵

$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\geq 0 \Longrightarrow \boldsymbol{A}\geq 0$ 半正定矩阵

$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}<0 \Longrightarrow \boldsymbol{A}<0$ 负定矩阵

$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\leq 0 \Longrightarrow \boldsymbol{A}\leq 0$ 半负定矩阵

二阶极值必要条件

$$f(\boldsymbol{x}+\Delta\boldsymbol{x})=f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^T\Delta\boldsymbol{x}+\frac{1}{2}\Delta\boldsymbol{x}^T\nabla^2 f(\boldsymbol{x})\Delta\boldsymbol{x}+O(\|\Delta\boldsymbol{x}\|^2)$$

$$\boxed{\nabla f(\boldsymbol{x}^*)=\mathbf{0} \qquad \nabla^2 f(\boldsymbol{x}^*)\geq 0}$$

局部极小值⇒“≥”半正定

严格局部极小值⇒“>”正定

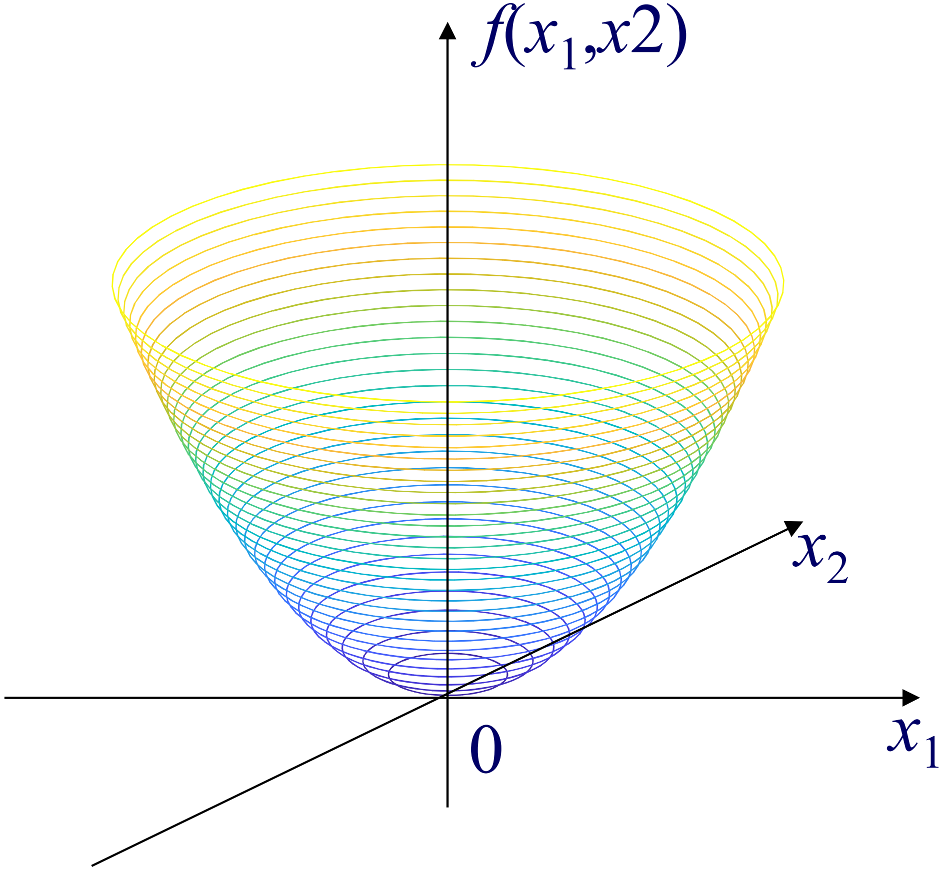
局部极大值⇒“≤”半负定

严格局部极大值⇒“<”负定

仅为必要条件！

例1：Hesse矩阵为正定阵

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$

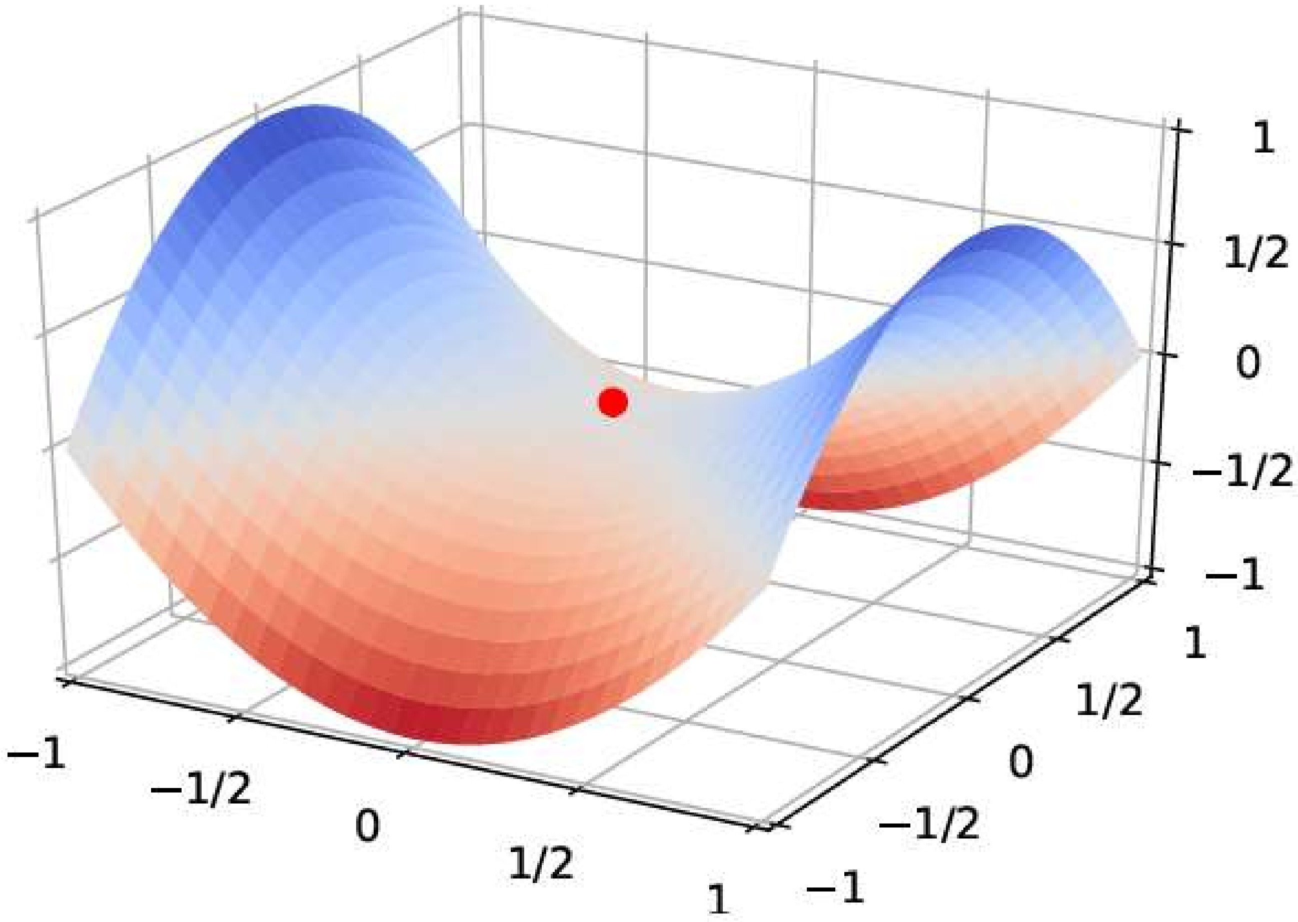


$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

例2：鞍点

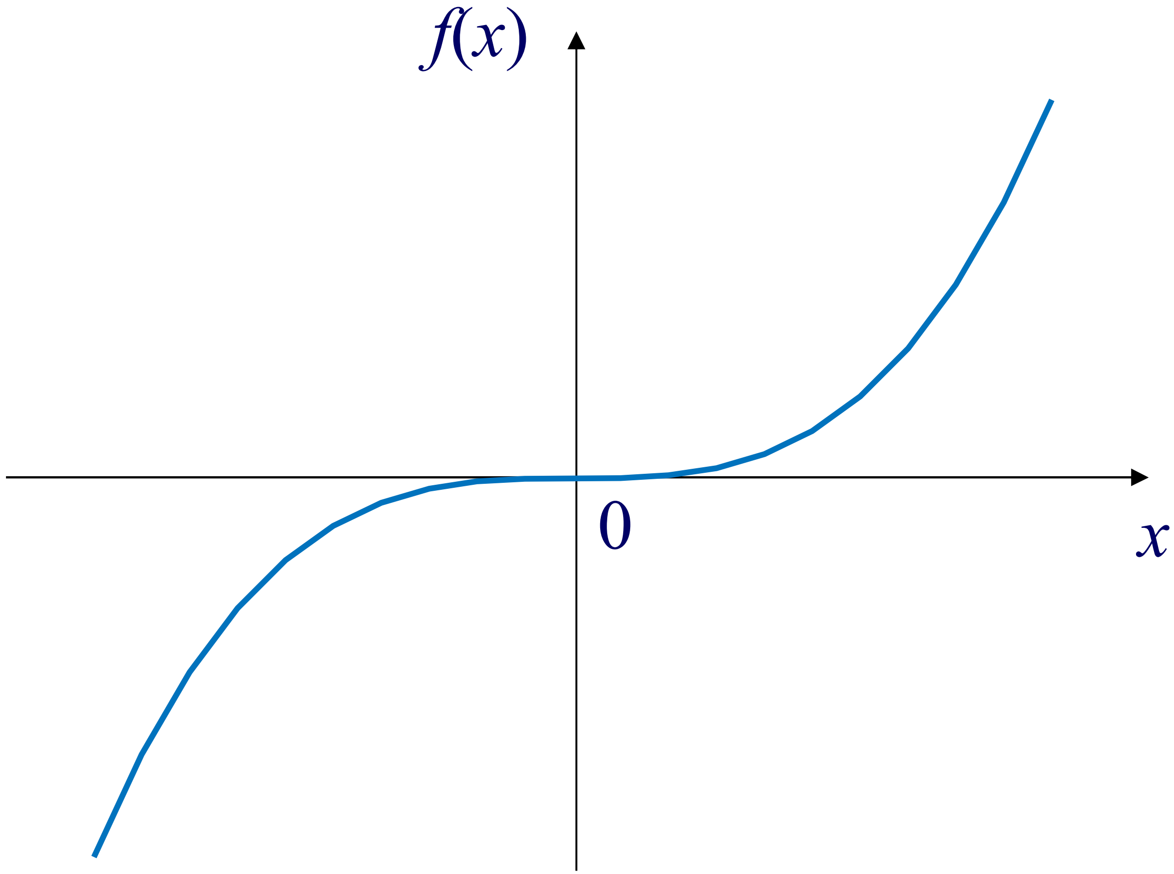
$$f(\boldsymbol{x}) = x_1^2 - x_2^2$$



$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

例3：必要非充分



$$f(x)=x^3$$

$$\nabla f(x)=3x^2$$

$$\nabla^2 f(x)=6x$$

如果进一步满足x^*邻域内的所有点的$\nabla^2 f(x)\geq 0$，则$\nabla f(x^*)=0$且$\nabla^2 f(x^*)\geq 0$是x^*是极小值的充分条件。（应用不方便！）

二阶极值充分条件

$$f(\boldsymbol{x}+\Delta \boldsymbol{x})=f(\boldsymbol{x})+\nabla f(\boldsymbol{x})^{T} \Delta \boldsymbol{x}+\frac{1}{2} \Delta \boldsymbol{x}^{T} \nabla^{2} f(\boldsymbol{x}) \Delta \boldsymbol{x}+\mathrm{O}(\|\Delta \boldsymbol{x}\|^{2})$$

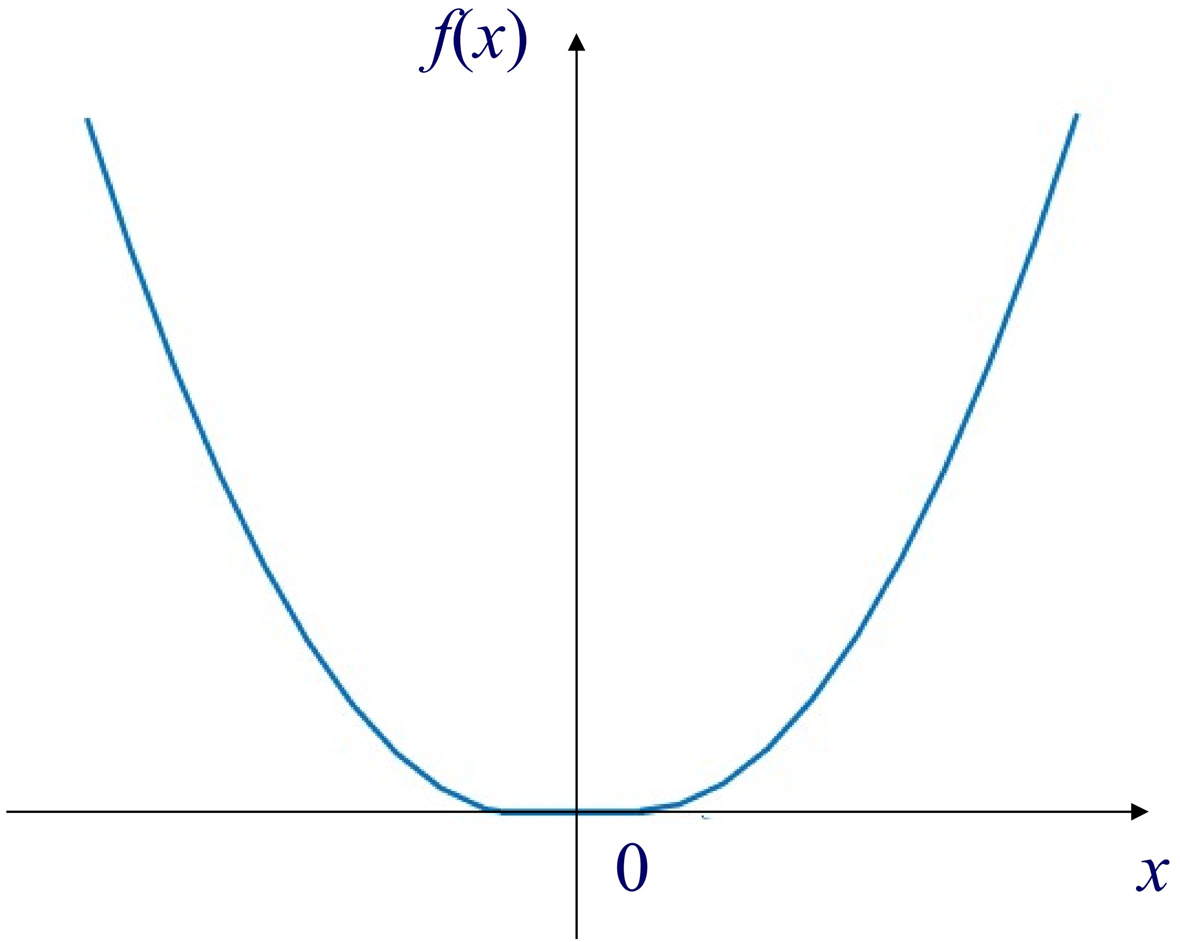
$$\boxed{\nabla f(\boldsymbol{x}^*)=\mathbf{0} \qquad \nabla^{2} f(\boldsymbol{x}^*)>0}$$

“>” 正定⇒严格局部极小值

“<” 负定⇒严格局部极大值

“=” 需检验更高阶导数项

例4 充分非必要



$\nabla f(0) = 0$

$\nabla^2 f(0) = 0$

小结

无约束极小值问题的最优性条件

■一阶必要条件：（局部极小值/局部极大值）

$$\nabla f(x^*)=0$$

■二阶必要条件：（局部极小值）

$$\nabla f(x^*)=0 \text{且} \nabla^2 f(x^*)\geq 0$$

■二阶充分条件：（严格局部极小值）

$$\nabla f(x^*)=0 \text{且} \nabla^2 f(x^*)>0$$

非线性规划的解析解

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束 （**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

等式约束极值问题

$$\min f(\boldsymbol{x})$$

$$\text{s.t.} \quad h_i(\boldsymbol{x}) = 0 \qquad i = 1, 2, \ldots, m$$

$$\boldsymbol{x} \in R^n$$

$$\min f(\boldsymbol{x})$$

$$\text{s.t.} \quad \boldsymbol{h}(\boldsymbol{x}) = 0$$

$$\boldsymbol{x} \in R^n$$

$$\boldsymbol{h}(\boldsymbol{x}) \triangleq \begin{bmatrix} h_1(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix}$$

Lagrange函数法

定义Lagrange函数

$$\min L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i h_i(\boldsymbol{x})$$

$$\boldsymbol{\lambda} = [\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_m]^T$$

求无约束极值问题：

$$\left.\frac{\partial L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \boldsymbol{x}}\right|_{x^*} = \mathbf{0} \quad \Longrightarrow \quad \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(\boldsymbol{x}^*) = \mathbf{0}$$

$$\left.\frac{\partial L(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right|_{\lambda^*} = \mathbf{0} \quad \Longrightarrow \quad h_i(\boldsymbol{x}^*) = 0 \qquad i = 1, 2, \ldots, m$$

非线性规划的解析解

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束 （**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

不等式约束极值问题

$$\min f(\boldsymbol{x})$$

s.t. $\quad g_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, l$

$$\boldsymbol{x} \in R^n$$

思考：能否如线性规划化为等式约束的问题？

求解思路

- 直观思想：邻域内不存在可行的更优解

- 不存在同时满足以下两个条件的方向
  - 可行点的可行方向
  - 目标函数的下降方向

- 最优性条件
  - **Fritz John**条件
  - **Kuhn-Tucker**条件

可行方向

可行点$\boldsymbol{x}^{(0)}$的可行方向$\boldsymbol{p}$是$\boldsymbol{x}^{(0)}$ 沿$\boldsymbol{p}$方向移动无限小步后仍在可行域内的方向（约束函数值增大方向），数学上可表述为：

$\exists\lambda_0 > 0$，对于$\forall\lambda \in (0, \lambda_0]$，有：

$$\boldsymbol{x}^{(0)} + \lambda\boldsymbol{p} \in \Omega, \qquad \Omega = \{\boldsymbol{x} \mid g_j(\boldsymbol{x}) \ge 0, j = 1, 2, \ldots, l\}$$

即 $$g_j(\boldsymbol{x}^{(0)} + \lambda\boldsymbol{p}) \ge 0 \qquad j = 1, 2, \ldots, l$$

可行方向判别条件（充分条件）

$\nabla g_j(\boldsymbol{x}^{(0)})^T \boldsymbol{p} > 0$　　$\forall j \in J(\boldsymbol{x}^{(0)})$　　$J(\boldsymbol{x}^{(0)}) = \{j \mid g_j(\boldsymbol{x}^{(0)})=0, j = 1,2,\ldots,l\}$

起作用约束集合

证明：　根据**Taylor**公式，有

$$g_j(\boldsymbol{x}^{(0)}+\lambda \boldsymbol{p}) = g_j(\boldsymbol{x}^{(0)}) + \lambda \nabla g_j(\boldsymbol{x}^{(0)})^T \boldsymbol{p} + O(\lambda)$$

$\| \boldsymbol{p} \| = 1$

当λ足够小时，如果 $j \notin J$，假设$\boldsymbol{g_j(x)}$连续，有

$$g_j(\boldsymbol{x}^{(0)} + \lambda \boldsymbol{p}) \geq 0$$

如果，$j \in J$，当 $\nabla g_j(\boldsymbol{x}^{(0)})^T \boldsymbol{p} > 0$时，也有

$$g_j(\boldsymbol{x}^{(0)} + \lambda \boldsymbol{p}) \geq 0$$

几何含义：与所有起作用约束梯度的夹角小于**90°**的方向。

下降方向

定义：令目标函数值 $f(x^{(0)})$ 下降的方向，满足

$\exists \lambda_0 > 0$，对于 $\forall \lambda \in (0, \lambda_0]$，有 $f(\boldsymbol{x}^{(0)} + \lambda \boldsymbol{p}) < f(\boldsymbol{x}^{(0)})$

判别条件： $\nabla f(\boldsymbol{x}^{(0)})^T \boldsymbol{p} < 0$ （充分条件）

证明：**Taylor公式**

$$f(\boldsymbol{x}^{(0)} + \lambda \boldsymbol{p}) = f(\boldsymbol{x}^{(0)}) + \lambda \nabla f(\boldsymbol{x}^{(0)})^T \boldsymbol{p} + O(\lambda)$$

几何含义：与目标函数的梯度夹角大于**90°**的方向

局部极小值存在的直观条件

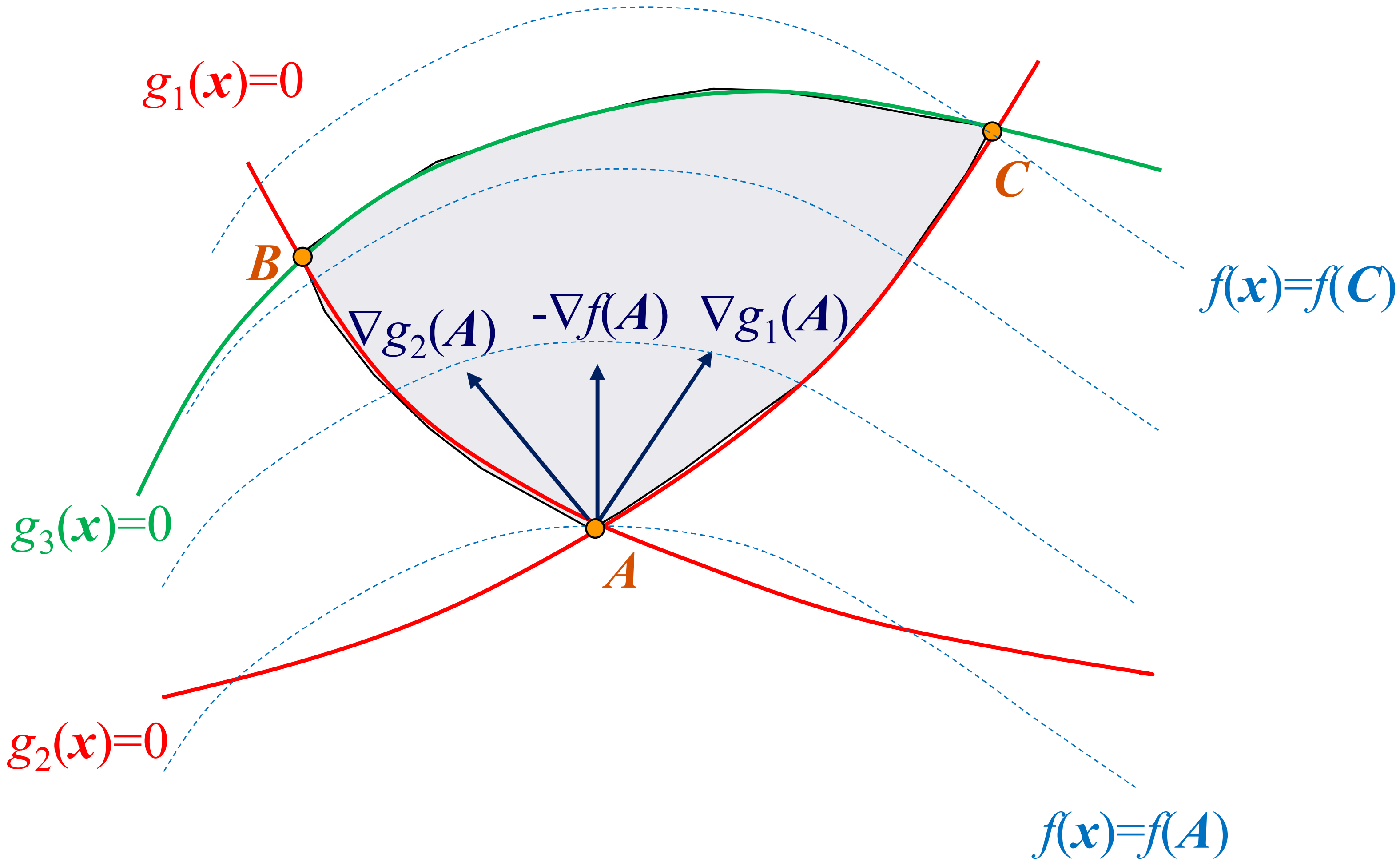
不存在 *可行下降方向* $\boldsymbol{p}$，即不存在同时满足下面两类不等式的方向。

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{p} < 0$$

$$\nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{p} > 0 \qquad j \in J(\boldsymbol{x}^*) \qquad J\text{为起作用约束集合}$$

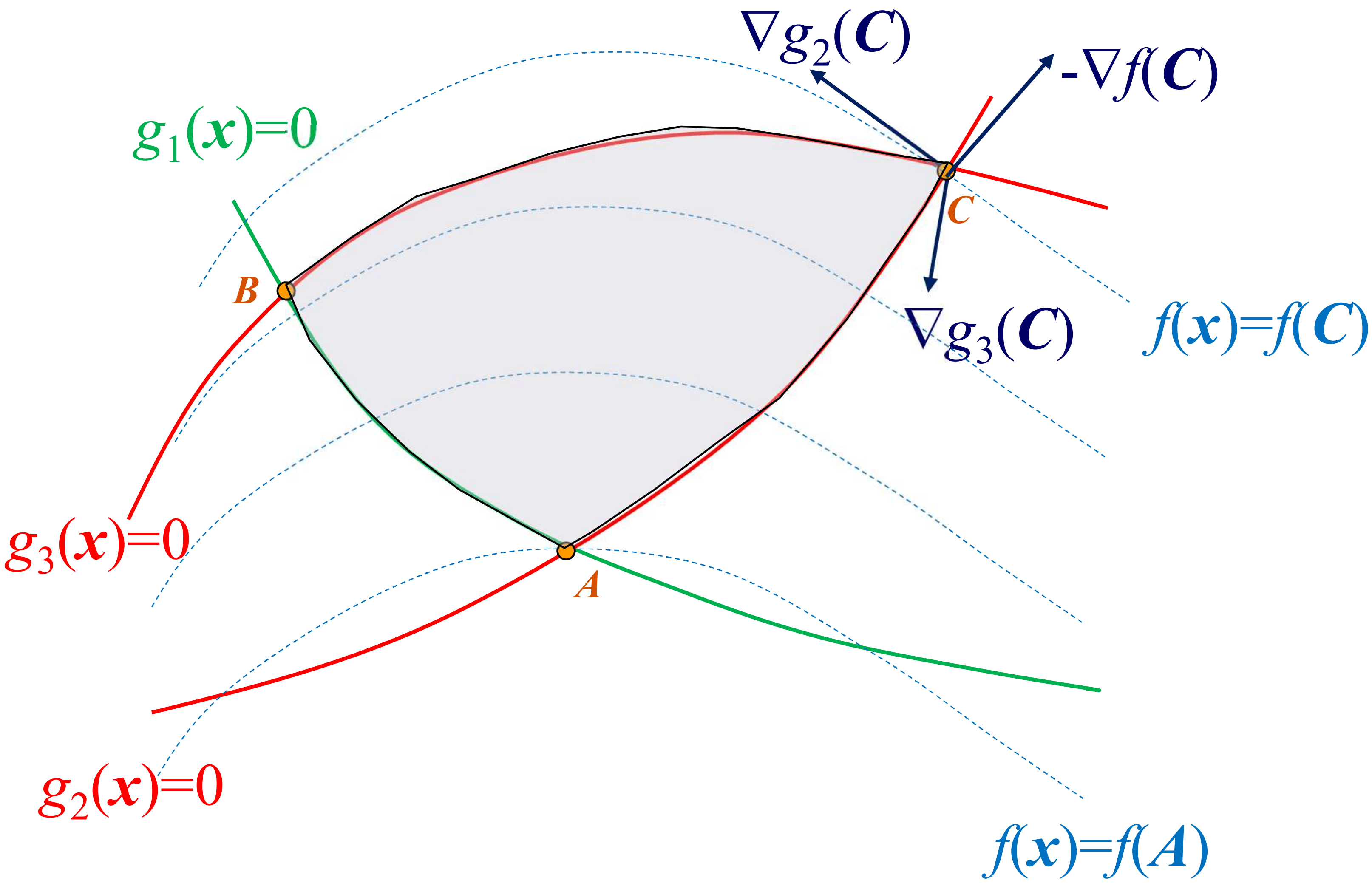
几何含义：不存在与$-\nabla f(\boldsymbol{x}^*)$和所有的 $\nabla g_{j\in J}(\boldsymbol{x}^*)$均成锐角的方向。

图示1：有可行下降方向



$g_1(\boldsymbol{x})=0$
B
C
$f(\boldsymbol{x})=f(\boldsymbol{C})$
$\nabla g_2(\boldsymbol{A})$
$-\nabla f(\boldsymbol{A})$
$\nabla g_1(\boldsymbol{A})$
$g_3(\boldsymbol{x})=0$
A
$g_2(\boldsymbol{x})=0$
$f(\boldsymbol{x})=f(\boldsymbol{A})$

图示2：无可行下降方向



$\nabla g_2(\boldsymbol{C})$
$-\nabla f(\boldsymbol{C})$
$g_1(\boldsymbol{x})=0$
B
C
$\nabla g_3(\boldsymbol{C})$
$f(\boldsymbol{x})=f(\boldsymbol{C})$
$g_3(\boldsymbol{x})=0$
A
$g_2(\boldsymbol{x})=0$
$f(\boldsymbol{x})=f(\boldsymbol{A})$

Gordan引理

➢ $\boldsymbol{a}_j$为一组已知向量，不存在向量$\boldsymbol{p}$，使

$$\boldsymbol{a}_j^T \boldsymbol{p} < 0 \qquad j = 1, 2, \ldots, l$$

同时成立的充要条件：

存在不全为零的非负实数$\mu_j \geq 0$，使

$$\sum_{j=1}^{l} \mu_j \boldsymbol{a}_j = 0$$

正线性相关（**positive linear dependence** ）

几何含义：$\boldsymbol{a}_j$不可能分布在任何超平面的同一侧

正线性相关⇒线性相关

Fritz John条件

- 若$\boldsymbol{x}^*$是局部极小点，则存在不全为零的非负实数$\mu_j \geq 0$，$j \in J(\boldsymbol{x}^*) \cup \{0\}$，使

$$\mu_0 \nabla f(\boldsymbol{x}^*) - \sum_{j \in J(\boldsymbol{x}^*)} \mu_j \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}$$

证明：**Gordan**引理的直接推论

$$\nabla f(\boldsymbol{x}^*)^T \boldsymbol{p} < 0$$

$$-\nabla g_j(\boldsymbol{x}^*)^T \boldsymbol{p} < 0 \qquad j \in J(\boldsymbol{x}^*)$$

问题：上述形式使用上是否方便？

等价Fritz John条件

➢ 若x^*是局部极小点，存在不全为零的非负实数$\mu_j \geq 0$，$j=0,1,2,\ldots,l$，使下列条件成立：

Fritz John条件

$$\mu_0 \nabla f(\boldsymbol{x}^*) - \sum_{j=1}^{l} \mu_j \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}$$

Lagrange函数驻点条件

$$\mu_j g_j(\boldsymbol{x}^*) = 0 \qquad j = 1, 2, \ldots, l$$

互补松弛条件

$$\mu_j \geq 0 \quad \sum_{j=0}^{l} \mu_j \neq 0 \qquad j = 0, 1, \ldots, l$$

强非负条件

上述条件隐含了如下事实：$j \notin J(\boldsymbol{x}^*) \Rightarrow g_j(\mathbf{x}^*) > 0 \Rightarrow \mu_j = 0$

Kuhn-Tucker条件

- 若**Fritz John**条件的$\mu_0>0$，则可推出**KKT**条件的数学形式：

$$
\text{Kuhn-Tucker条件}\quad\begin{cases}
\nabla f(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\mathbf{0} & & \text{Lagrange驻点条件}\\
\mu_j^* g_j(\boldsymbol{x}^*)=0 & j=1,2,\ldots,l & \text{互补松弛条件}\\
\mu_j^*\geq 0 & j=1,2,\ldots,l & \text{非负条件}
\end{cases}
$$

Karush(1939), Kuhn-Tucker (1951)

问题：什么条件下$\mu_0>0$？

$$\mu_0 \nabla f(\boldsymbol{x}^*) - \sum_{j \in J(\boldsymbol{x}^*)} \mu_j \nabla g_j(\boldsymbol{x}^*) = \boldsymbol{0}$$

正则条件

- 正则条件(**regular condition**)：起作用约束$\nabla \boldsymbol{g}_{j\in J}(\boldsymbol{x})$线性无关。

- 性质：若极小值x^*满足正则条件，则**KKT**条件成立。

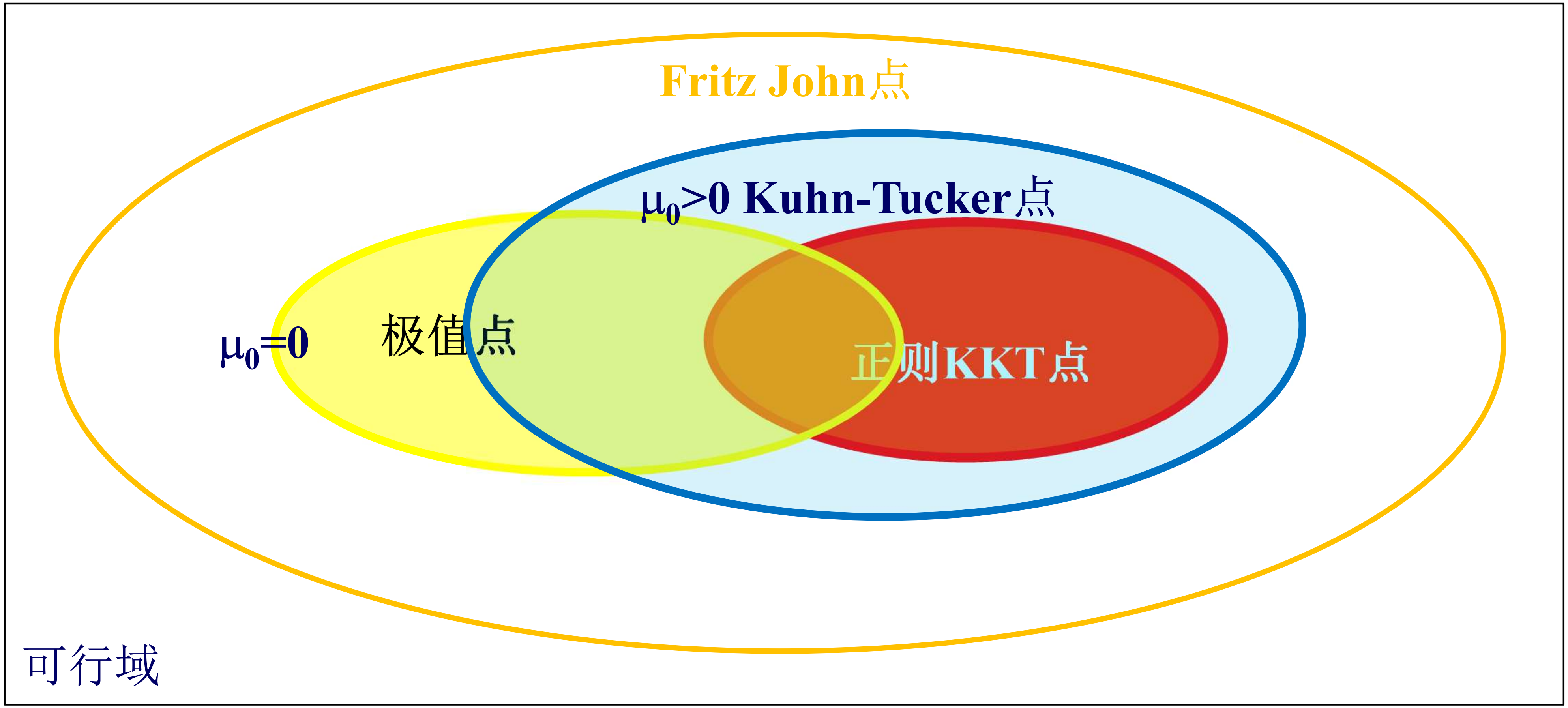
  证明：x^*满足正则条件，**Fritz John**条件中的$\mu_0>0$。

◆ **Kuhn-Tucker**定理：若x^*是局部极小点，且满足正则条件[*约束规格*]，则**Kuhn-Tucker**条件成立。

- 约束规格(**constraint qualification**) :使结论成立的某种前提。

  注意：约束规格不唯一，正则条件只是其中之一。

集合关系



非线性规划的解析解

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束 （**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

一般约束极值问题

$$\min f(\boldsymbol{x})$$

s.t. $h_i(\boldsymbol{x}) = 0 \qquad i = 1, 2, \ldots, m$

$$g_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, l$$

$$\boldsymbol{x} \in R^n$$

求解思路：化为不等式约束极值问题

等价不等式约束极值问题

$$\min f(\boldsymbol{x})$$

s.t.

$$h_i(\boldsymbol{x}) \geq 0 \qquad i = 1, 2, \ldots, m$$

$$-h_i(\boldsymbol{x}) \geq 0 \qquad i = 1, 2, \ldots, m$$

$$g_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, l$$

$$\boldsymbol{x} \in R^n$$

Fritz John条件推导

➢根据不等式约束Fritz John条件，有

$$\mu_0^*\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}\mu'^*_i\nabla h_i(\boldsymbol{x}^*)+\sum_{i=1}^{m}\mu''^*_i\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\mathbf{0}$$

$$\Downarrow$$

$$\mu_0^*\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}(\mu'^*_i-\mu''^*_i)\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\mathbf{0}$$

$$\Downarrow$$

$$\mu_0^*\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}\gamma_i^*\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\mathbf{0}$$

$\mu'^*_i\geq 0\quad \mu''^*_i\geq 0\quad \Rightarrow\quad \gamma_i^*=\mu'^*_i-\mu''^*_i$无符号约束 $\quad i=1,2,\ldots,p$

注意μ_0、μ_j、γ_i不可同时为0！

一般约束下的Fritz John条件

- 若x^*是局部极小点，存在不全为零的μ_j ($j=0,1,2,\ldots,m$)和γ_i ($i=0,1,2,\ldots,p$)，满足

$$\mu_0^*\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}\gamma_i^*\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\boldsymbol{0}$$ 稳定性条件

$$\mu_j^* g_j(\boldsymbol{x}^*)=0 \qquad j=1,2,\ldots,m$$ 互补松弛条件

$$\mu_j^*\geq 0 \qquad j=0,1,\ldots,m \qquad \sum_{j=0}^{l}\mu_j^*+\sum_{i=1}^{m}|\gamma_i^*|\neq 0$$ 强非负条件

$$\gamma_i^* h_i(\boldsymbol{x}^*)=\boldsymbol{0} \qquad i=1,2,\ldots,p$$ 等式互补松弛条件

等式约束自然满足

一般约束下的KKT条件

- 若$\boldsymbol{x}^*$是局部极小点， 等式约束梯度$\nabla h_i(\boldsymbol{x}^*)$和起作用约束梯度$\nabla g_j(\boldsymbol{x}^*)$线性无关，则下列条件成立：

$$\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^{m} \gamma_i^* \nabla h_i(\boldsymbol{x}^*) - \sum_{j=1}^{l} \mu_j^* \nabla g_j(\boldsymbol{x}^*) = \mathbf{0}$$

Lagrange驻点条件

$$\mu_j^* g_j(\boldsymbol{x}^*) = 0 \qquad j = 1, 2, \ldots, l$$

互补松弛条件

$$\mu_j^* \geq 0 \qquad j = 1, 2, \ldots, l$$

非负条件

广义Lagrange函数法

定义Lagrange函数

$$L(\boldsymbol{x},\boldsymbol{\gamma},\boldsymbol{\mu})=f(\boldsymbol{x})-\sum_{i=1}^{m}\gamma_i h_i(\boldsymbol{x})-\sum_{j=1}^{l}\mu_j g_j(\boldsymbol{x})$$

γ_i*和μ_j*称为广义Lagrange乘子

类似无约束极值问题，得到：

$$\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}\gamma_i^*\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=\boldsymbol{0}$$

Lagrange驻点条件

$\mu_j^* g_j(\boldsymbol{x}^*)=0 \quad j=1,2,\ldots,l$ 互补松弛条件

$\mu_j^* \geq 0 \quad j=1,2,\ldots,l$ 非负条件

$h_i(\boldsymbol{x}^*)=0 \quad i=1,2,\ldots,m$

$g_j(\boldsymbol{x}^*)\geq 0 \quad j=1,2,\ldots,l$

可行性条件

KKT条件的矩阵形式

$$\boldsymbol{\gamma}^* = \begin{bmatrix} \gamma_1^* \\ \gamma_2^* \\ \vdots \\ \gamma_m^* \end{bmatrix} \qquad \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} h_1(\boldsymbol{x}) \\ h_2(\boldsymbol{x}) \\ \vdots \\ h_m(\boldsymbol{x}) \end{bmatrix} \qquad \boldsymbol{\mu}^* = \begin{bmatrix} \mu_1^* \\ \mu_2^* \\ \vdots \\ \mu_l^* \end{bmatrix} \qquad \boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} g_1(\boldsymbol{x}) \\ g_2(\boldsymbol{x}) \\ \vdots \\ g_l(\boldsymbol{x}) \end{bmatrix}$$

$$\nabla \boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x}) & \nabla h_2(\boldsymbol{x}) & \cdots & \nabla h_m(\boldsymbol{x}) \end{bmatrix}$$

$$\nabla \boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} \nabla g_1(\boldsymbol{x}) & \nabla g_2(\boldsymbol{x}) & \cdots & \nabla g_l(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_2}{\partial x_1} & \cdots & \frac{\partial g_l}{\partial x_1} \\ \frac{\partial g_1}{\partial x_2} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_l}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial x_n} & \frac{\partial g_2}{\partial x_n} & \cdots & \frac{\partial g_l}{\partial x_n} \end{bmatrix} = \left(\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}} \right)^T$$

梯度矩阵与雅可比矩阵是转置关系

KKT条件的矩阵形式

$$\nabla f(\boldsymbol{x}^*) - \nabla \boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{\gamma}^* - \nabla \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = \boldsymbol{0}$$

Lagrange驻点条件

$$\boldsymbol{\mu}^* \odot \boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0} \Longleftrightarrow \mu_j^* g_j(\boldsymbol{x}^*)=0 \quad j = 1, 2, \ldots, l$$

Hadamard product

互补松弛条件

$$\boldsymbol{\mu}^* \geq 0$$

非负条件

$$\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$$

$$\boldsymbol{g}(\boldsymbol{x}^*) \geq \boldsymbol{0}$$

可行性条件

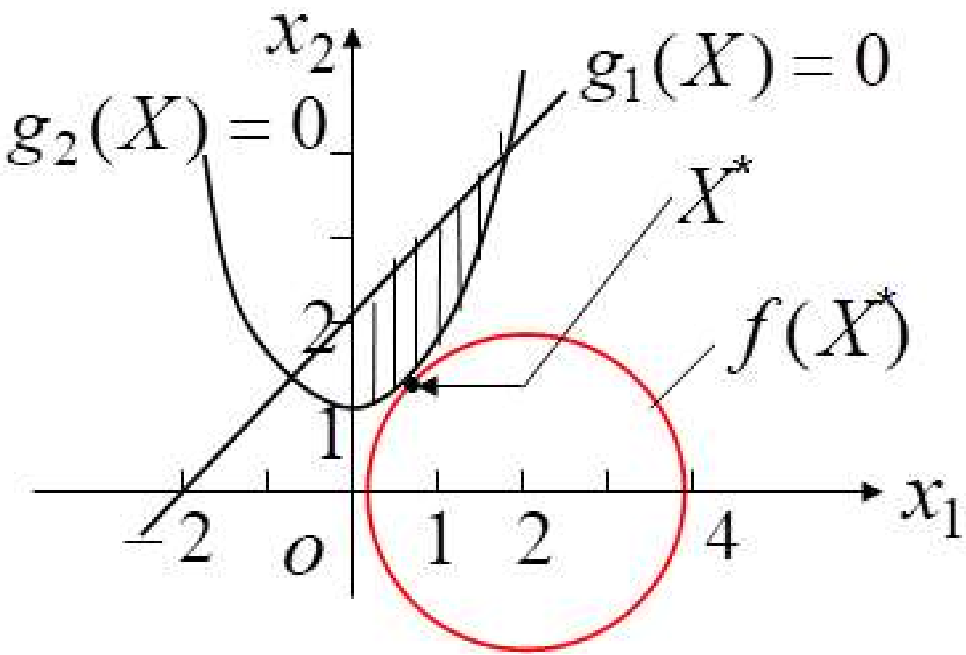
举例

$$\min f(x_1, x_2) = (x_1 - 2)^2 + x_2^2$$

s.t. $x_2 \le x_1 + 2$

$x_2 \ge x_1^2 + 1$

$x_1 \ge 0 \quad x_2 \ge 0$



目标函数、约束函数

$$f(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 2) \\ 2x_2 \end{bmatrix}$$

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} x_1 - x_2 + 2 \\ -x_1^2 + x_2 - 1 \\ x_1 \\ x_2 \end{bmatrix}$$

$$\nabla \boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} 1 & -2x_1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix}$$

KKT条件的矩阵形式

$$\nabla f(\boldsymbol{x}^*) - \nabla \boldsymbol{h}(\boldsymbol{x}^*)\gamma^* - \nabla \boldsymbol{g}(\boldsymbol{x}^*)\boldsymbol{\mu}^* = \mathbf{0}$$

$$\Longrightarrow \begin{bmatrix} 2(x_1 - 2) \\ 2x_2 \end{bmatrix} - \begin{bmatrix} 1 & -2x_1 & 1 & 0 \\ -1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\mu}^* \odot \boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0} \quad \Longrightarrow \quad \begin{bmatrix} \mu_1(x_1 - x_2 + 2) \\ \mu_2(-x_1^2 + x_2 - 1) \\ \mu_3 x_1 \\ \mu_4 x_2 \end{bmatrix} = \mathbf{0}$$

$$\boldsymbol{g}(\boldsymbol{x}^*) \geq \mathbf{0} \qquad \boldsymbol{\mu}^* \geq 0$$

KKT条件的代数形式

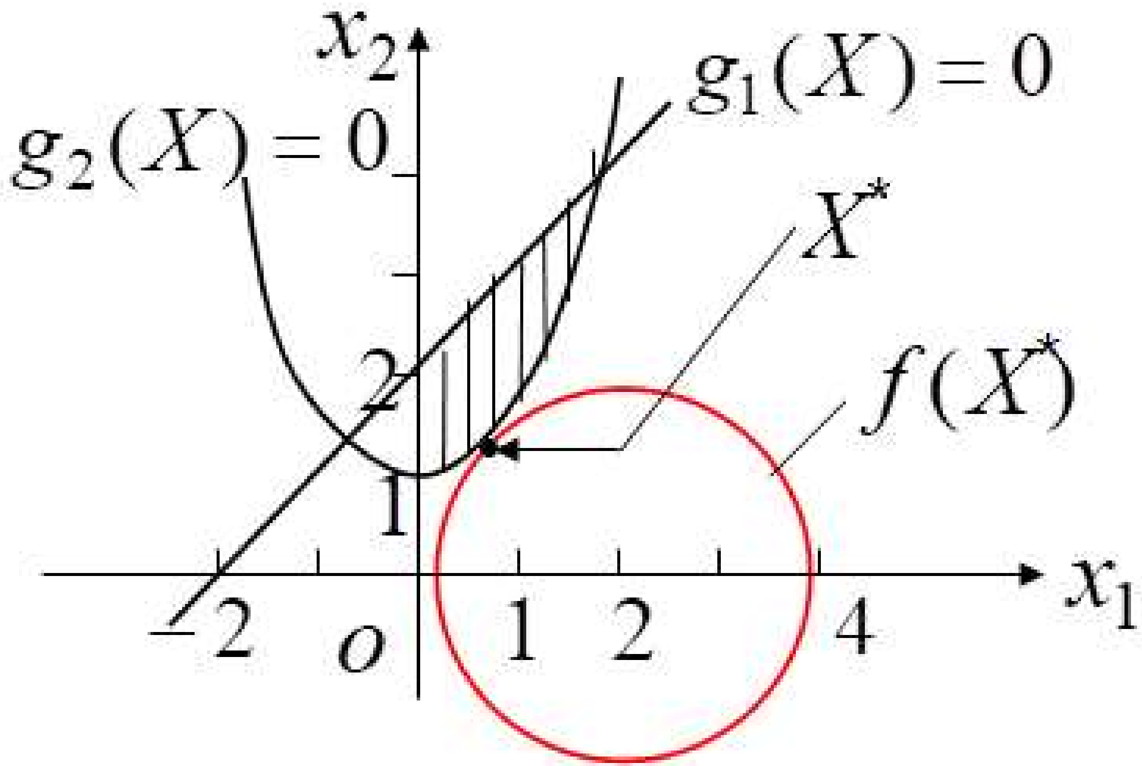
$$\begin{cases} 2(x_1-2)-\mu_1+2\mu_2 x_1-\mu_3=0 \\ 2x_2+\mu_1-\mu_2-\mu_4=0 \\ \mu_1(x_1-x_2+2)=0 \\ \mu_2(-x_1^2+x_2-1)=0 \\ \mu_3 x_1=0 \\ \mu_4 x_2=0 \end{cases}$$

6个变量、6个方程

$$\begin{cases} \mu_j \geq 0 \qquad j=1,2,.3,4 \\ x_2 \leq x_1+2 \qquad x_1 \geq 0 \\ x_2 \geq x_1^2+1 \qquad x_2 \geq 0 \end{cases}$$

不等式条件

图示



求解

观察可得： $\mu_1 = \mu_3 = \mu_4 = 0$

所以有： $(1+\mu_2)x_1 - 2 = 0$

$$2x_2 - \mu_2 = 0$$

$$-x_1^2 + x_2 - 1 = 0$$

求解得： $\mu_2^* = 2.6219 \quad x_1^* = 0.5536 \quad x_2^* = 1.3064$

$$f(\boldsymbol{x}^*) = 3.7989$$

非线性规划的解析解

- 非线性规划问题及其数学模型
- 非线性规划解析解法
  - 无约束极值问题
    - 一阶条件
    - 二阶条件
  - 有约束极值问题：
    - 等式约束
    - 不等式约束（**Kuhn-Tucker**条件）
    - 一般约束
- 凸规划

凸规划

$$\min f(\boldsymbol{x})$$ 凸函数

s.t. $h_i(\boldsymbol{x}) = 0 \quad i = 1, 2, \ldots, m$ 线性函数

$g_j(\boldsymbol{x}) \geq 0 \quad j = 1, 2, \ldots, l$ 凹函数

$$\boldsymbol{x} \in R^n$$

凸函数与凹函数

- 凸函数为满足下列条件的函数：

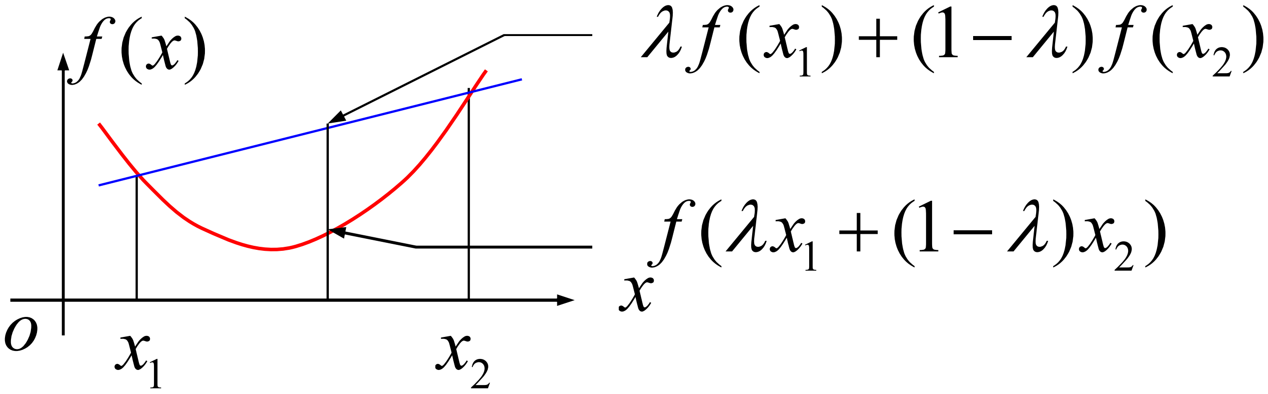
$$f[\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2] \le \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \qquad 0 < \lambda < 1 \qquad \boldsymbol{x}_1, \boldsymbol{x}_2 \in R^n$$

- “<”：严格凸函数
- “≥”：凹函数
- “>”：严格凹函数

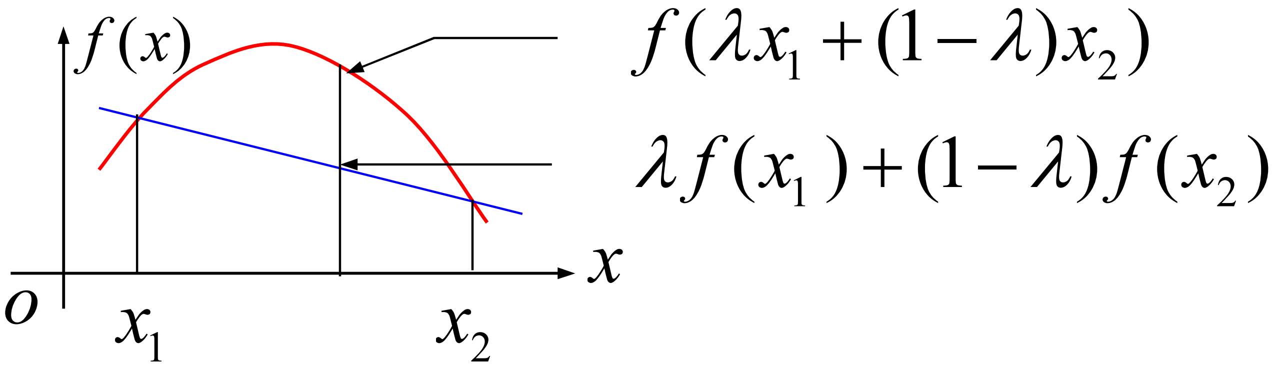
注意：线性函数既是凸函数，又是凹函数，所以线性规划是凸规划。

凸函数和凹函数图示

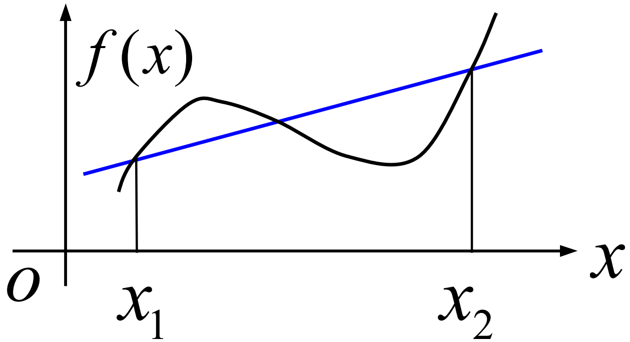
凸函数



凹函数



非凸非凹函数



凸函数判定条件（充要条件）

- 一阶条件

$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in R^n$　恒有　$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^T (\boldsymbol{x}_2 - \boldsymbol{x}_1)$

几何意义：任何一点的切线在凸函数曲线的下方

- 二阶条件

$\forall \boldsymbol{x} \in R^n$　恒有　$\nabla^2 f(\boldsymbol{x}) \geq 0$

几何意义：函数曲线向上弯曲

- **”>”**：严格凸函数
- **”≤”**：凹函数
- **“<“**：严格凹函数
- **“=“**：线性函数

$$\left[\frac{d(\boldsymbol{k}^T\boldsymbol{x})}{d\boldsymbol{x}}\right]^T=\boldsymbol{k}$$

$$\left[\frac{d\left(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\right)}{d\boldsymbol{x}}\right]^T=2\boldsymbol{A}\boldsymbol{x}$$

A对称

二次函数为凸函数的充要条件

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\mathbf{H}\boldsymbol{x}+\boldsymbol{c}^T\boldsymbol{x}+r$$

H为对称矩阵

$$\nabla f(\boldsymbol{x})=\mathbf{H}\boldsymbol{x}+\boldsymbol{c}$$

$$\nabla^2 f(\mathbf{X})=\mathbf{H}\geq 0$$

H为半正定矩阵

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^T (\boldsymbol{x}_2 - \boldsymbol{x}_1)$$

凸函数的性质

- 凸函数非负线性组合仍为凸函数

$$\gamma_1 f_1(\boldsymbol{x}) + \gamma_2 f_2(\boldsymbol{x}) \qquad \gamma_i \geq 0$$

- 若$f(x)$是定义在凸集R_C上的凸函数，则其β水平集S_β为凸集。

$$S_\beta = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq \beta, \boldsymbol{x} \in R_C\}$$ β水平集

- 对于凸函数$f(x)$，若存在$x^* \in R^n$满足

$$\nabla f(\boldsymbol{x}^*)^T (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \qquad \forall \boldsymbol{x} \in R^n$$ 充要条件！

则x^*为$f(x)$的全局最小点。

推论1：对于凸目标函数，$\nabla f(x^*)=0$ 是x^*为极小值的充要条件。

推论2：对于凸目标函数，局部极小点($\nabla f(X^*)=0$)也是全局最小点。

凸集性质

- 凸规划的可行域为凸集

$$h_i(\boldsymbol{x}) = 0 \qquad -g_j(\boldsymbol{x}) \le 0$$

各约束的**0**水平集为凸集

凸集的交集为凸集

- 如果最优解存在，最优解集合也为凸集

$$f[\lambda \boldsymbol{x}_1^* + (1-\lambda)\boldsymbol{x}_2^*] \le \lambda f(\boldsymbol{x}_1^*) + (1-\lambda) f(\boldsymbol{x}_2^*) = f(\boldsymbol{x}_1^*) = f(\boldsymbol{x}_2^*) \qquad 0 < \lambda < 1$$

$$\Rightarrow \quad f[\lambda \boldsymbol{x}_1^* + (1-\lambda)\boldsymbol{x}_2^*] = f(\boldsymbol{x}_1^*) = f(\boldsymbol{x}_2^*)$$

最优解的连线段均为最优解

推论：线性规划问题的最优解集为所有最优顶点构成的多边形。（归纳法证）

$$\boldsymbol{x}^* = \sum_{i=1}^{r} a_i \boldsymbol{x}_i^* \qquad \sum_{i=1}^{r} a_i = 1 \qquad 0 \le a_i \le 1 \qquad i = 1, \cdots, r$$

凸规划性质

- 任何局部极值解也是全局最优解　（目标函数为凸函数）

  局部极小点和全局最小点连线的目标函数值相同

- 若目标函数为严格凸函数，则如果全局最优解存在，必为唯一全局最优解。（反证法）

$$f[\lambda \boldsymbol{x}_1^* + (1-\lambda)\boldsymbol{x}_2^*] < \lambda f(\boldsymbol{x}_1^*) + (1-\lambda) f(\boldsymbol{x}_2^*) = f(\boldsymbol{x}_1^*) = f(\boldsymbol{x}_2^*)$$

  最优解的唯一性为数值解法提供了方便。

- 凸规划下的**KKT**条件为最优解的充要条件

KT条件的充分性证明

$$\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{m}\gamma_i^*\nabla h_i(\boldsymbol{x}^*)-\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)=0 \qquad \mu_j^* g_j(\boldsymbol{x}^*)=0 \qquad \mu_j^*\ge 0$$

***f*(*x*)是凸函数** $\Rightarrow$

$$f(\boldsymbol{x})\ge f(\boldsymbol{x}^*)+\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*)$$

$$=f(\boldsymbol{x}^*)+\sum_{i=1}^{m}\gamma_i^*\nabla h_i(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*)+\sum_{j=1}^{l}\mu_j^*\nabla g_j(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*)$$

$h_i(x)$线性函数且$h_i(x)=0$ $\Rightarrow$

$$\nabla h_i(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*)=h_i(\boldsymbol{x})-h_i(\boldsymbol{x}^*)=0$$

$g_j(x)$是凹函数 $\Rightarrow$

$$\nabla g_j(\boldsymbol{x}^*)^T(\boldsymbol{x}-\boldsymbol{x}^*)\ge g_j(\boldsymbol{x})-g_j(\boldsymbol{x}^*)$$

根据互补松弛性条件 $\mu_j^* g_j(x^*)=0$ 且$g_j(x)\ge 0$，有

$$f(\boldsymbol{x})\ge f(\boldsymbol{x}^*)+\sum_{j=1}^{l}\mu_j^*\left[g_j(\boldsymbol{x})-g_j(\boldsymbol{x}^*)\right]=f(\boldsymbol{x}^*)+\sum_{j=1}^{l}\mu_j^* g_j(\boldsymbol{x})\ge f(\boldsymbol{x}^*)$$

举例

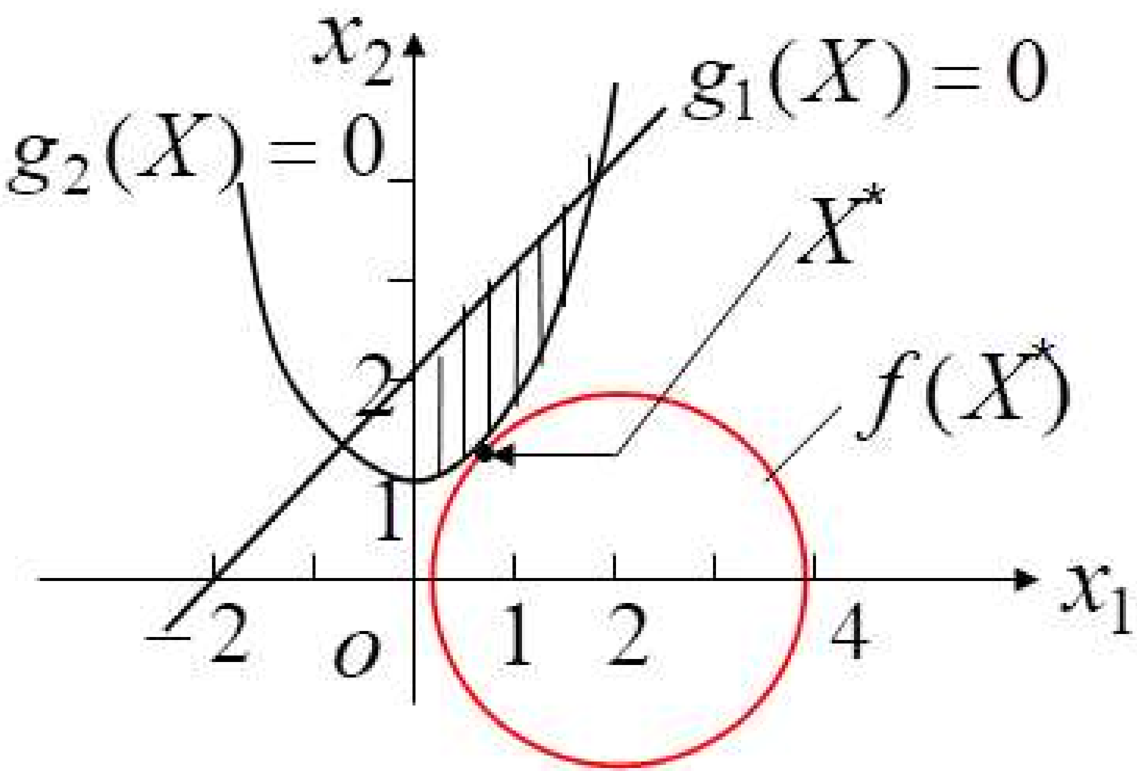
$$\min f(x_1, x_2) = (x_1 - 2)^2 + x_2^2$$

s.t. $x_2 \le x_1 + 2$

$x_2 \ge x_1^2 + 1$

$x_1 \ge 0$

$x_2 \ge 0$



凸规划判断

$$f(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2 \Rightarrow \nabla f(\boldsymbol{x}) = \begin{bmatrix} 2(x_1 - 2) \\ 2x_2 \end{bmatrix} \Rightarrow \nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0$$

$$g_1(\boldsymbol{x}) = x_1 - x_2 + 2 \Rightarrow \nabla g_1(\boldsymbol{x}) = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \nabla^2 g_1(\boldsymbol{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$g_2(\boldsymbol{x}) = -x_1^2 + x_2 - 1 \Rightarrow \nabla g_2(\boldsymbol{x}) = \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix} \Rightarrow \nabla^2 g_2(\boldsymbol{x}) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} \le 0$$

$$g_3(\boldsymbol{x}) = x_1 \Rightarrow \nabla g_3(\boldsymbol{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \nabla^2 g_3(\boldsymbol{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$g_4(\boldsymbol{x}) = x_2 \Rightarrow \nabla g_4(\boldsymbol{x}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \nabla^2 g_4(\boldsymbol{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

小结

- 无约束优化问题
  - 目标函数为凸函数
    - $\nabla f(x^*)=0$是极小值的充要条件
    - 局部极小值为全局极小值
  - 目标函数为严格凸函数
    - 局部极小值为唯一的极小值
- 凸规划
    - **KKT**条件为极小值的充要条件
    - 局部极小值为全局极小值

非线性规划的matlab求解

- 线性规划(LP)：**linprog**
- 混合整数线性规划 (MILP)：**intlinprog**
- 二次规划(QP)：**quadprog**
- 二阶锥规划(SOCP)：**coneprog**
- 半定规划(SDP)：**Yalmip**中调用SDP求解器

> 无约束极值问题：**fminunc**
> 有约束极值问题：**fmincon**

LPs⊂QPs ⊂ QCQPs ⊂ SOCPs ⊂ SDPs ⊂ 锥规划CPs

二次规划

■ （凸）二次规划（**QP，Quadratic Program**）：

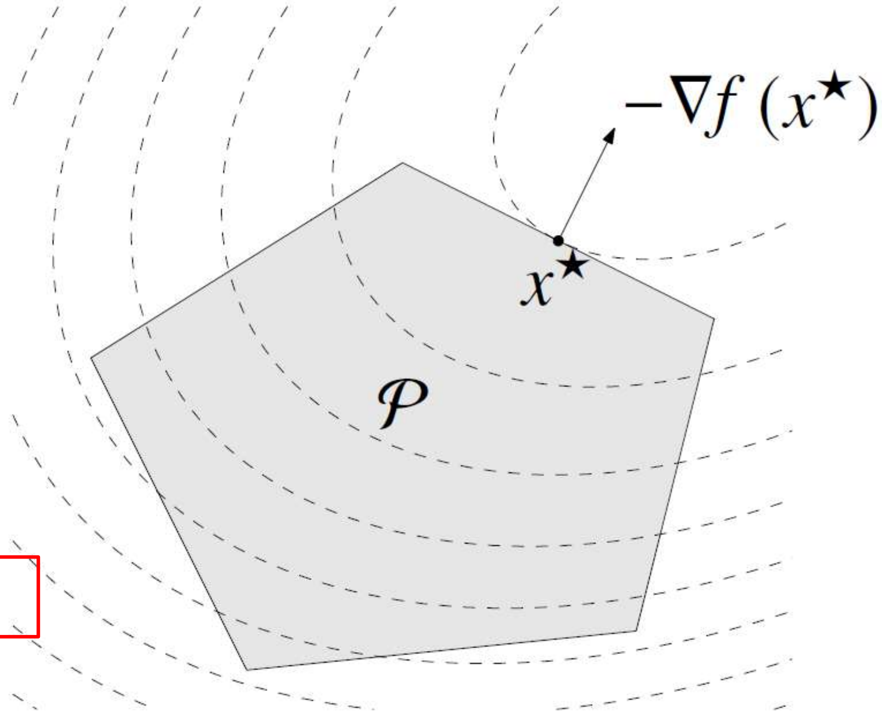
$$\min \frac{1}{2} \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} + \boldsymbol{q}^T \boldsymbol{x} + r$$

$$\text{s.t.} \quad \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h}$$

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

线性约束

$$\boldsymbol{x} \in R^n$$



$\boldsymbol{P} \in S_+^n$ $\quad \boldsymbol{G} \in R^{l\times n}$ $\quad \boldsymbol{A} \in R^{m\times n}$

半正定对称矩阵集合

LPs⊂QPs

二次约束二次规划

■ 二次约束二次规划问题

（**QCQP，Quadratically Constrained Quadratic Program)：**

$$\min \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}+\boldsymbol{q}^T\boldsymbol{x}+r$$

$$\text{s.t.}\quad \frac{1}{2}\boldsymbol{x}^T\boldsymbol{P}_j\boldsymbol{x}+\boldsymbol{q}_j^T\boldsymbol{x}+r_j\le 0 \qquad j=1,\cdots,l$$

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

线性约束

$$\boldsymbol{x}\in R^n$$

$$\boldsymbol{P},\boldsymbol{P}_j\in S_+^n \qquad \boldsymbol{A}\in R^{m\times n}$$

QPs⊂QCQPs

二阶锥规划

■ 二阶锥规划问题

（**SOCP，Second-Order Cone Program):**

$$\min \boldsymbol{c}^T \boldsymbol{x}$$

$$\text{s.t.} \quad \| \boldsymbol{A}_j \boldsymbol{x} + \boldsymbol{b}_j \|_2 \le \boldsymbol{c}_j^T \boldsymbol{x} + d_j \qquad j = 1, \cdots, l$$

二阶锥约束

$$\boldsymbol{F}\boldsymbol{x} = \boldsymbol{g}$$

$$\boldsymbol{x} \in R^n$$

半定规划

- 半定规划（SDP，Semidefinite program):
  - LMI标准模型：

$$\min \boldsymbol{c}^T \boldsymbol{x}$$

$$\text{s.t.}\quad x_1\boldsymbol{F}_1 + \cdots + x_n\boldsymbol{F}_n + \boldsymbol{G} \preceq 0$$

线性矩阵不等式
LMI(linear matrix inequality)

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

$A \succeq B \Leftrightarrow A - B$半正定

$$\boldsymbol{x} \in R^n$$

$$\boldsymbol{G}, \boldsymbol{F}_j \in S^k \qquad \boldsymbol{A} \in R^{p\times n}$$

半定规划可视为线性规划在矩阵参数空间的推广！