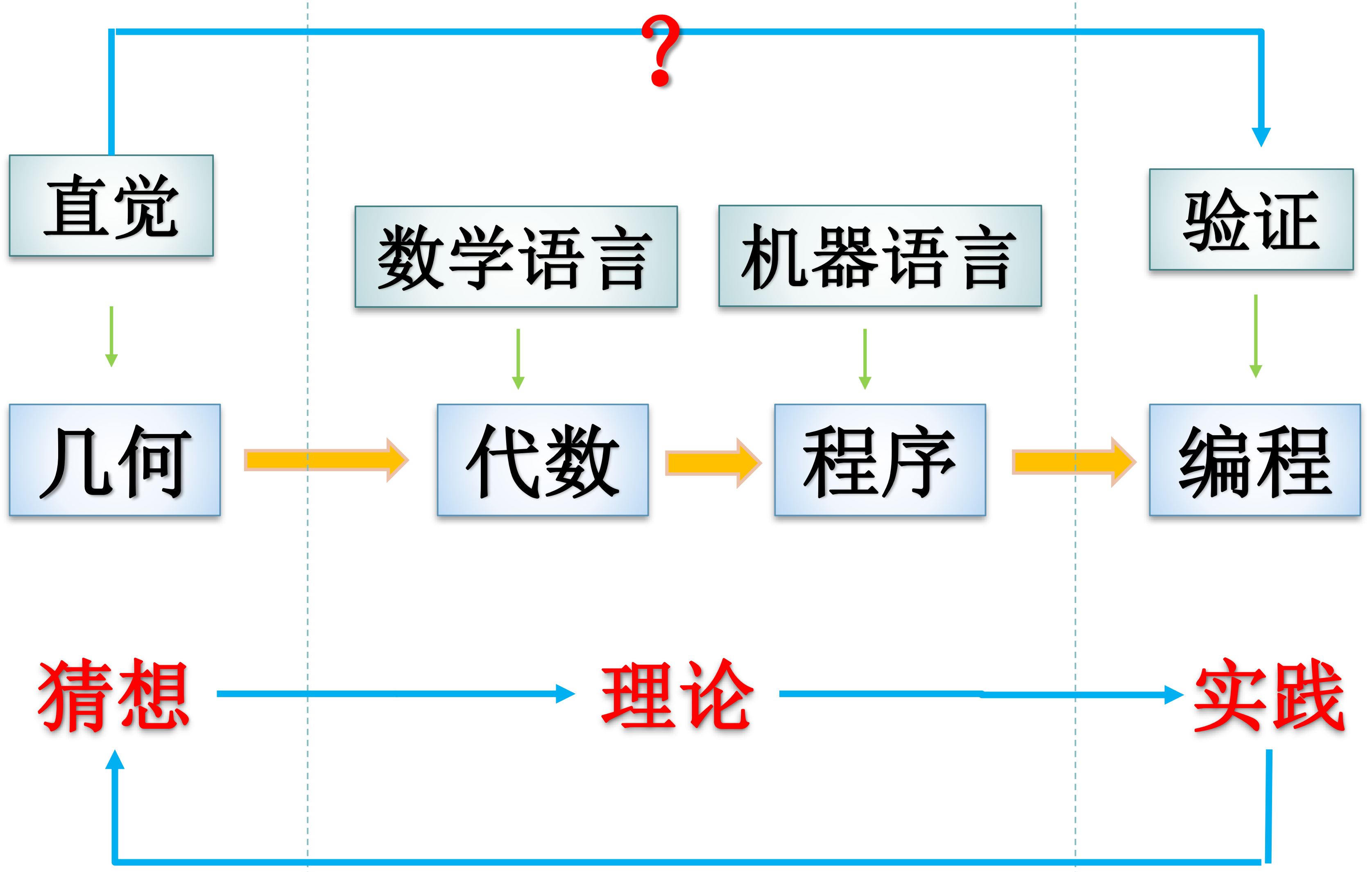
2019

最优化理论与方法

研究生学位课

陈军华（副教授/主任）

- **课程学习的两个目标**
  - 最优化的概念、理念、思路、方法
  - 在计算上编程实现
- **基础知识**
  - 线性规划
  - 微积分
  - 矩阵、线代
  - **Matlab, Gams(cplex); Python, C#, C++.**


?
直觉
数学语言
机器语言
验证
几何
代数
程序
编程
猜想
理论
实践

参考书籍

- **Linear and Nonlinear Programming, David G. Luenberger Stanford University,2008**
- 陈宝林，最优理论与方法，清华大学出版社，**1989**
- 袁亚湘，孙文瑜，最优化理论与方法，科学出版社，**1997**
- **Dimitri P.Bertsekas,Nonlinear Programming, second Edition, Athena Scientific, Belmont, MA,1999**
- 陈军华，铁路运输经典问题建模与实现，**2019**

- 龚纯，王正林，精通**MATLAB**最优化计算，第**2**版，**2012**
- 宋叶志，徐导，**C#**科学计算讲义，**2012**

- **学习资料-云盘分享**

链接：ttps://pan.baidu.com/s/1g4D42OD3WAaz8lkWbLH7_Q
提取码：ml0o

学习框架

- 基本理论
  - 凸函数及其性质
  - 函数极值存在条件
  - 等式约束的最优性条件
  - K-T条件
- 线性规划
- 基本方法
  - 单峰函数极值问题
    - 试探法（区间分割）
      - 二分法
      - 0.618
      - Fibonacci
    - 函数逼近法（插值法）
      - 牛顿法
      - 多项式逼近法
  - ……………………………

- **学习框架**

- 基本方法
  - 单峰函数极值问题
    - 试探法（区间分割）
      - 二分法
      - 0.618
      - Fibonacci
    - 函数逼近法（插值法）
      - 牛顿法
      - 多项式逼近法
  - 多维无约束极值问题
    - 解析法
      - 梯度法
      - 二阶导数法
      - 共轭梯度法
      - 变尺度法
    - 直接法
      - 坐标轮换法
      - 单纯形法
  - 有约束非线性规范问题
    - NLP→LP
      - 线性逼近法
      - 二次规划法
    - 约束→无约束
      - 惩罚函数法

- **Example 1**
  - 某公司经营两种物资，第一种每吨售价30元，第二种每吨售价是450元，据统计每售出1吨一、二种物资的平均营业时间为0.5h和（2+0.25x_2）h，其中x_2是第二种物资的出售量。若该公司在这段时间里的总营业时间为800小时，试决定营业额最大的营业计划。

- **Example 1**
  - 解：设x_1，x_2分别为第一、二种物资的出售量。

    其目标是：$\max f(x) = 30x_1 + 450x_2$

    约束条件：$0.5x_1 + (2+0.25x_2)x_2 \le 800$

    $x_1, x_2 \ge 0$

$$
\begin{array}{llcccccl}
 & \text{Maximize} & 30x_1 & + & 450x_2 & & & \\
\text{即：} & s.t. & 0.5x_1 & + & 2x_2 & + & 0.25x_2^2 & \le 800 \\
 & & x^1 & , & x^2 & & & \ge \ 0
\end{array}
$$

象这种含有非线性函数的规划模型就是非线性模型。

- **Example 2**
  - 一投资者拥有资金5万元，准备向两个项目分别以x_1，x_2(万元)投资，两个项目的效益分别为20%，16%。所承担的风险损失数为：$2x_1^2+x_2^2(x_1+x_2)^2$元，投资者希望知道怎样的投资方案既可使收益最大，而所承担的风险最小。

- **Example 2**

  解：该问题要求收益最大，风险最小，这是一个多目标问题，为了求解，引进一个风险损失系数$\theta \geq 0$，将多目标化为单目标问题：

$$\max \mathrm{p} = 2000x_1 + 1600x_2 - \theta\left[2x_1^2 + x_2^2(x_1 + x_2)^2\right]$$

$$s.t.\quad x_1 + x_2 = 5$$

$$x_1, x_2 \geq 0$$

- **建立NLP的模型的步骤**
  - **step1:设变量;**
  - **step2:找目标，建立目标函数;**
  - **step3:寻找全部约束;**
  - **step4:非负约束。**

- **NLP的通用模型**

由上面的例子知NLP有各种形式，我们可以用通用模型表示：

$$
\begin{aligned}
&\min f(x) \\
s.t.\ \ & g_i(x) \geq 0 \quad i = 1, \dots, m \\
& h_i(x) = 0 \quad i = 1, \dots, p \\
& x \in E^n
\end{aligned}
$$

- 其中$m \geq 0, p \geq 0, f(x), g_i(x), h_i(x)$至少有一个是$x$的非线性函数。
- $f(x)$:代价函数（目标函数） Cost function
- x：决策变量
- X：约束集

- **NLP的通用模型**
  - 若令R表示NLP的可行解集合（即可行域），模型又可表示成集合的形式：

$$\min f(x)$$

NLP：

$$\mathbf{R}=\left\{x\middle|g_i(x)\geq 0, i=1,\ldots m; h_j(x)=0, j=1,\ldots,p; x\in E^n\right\}$$

  - 求解非线性规划与求解线性规划一样，就是在可行域中求出目标函数最小值的点，只是对于复杂的规划不易求得其精确解，只能借助于迭代法求得近似最优解。

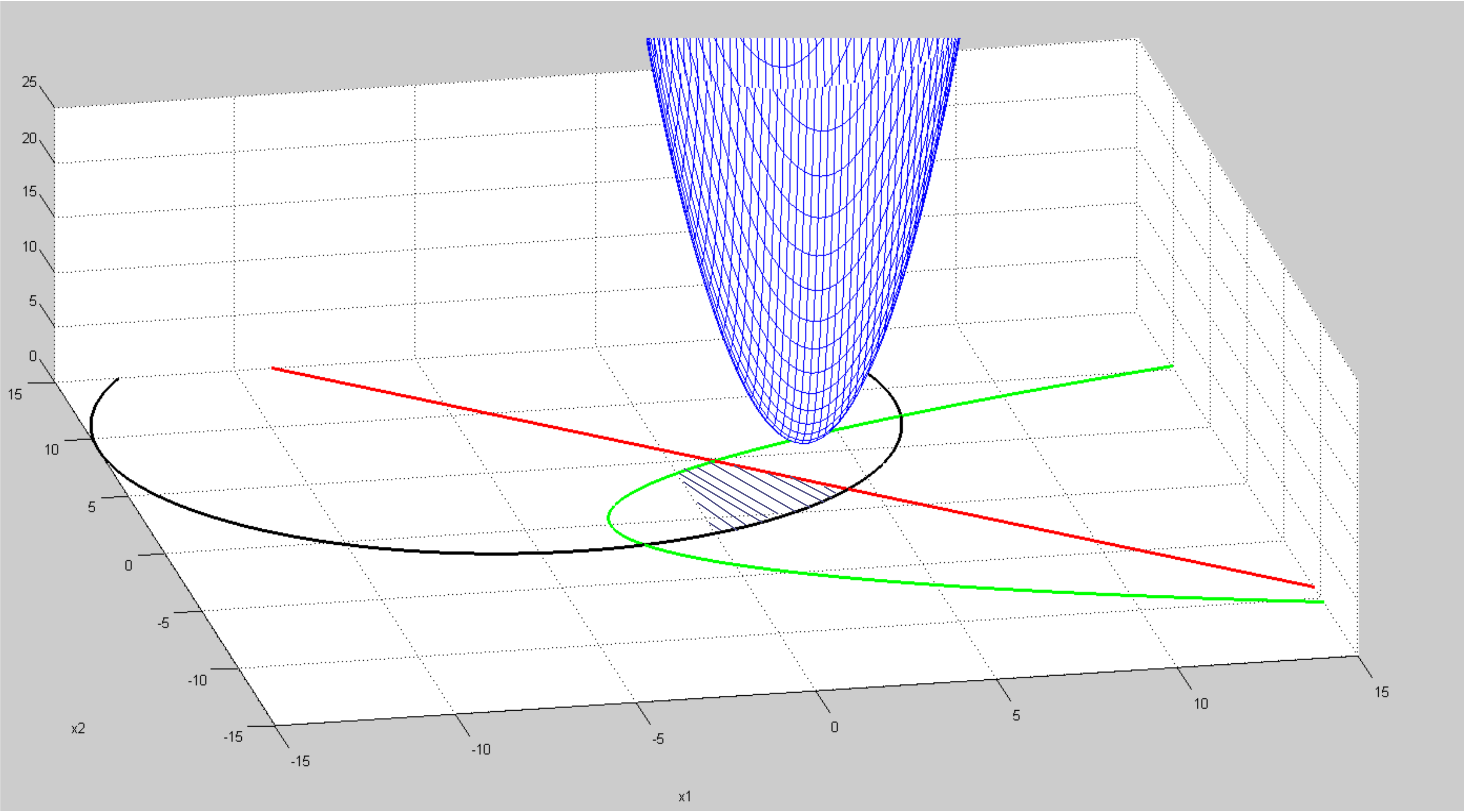
NLP的图解法

- 例：$\min f(x) = 2(x_1 - 3.5)^2 + 4(x_2 - 4)^2 + 2$

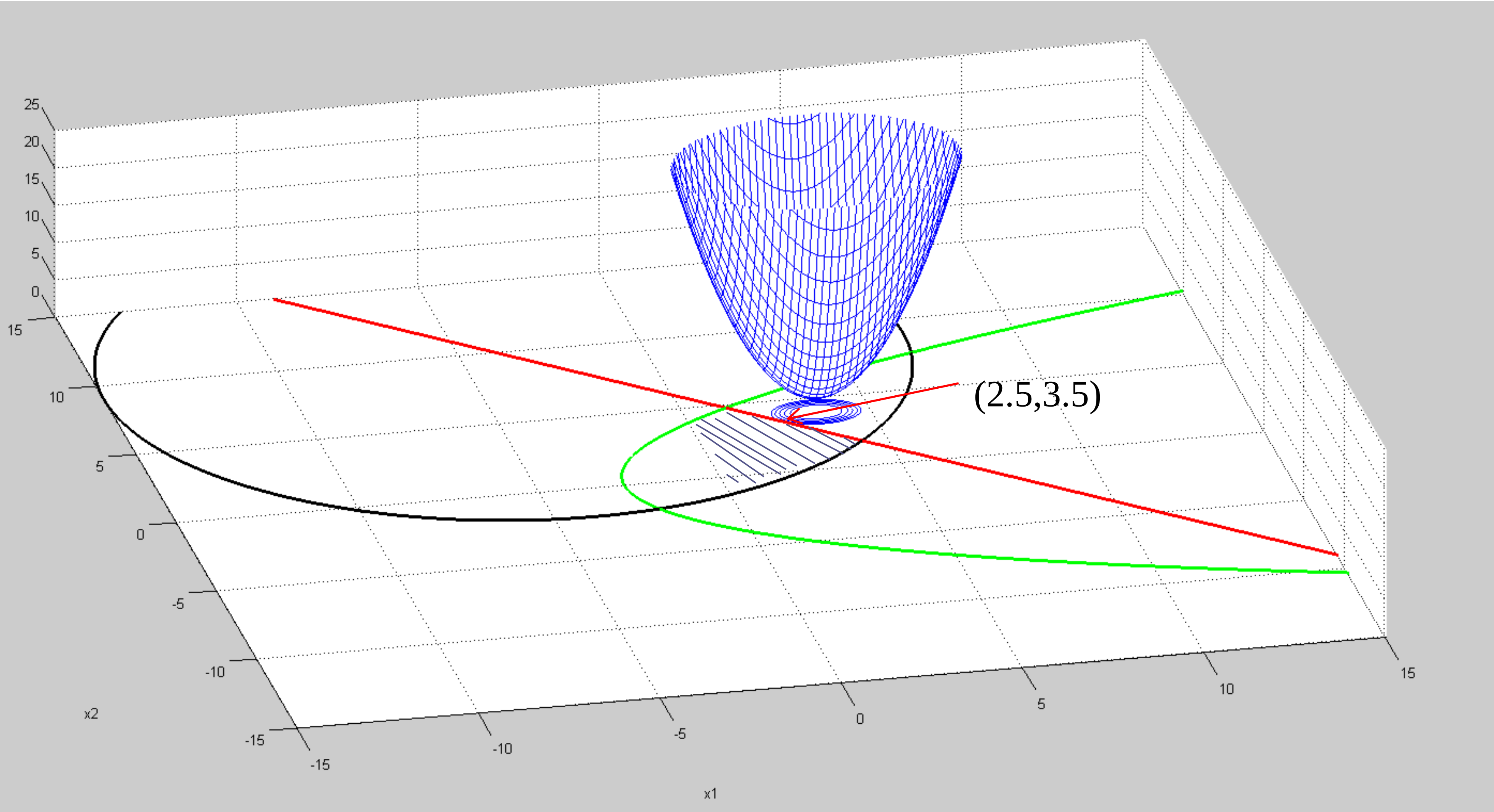
$$
\begin{aligned}
s.t.\quad & x_1 + x_2 - 6 \leq 0 \\
& 6x_1 - x_2^2 + 16 \geq 0 \\
& (x_1 + 4)^2 + (x_2 - 9)^2 - 121 \leq 0 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

- 先在三维空间中划出可行域R,然后画出目标函数的立体图。

- **NLP的图解法(ex1)**



- **NLP的图解法**



- **NLP的最优解与局部最优解**
  - **1. 最优解**

定义1：若$x^*\in R$是NLP的最优解点,当且仅当对于任意的$x\in R$均有:

$$f(x)\geq f(\boldsymbol{x}^*) \qquad (4)$$

若上式为严格不等式,则$\boldsymbol{x}^*$为NLP的严格最优解点。

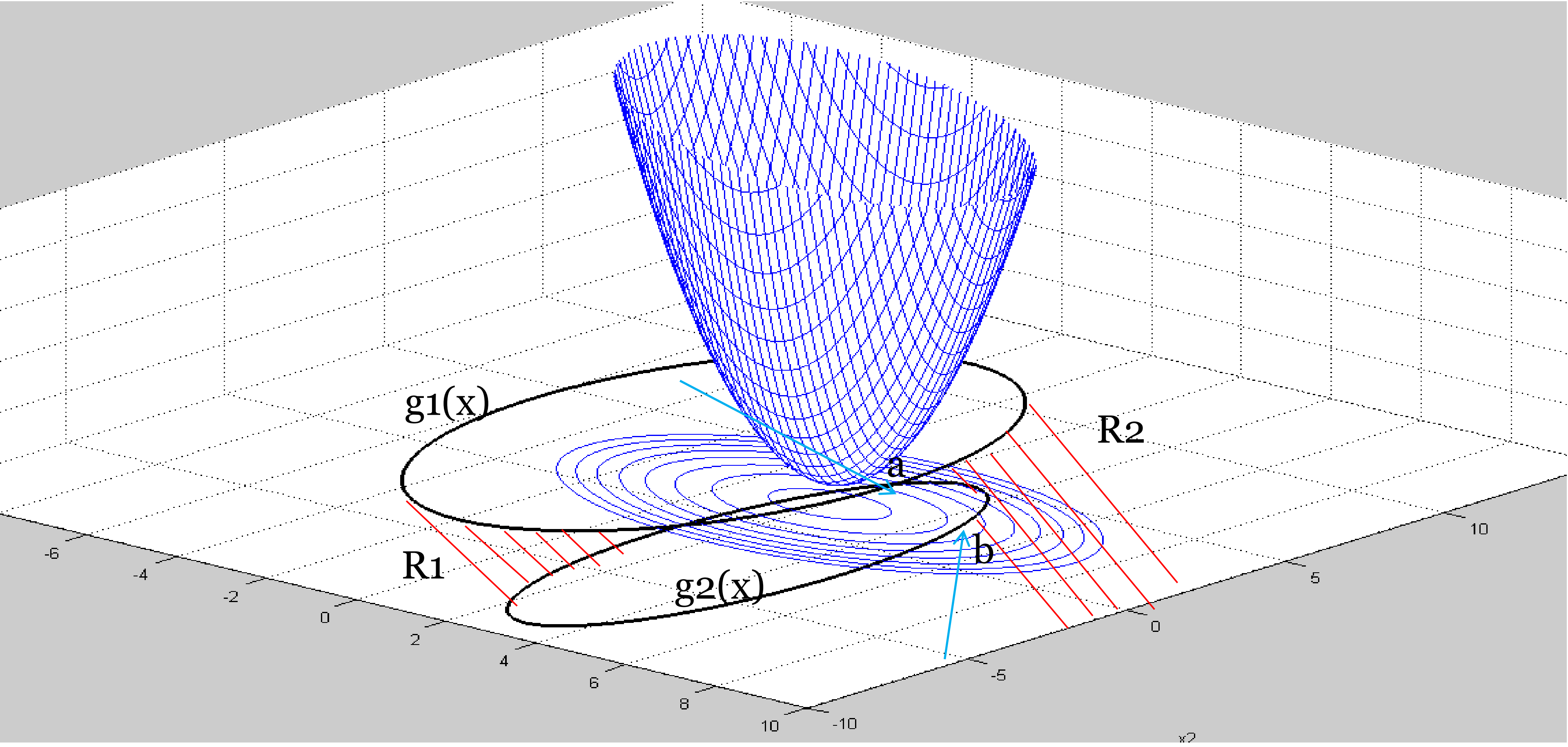
若NLP的最优节点不是唯一，则记最优解集合为$\boldsymbol{R}^*$。

- **NLP的最优解与局部最优解（ex2）**

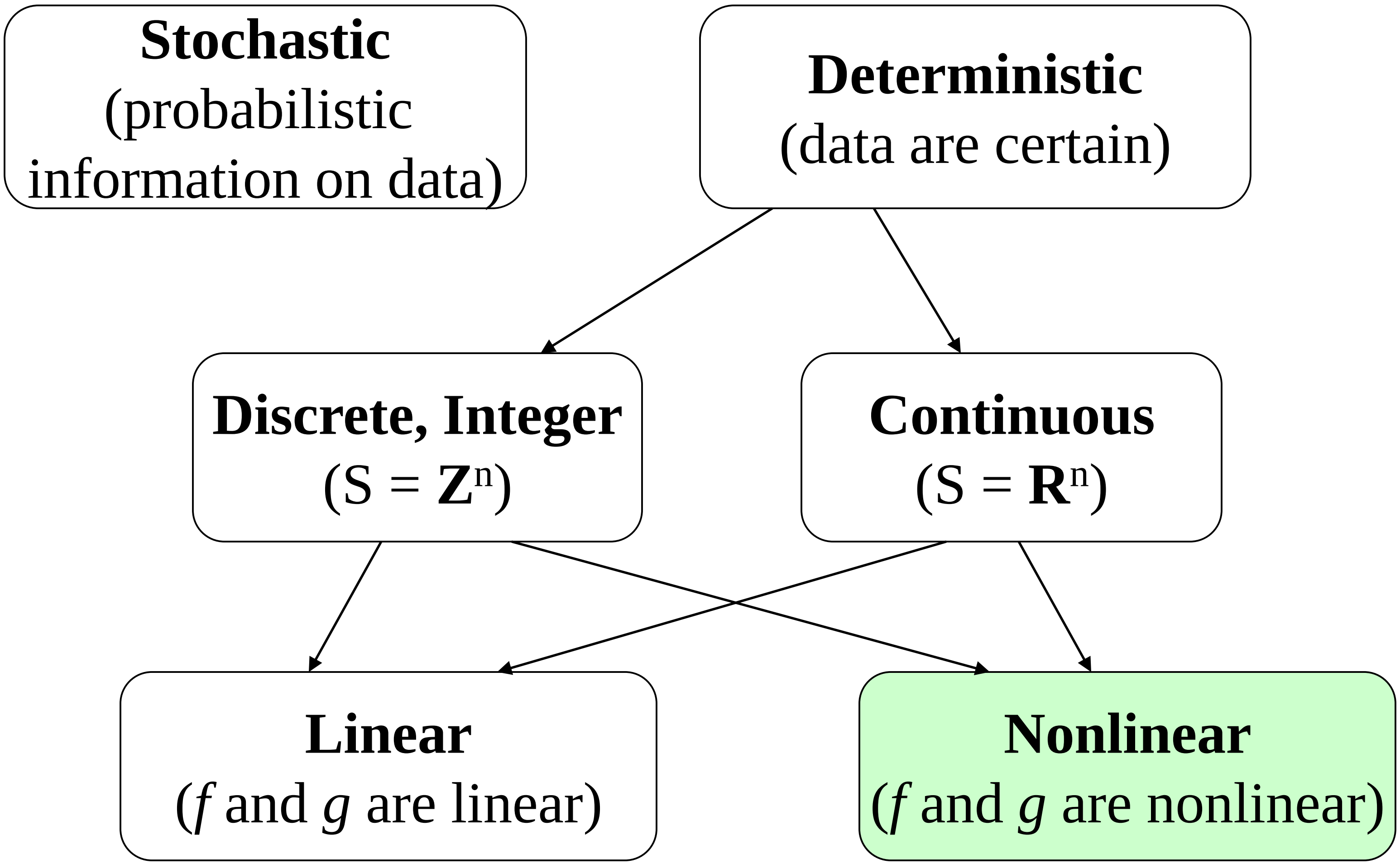
设：NLP $\min\limits_{x\in E^2} f(x)$

$$st.\ g_1(x) \geq 0$$

$$g_2(x) \geq 0$$



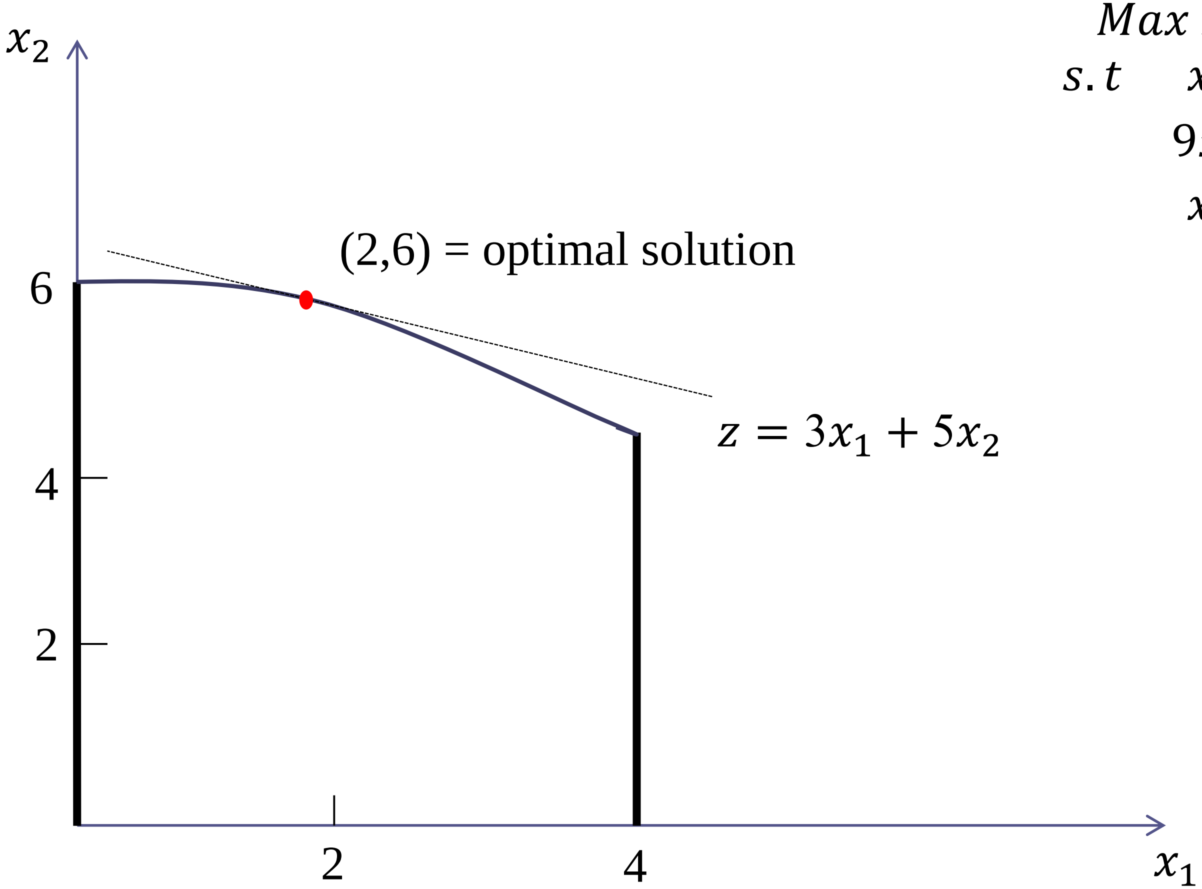
- 最优化类型


Stochastic
(probabilistic information on data)
Deterministic
(data are certain)
Discrete, Integer
($S = \mathbf{Z}^n$)
Continuous
($S = \mathbf{R}^n$)
Linear
(f and g are linear)
Nonlinear
(f and g are nonlinear)

- **非线性规划的特点**
  - NLP的可行域不一定是凸集，甚至是由几个不连通的部分组成。
  - NLP有局部最优解和全局最优解之分，这就要求我们寻优的时候不能停留在局部最优解点，要大区域搜索。
  - NLP的最优解不一定在R的极点上，也不一定在R边界上，它的寻优范围比LP大。
  - NLP的求解中，一般不易求得精确解，大都是近似最优解。

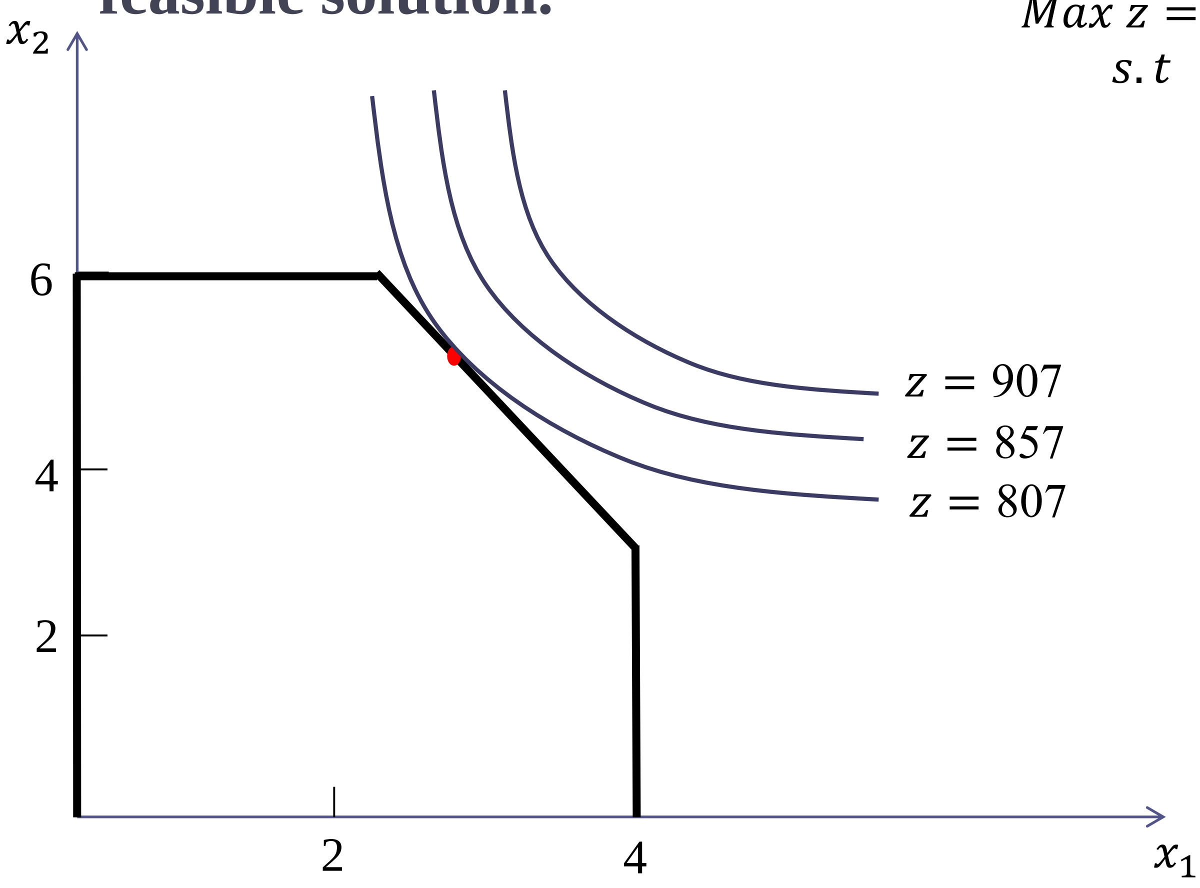
- # Example 1_3
  - **An example with nonlinear constraints when the optimal solution is not a corner point feasible solution.**

$$
\begin{array}{llcccc}
 & Max\ z = 3x_1 + 5x_2 & & & & \\
s.t & x_1 & & & \le & 4 \\
 & 9x_1^2 & + & 5x_2^2 & \le & 216 \\
 & x_1 & , & x_2 & \ge & 0
\end{array}
$$



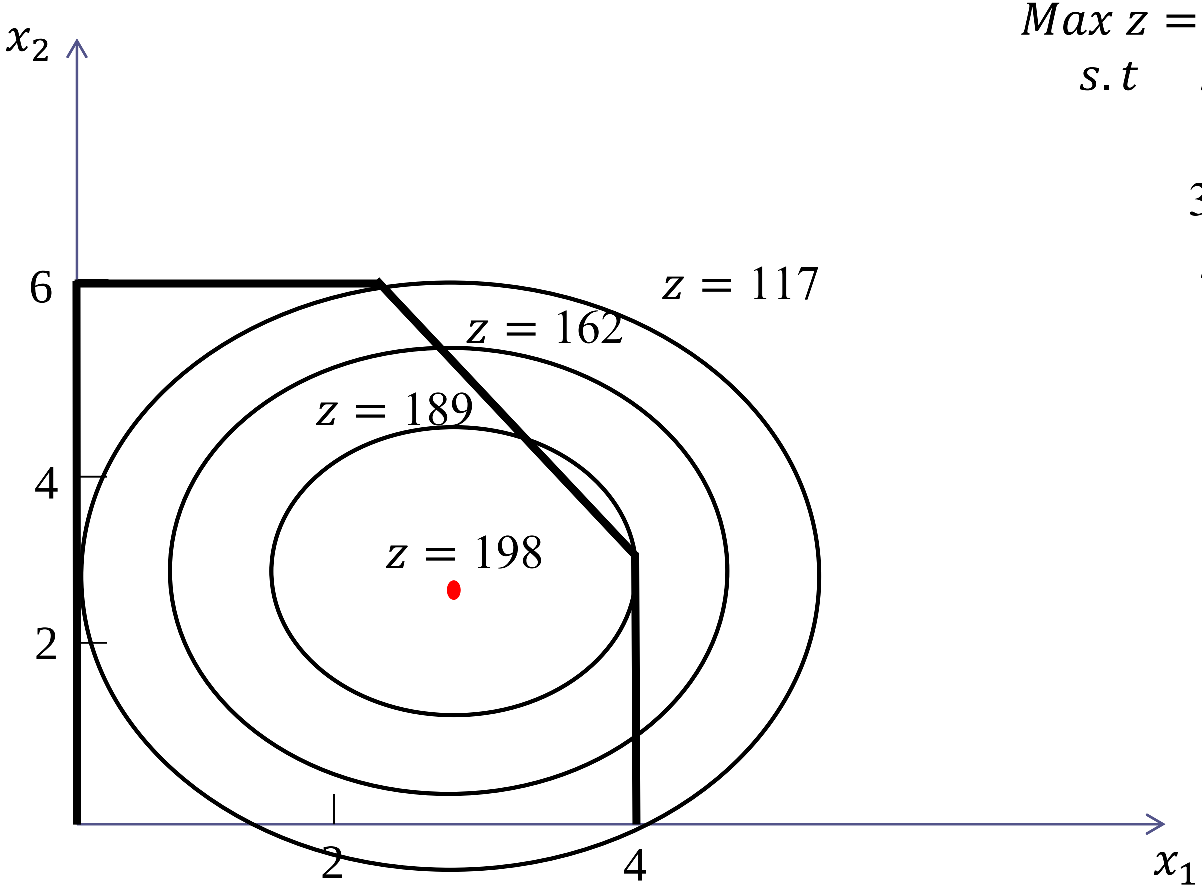
- **Example 1_4**
  - **An example with linear constraints but nonlinear objective function when the optimal solution is not a corner point feasible solution.**

$$
\begin{array}{llllll}
Max\ z = 126x_1 - 9x_1^2 + 182x_2 - 13x_2^2 \\
s.t & x_1 & & & \le & 4 \\
& & + & 2x_2 & \le & 12 \\
& 3x_1 & + & 2x_2 & \le & 18 \\
& x_1 & , & x_2 & \ge & 0
\end{array}
$$



Example 1_5

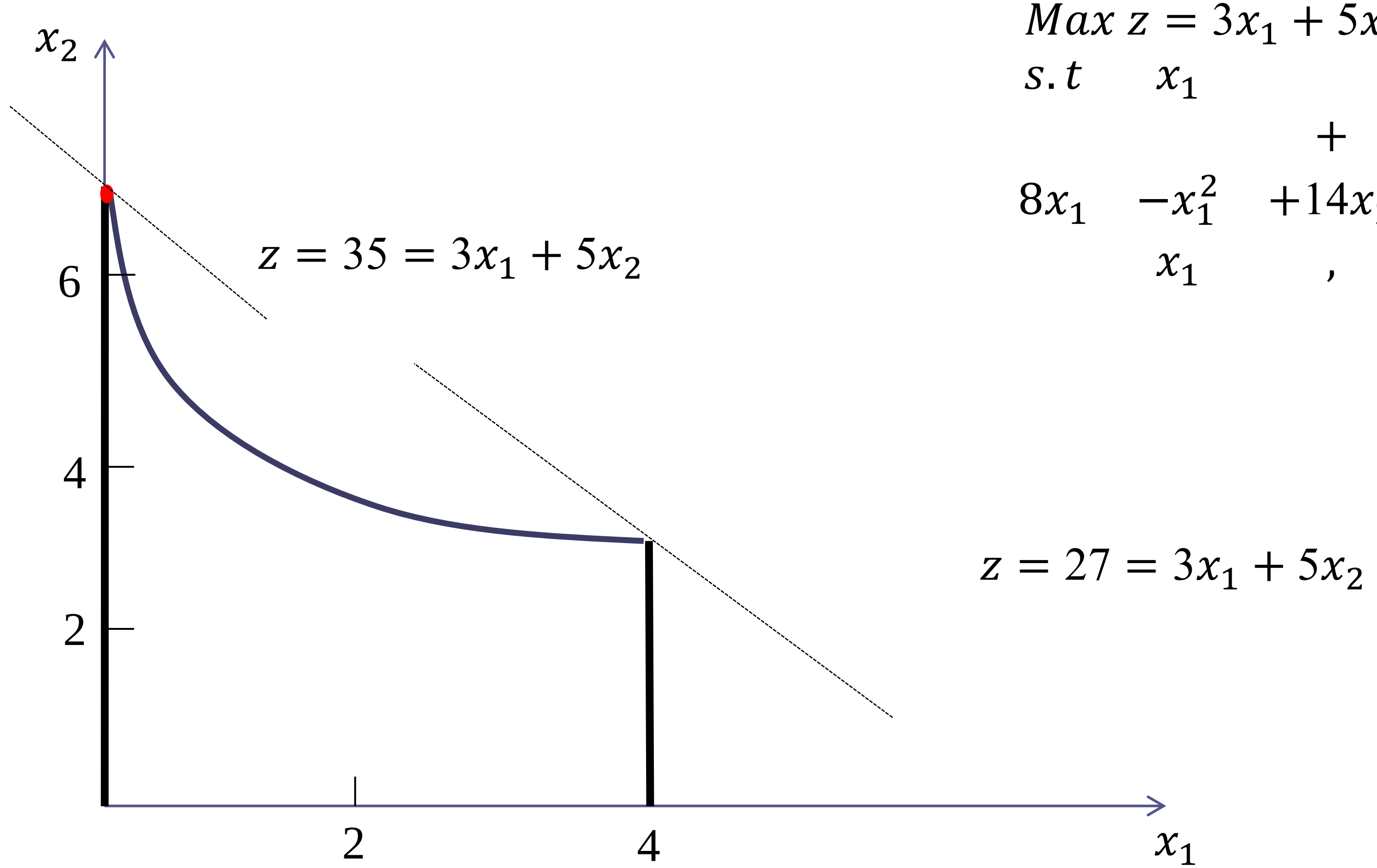
- An example when the optimal solution is inside the boundary of the feasible region.



$$
\begin{array}{lcccccc}
Max\ z = 54x_1 - 9x_1^2 + 78x_2 - 13x_2^2 \\
s.t & x_1 & & & \leq & 4 \\
 & & + & 2x_2 & \leq & 12 \\
 & 3x_1 & + & 2x_2 & \leq & 18 \\
 & x_1 & , & x_2 & \geq & 0
\end{array}
$$

- ## Example 1_6
  - **An example when a local maximum is not a global maximum (the feasible region is not a convex set).**

$$
\begin{array}{llllll}
Max\ z = 3x_1 + 5x_2 & & & & & \\
s.t & x_1 & & & \le & 4 \\
 & & + & 2x_2 & \le & 14 \\
8x_1 & -x_1^2 & +14x_2 & -x_2^2 & \le & 49 \\
 & x_1 & , & x_2 & \ge & 0
\end{array}
$$



基本理论与基础知识

- **NLP的最优解与局部最优解**
  - **1. 最优解**

定义1： 若$x^*\in R$是NLP的最优解点,当且仅当对于任意的$x\in R$均有:

$$f(x)\geq f(\boldsymbol{x}^*) \qquad (4)$$

若上式为严格不等式,则$\boldsymbol{x}^*$为NLP的严格最优解点。

若NLP的最优点不唯一，则记最优解集合为$\boldsymbol{R}^*$。

- **邻域的概念**
  - 定义

  x轴上某点x_0，正数$\varepsilon > 0$，则（$x_0 - \varepsilon, x_0 + \varepsilon$）称为$x_0$点的一个邻域。

  - 定义-扩展到N维

  设点$x_0 \in E_n$，实数$\varepsilon > 0$，则集合$\{x \mid \|x - x_0\| < \varepsilon\}$称为$x_0$点的一个以半径为$\varepsilon$的邻域，记为$N_\varepsilon(x^0)$

几何意义？

- **局部最优解概念**
  - 定义

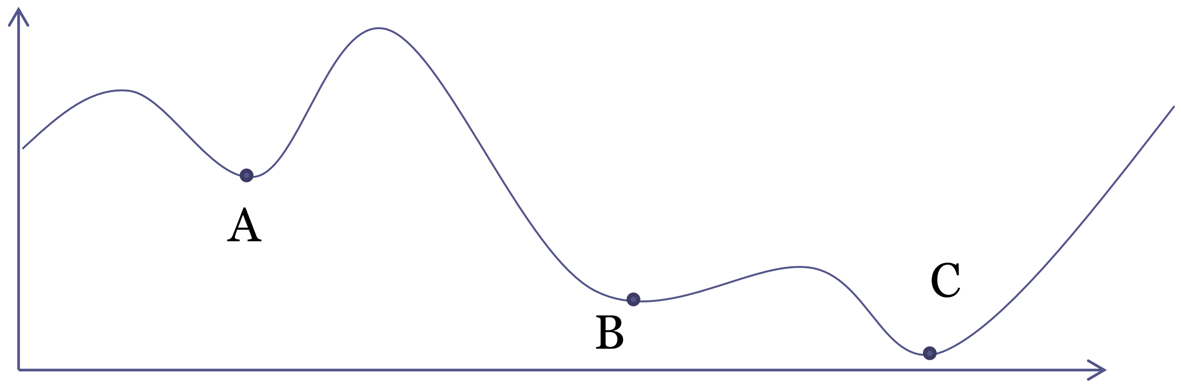
设n维欧氏空间的一点$\boldsymbol{x}^* \in \boldsymbol{R}$的某个邻域$\boldsymbol{N}_{\boldsymbol{\varepsilon}}(\boldsymbol{x}^*) = \{\boldsymbol{x} | \|\boldsymbol{x} - \boldsymbol{x}_0\| < \boldsymbol{\varepsilon}\}$,使得对所有的$\boldsymbol{x} \in \boldsymbol{R} \cap \boldsymbol{N}_{\boldsymbol{\varepsilon}}(\boldsymbol{x}^*)$都有：

$$\boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\boldsymbol{x}^*)$$

则称$\boldsymbol{x}^*$为NLP的局部最优解点。

若上式换成严格不等式$\boldsymbol{f}(\boldsymbol{x}) > \boldsymbol{f}(\boldsymbol{x}^*)$则称$\boldsymbol{x}^*$为NLP的严格局部最优点。局部最优点是否是全局最优点，需要进行比较或鉴别才能知道。

- **局部最优解与全局最优解**



$\boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\boldsymbol{x}^*) \forall \boldsymbol{x}$满足$\|\boldsymbol{x}-\boldsymbol{x}^*\| < \boldsymbol{\varepsilon}$，则$\boldsymbol{x}$为局部最优

$\boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\boldsymbol{x}^*) \forall \boldsymbol{x} \in \boldsymbol{X}$，则$\boldsymbol{x}$为全局最优

以上“$\geq$”换为“$>$”则为严格最优

- 如果目标函数有多个局部最优解，如何取得全局最优?

1、大部分的最优化理论和算法求的是局部最优解。

2、如果目标函数为凸函数，局部最优即为全局最优。如果目标函数为严格凸函数，最优解也唯一。故凸函数与凹函数亦称为单峰函数。

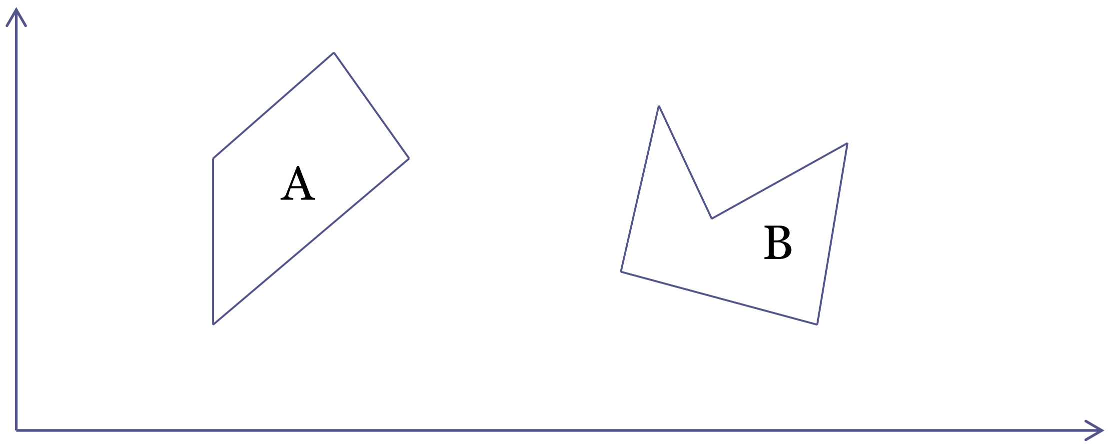
3、使用随机优化算法，如：模拟退火、遗传算法理论上可以找到全局最优解。

- ## 凸集
  - 定义

$\forall\, \boldsymbol{x^1}, \boldsymbol{x^2} \in \boldsymbol{R}, \boldsymbol{x^1}, \boldsymbol{x^2}$为两点连线的点，那么：

$$\boldsymbol{\lambda x^1 + (1-\lambda)x^2 \in R^n}$$

称R为凸集



例：$\boldsymbol{X = \{x \mid \|x\| \leq 5, x \in R^2\}}$

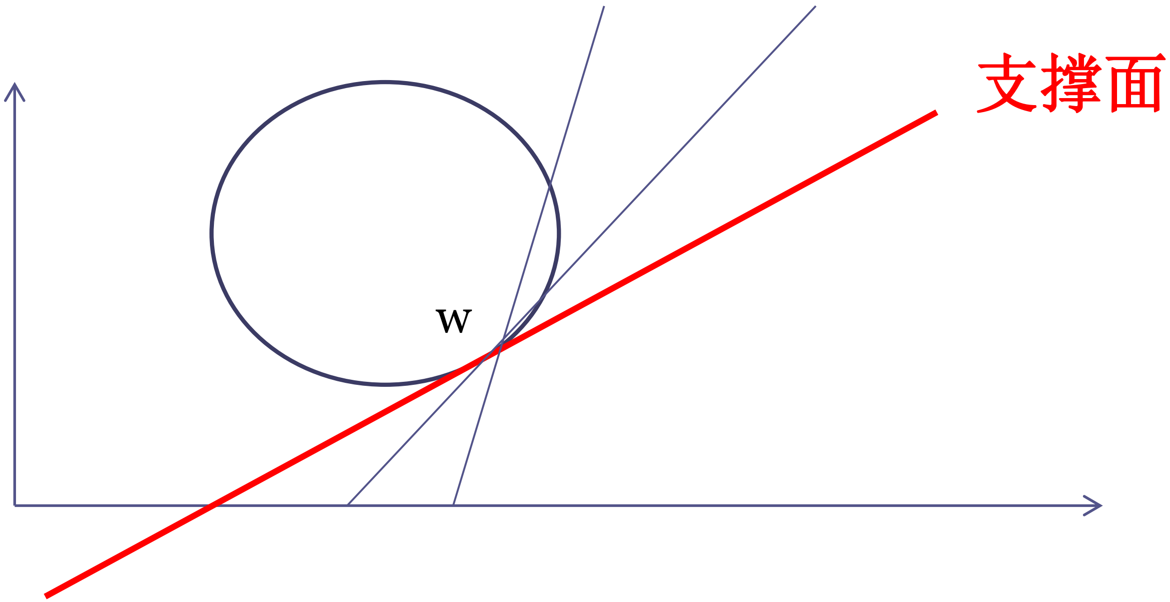
$\boldsymbol{Y = \{y \mid 2 \leq \|y\| \leq 6\,, y \in R^2\}}$

超平面

- 定义

$X = \{x | c^T x = z\} x \neq 0, z$ 为常数，那么：　　称**X**为超平面

- 超平面将空间分成两部分：$c^T x \geq z, c^T x \leq z$



- **凸函数及其性质**
  - 定义
  - 设$\boldsymbol{R} \subseteq \boldsymbol{E}^n$为凸集，若对任意$\boldsymbol{x}^1 \in \boldsymbol{R}, \boldsymbol{x}^2 \in \boldsymbol{R}, \boldsymbol{x}^1 \neq \boldsymbol{x}^2$ 以及数$\boldsymbol{\lambda}, 0 \leq \boldsymbol{\lambda} \leq 1$ 有：

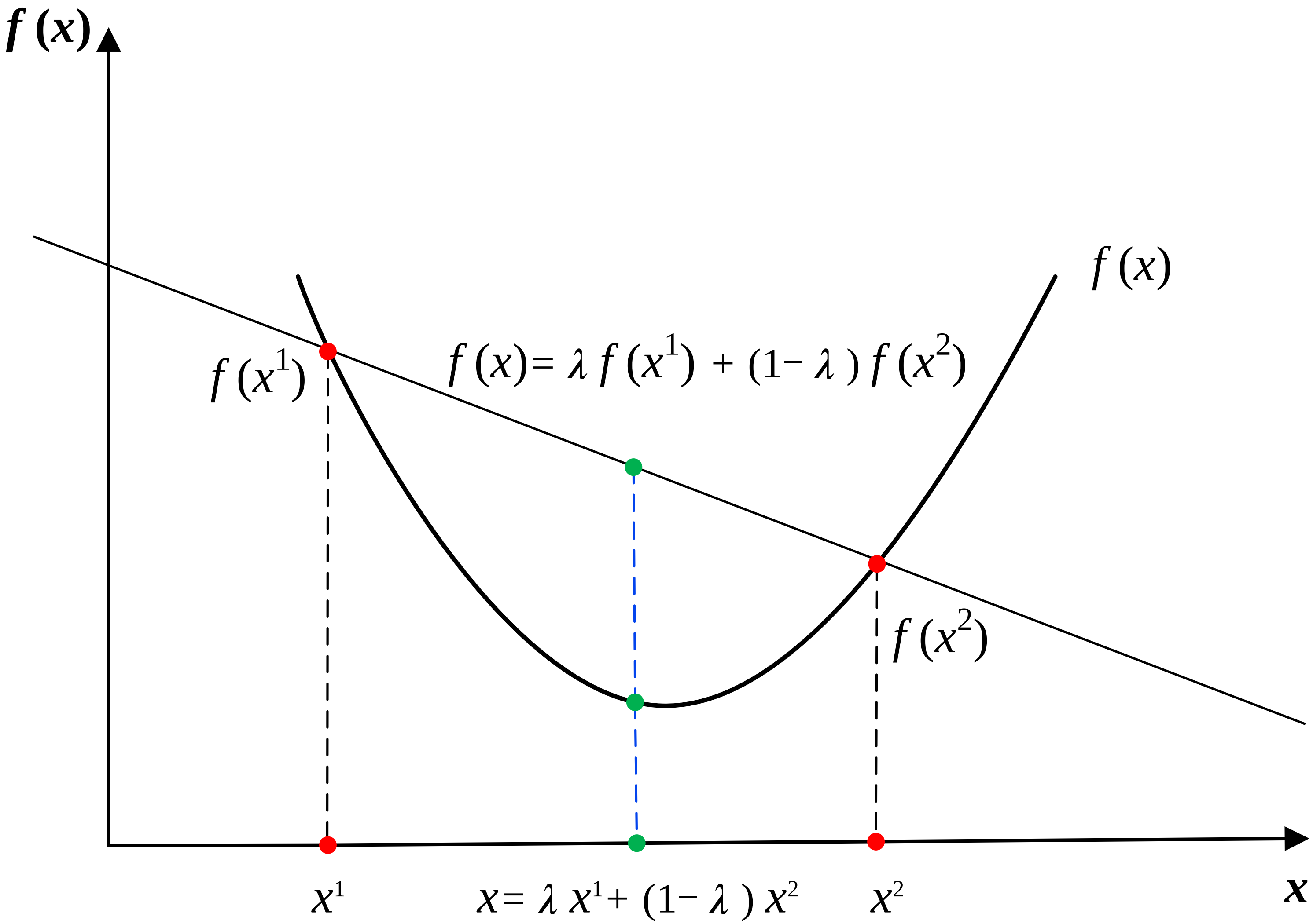
$$\boldsymbol{f}\left(\boldsymbol{x}^1 + \boldsymbol{\lambda}\left(\boldsymbol{x}^2 - \boldsymbol{x}^1\right)\right) \leq \boldsymbol{f}\left(\boldsymbol{x}^1\right) + \boldsymbol{\lambda}[\boldsymbol{f}\left(\boldsymbol{x}^2\right) - \boldsymbol{f}(\boldsymbol{x}^1)]$$

称$\boldsymbol{f}(\boldsymbol{x})$在**R**上为凸函数。上式可写成：

$$\boldsymbol{f}\left(\boldsymbol{\lambda x}^2 + (\mathbf{1} - \boldsymbol{\lambda})\boldsymbol{x}^1\right) \leq \boldsymbol{\lambda f}\left(\boldsymbol{x}^2\right) + (\mathbf{1} - \boldsymbol{\lambda})\boldsymbol{f}(\boldsymbol{x}^1)$$

若换成严格不等式，则$\boldsymbol{f}(\boldsymbol{x})$为严格凸函数。

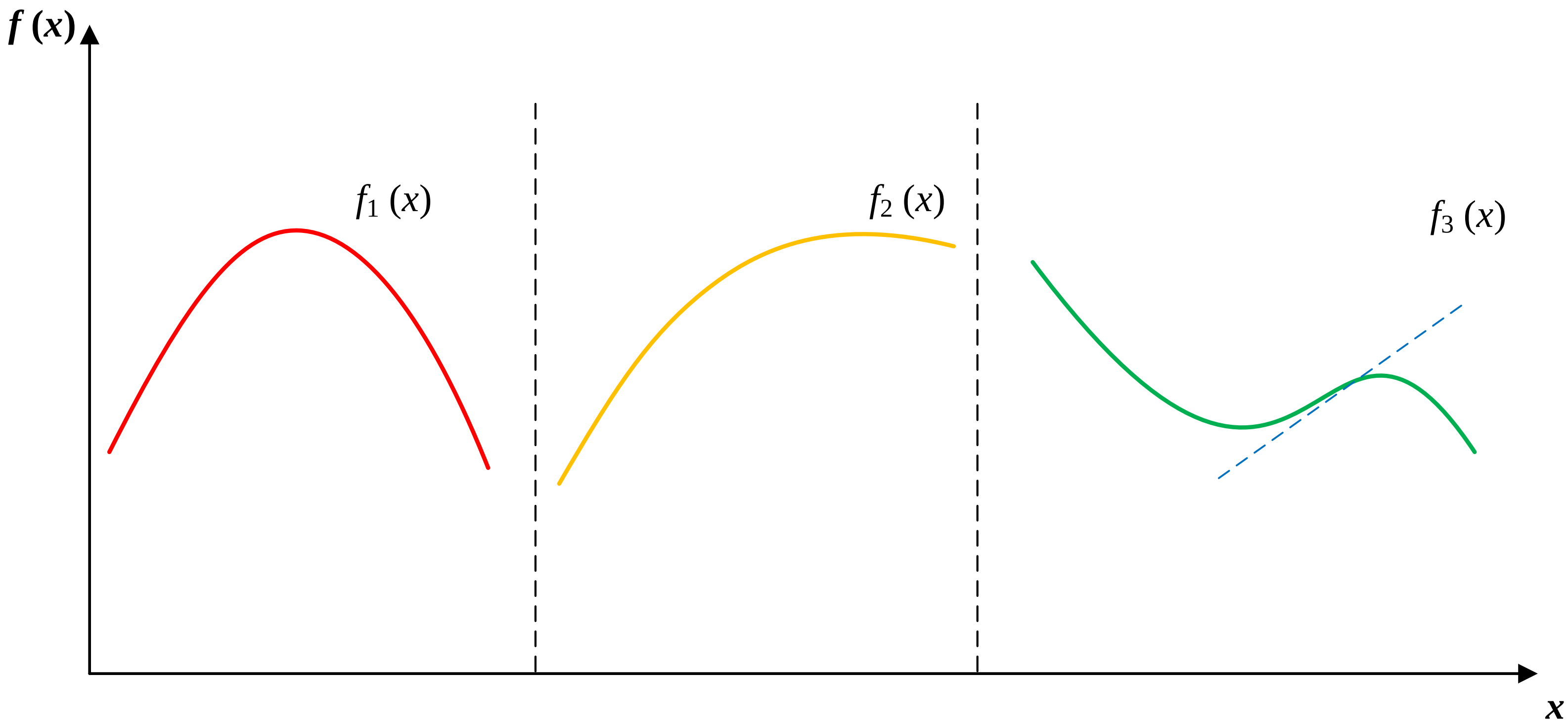
- **凸函数及其性质**



$f(x)$
$f(x)$
$f(x^1)$
$f(x)=\lambda f(x^1)+(1-\lambda)f(x^2)$
$f(x^2)$
x^1
$x=\lambda x^1+(1-\lambda)x^2$
x^2
x

• 凸函数及其性质

- 若将上式的不等式反转，则所定义的函数是**凹函数**和**严格凹函数**。



上面的（1）、（2）图为凹函数，（3）既不是凸函数也不是凹函数（分区域论）。**线性函数**在整个E^n空间中，既为凸函数，又为凹函数，但既非严格凸函数又非严格凹。

- **凸函数的性质**
  - 1.设$f(x)$是凸集R上的一凸（凹）函数，则对于任意$\lambda \geq 0$，函数$\lambda f(x)$在R上是凸（凹）的。由定义即可证得。
  - 2.设$f_1(x), f_2(x)$是凸集R上的凸（凹）函数，则$f(x) = f_1(x) + f_2(x)$在R上也是凸（凹）函数。

证明性质2.

- **凸函数的性质2证明**
  - 证明：设 $f(x)=f_1(x)+f_2(x)$

$\because f_1(x), f_2(x)$是凸的。

$$\therefore f_1\left(x^1+\lambda\left(x^2-x^1\right)\right)\le f_1(x^1)+\lambda[f_1\left(x^2\right)-f_1\left(x^1\right)] \quad \mathbf{(1)}$$

$$f_2\left(x^1+\lambda\left(x^2-x^1\right)\right)\le f_2(x^1)+\lambda[f_2\left(x^2\right)-f_2\left(x^1\right)] \quad \mathbf{(2)}$$

(1)+(2) 得:

$$f\left(x^1+\lambda(x^2-x^1)\right)=f_1\left(x^1+\lambda(x^2-x^1)\right)+\ f_2\left(x^1+\lambda(x^2-\right.$$

- **凸函数的性质**

通过反复运用上两性质可得：

有限个在凸集R上的凸（凹）函数的非负线性组合的函数：

$$f(x)=\lambda_1 \boldsymbol{f}_1(x)+\lambda_2 \boldsymbol{f}_2(x)+\cdots+\lambda_m \boldsymbol{f}_m(x)$$

仍然是凸（凹）函数，其中$\lambda_i \geq 0$，$i=1$，2，…，m。

- **Example**
  - 判别函数$f(x)=\frac{1}{x}+2x^3+3x^4+(x-5)^2+2e^x+chx$在区间$(0,+\infty)$上是否是凸函数。$chx=(\frac{e^x+e^{-x}}{2})$

凸函数的性质

- 性质3
- 设$f(x)$是可微函数，则$f(x)$在开凸集R上为凸函数（或严格凸函数）的充分必要条件是：对任意$x^1, x^2 \in R$有：

$$f(x^2) - f(x^1) \geq \nabla^{\mathrm{T}} f(x^1)(x^2 - x^1) \text{ 或}$$

$$f(x^2) - f(x^1) > \nabla^{\mathrm{T}} f(x^1)(x^2 - x^1)\ x^1 \neq x^2$$

其中，$\nabla^{\mathrm{T}} f(x^1)$表示函数在点$x^1$上的梯度（一阶偏导数）

$$\nabla\ \ f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots \frac{\partial f(x)}{\partial x_n}\right)^T$$

• 方向导数

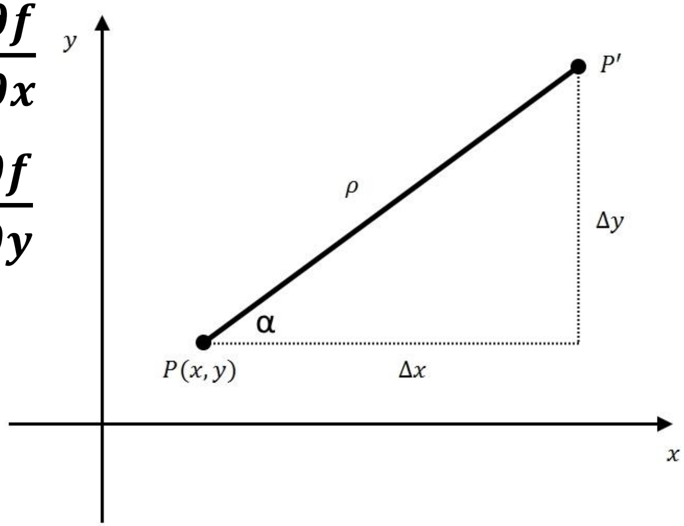
- 设函数$Z=f(x,y)$在p点的某一领域内有定义，称$f(x,y)$在点p沿方向d的变化率为$f(x,y)$在点p沿方向d的方向导数。

$$\frac{\partial f}{\partial d}=\lim_{\rho\to 0}\frac{f(x+\triangle x,y+\triangle y)}{\rho}$$

$\because f(x,y)$在p点沿x方向的方向导数是$\frac{\partial f}{\partial x}$

$f(x,y)$在p点沿y方向的方向导数是$\frac{\partial f}{\partial y}$

$$\therefore \frac{\partial f}{\partial d}=\frac{\partial f}{\partial x}cos\alpha+\frac{\partial f}{\partial y}sin\alpha$$



- **梯度**
  - 梯度是与方向导数关联的一个概念，假设$\boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$在某一领域内具有一阶连续偏导数，则对于平面上的每一点$\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y})$可定义一个向量：

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\boldsymbol{i} + \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\boldsymbol{j}$$

  - 就叫做 $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$在$p$点的梯度。方向导数表示$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})$在点沿$\boldsymbol{d}$方向的变化率，而梯度则表示函数$\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{C}$(等值线)在$p$点的法线方向。
  - NLP中研究的都是n维的情形。

- **充要性证明**
  - 证明：必要性

设$f(x)$是凸函数，任选$\lambda(0 \le \lambda \le 1)$有

$$f\left(x^1 + \lambda\left(x^2 - x^1\right)\right) \le \lambda f\left(x^2\right) + (1-\lambda) f\left(x^1\right) \quad \cdots\cdots (1)$$

整理：$f\left(x^1 + \lambda\left(x^2 - x^1\right)\right) - f\left(x^1\right) \le \lambda\left[f\left(x^2\right) - f\left(x^1\right)\right]$

即：$$\frac{f\left(x^1 + \lambda\left(x^2 - x^1\right)\right) - f\left(x^1\right)}{\lambda} \le f\left(x^2\right) - f\left(x^1\right) \cdots\cdots (2)$$

取极限：$$\lim_{\lambda \to 0} \frac{f\left(x^1 + \lambda\left(x^2 - x^1\right)\right) - f\left(x^1\right)}{\lambda\left(x^2 - x^1\right)}\left(x^2 - x^1\right) \le f\left(x^2\right) - f\left(x^1\right)$$

$$\nabla^{\mathrm{T}} f\left(x^1\right)\left(x^2 - x^1\right) \le f\left(x^2\right) - f\left(x^1\right)$$

必要性得证.

- **充要性证明**
  - 若设$f(x)$在Z上为严格凸函数，**(2)式换成严格不等式后不能两边取极限**. 现假设中间变量: $\boldsymbol{z}=\frac{1}{2}\boldsymbol{x}^1+\frac{1}{2}\boldsymbol{x}^2, \boldsymbol{z}\in\boldsymbol{R}$.

    由于严格凸也是凸函数，则由(1)式得:

$$\boldsymbol{f}(\boldsymbol{z})<\frac{1}{2}\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\frac{1}{2}\boldsymbol{f}(\boldsymbol{x}^2)$$

考虑Z，x^1两点，对凸函数$f(x)$有： $\nabla^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^1)(\boldsymbol{z}-\boldsymbol{x}^1)\leq\boldsymbol{f}(\boldsymbol{z})-\boldsymbol{f}(\boldsymbol{x}^1)$

$$\boldsymbol{f}(\boldsymbol{z})\geq\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\nabla^{\mathrm{T}}\boldsymbol{f}\left(\boldsymbol{x}^1\right)\left(\boldsymbol{z}-\boldsymbol{x}^1\right)$$

$$\therefore\frac{1}{2}\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\frac{1}{2}\boldsymbol{f}\left(\boldsymbol{x}^2\right)>\boldsymbol{f}(\boldsymbol{z})\geq\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\nabla^{\mathrm{T}}\boldsymbol{f}\left(\boldsymbol{x}^1\right)\left(\boldsymbol{z}-\boldsymbol{x}^1\right)$$

$$=\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\frac{1}{2}\nabla^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^1)(\boldsymbol{x}^2-\boldsymbol{x}^1)$$

$$\therefore\boldsymbol{f}\left(\boldsymbol{x}^2\right)>\boldsymbol{f}\left(\boldsymbol{x}^1\right)+\nabla^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^1)(\boldsymbol{x}^2-\boldsymbol{x}^1)$$

- **性质3充分性证明**

设$\boldsymbol{x}^1, \boldsymbol{x}^2 \in \boldsymbol{R}, 0 \le \lambda \le 1$,令$\boldsymbol{z} = \lambda \boldsymbol{x}^2 + (1-\lambda)\boldsymbol{x}^1$, 则$\boldsymbol{z} \in \boldsymbol{R}$

$$\because f(\boldsymbol{x}^2) \ge f(\boldsymbol{x}^1) + \nabla^{\mathrm{T}} f(\boldsymbol{x}^1)(\boldsymbol{x}^2 - \boldsymbol{x}^1)$$

则有：$f(\boldsymbol{x}^2) \ge f(\boldsymbol{z}) + \nabla^{\mathrm{T}} f(\boldsymbol{z})(\boldsymbol{x}^2 - \boldsymbol{z})$

$$f(\boldsymbol{x}^1) \ge f(\boldsymbol{z}) + \nabla^{\mathrm{T}} f(\boldsymbol{z})(\boldsymbol{x}^1 - \boldsymbol{z})$$

将上两式分别乘以$\lambda, (1-\lambda)$相加得：

$$\lambda f(\boldsymbol{x}^2) + (1-\lambda) f(\boldsymbol{x}^1)$$

$$\ge \lambda f(\boldsymbol{z}) + (1-\lambda) f(\boldsymbol{z}) + \nabla^{\mathrm{T}} f(\boldsymbol{z}) \left[\lambda(\boldsymbol{x}^2 - \boldsymbol{z}) + \left((1-\lambda)(\boldsymbol{x}^1 - \boldsymbol{z})\right)\right]$$

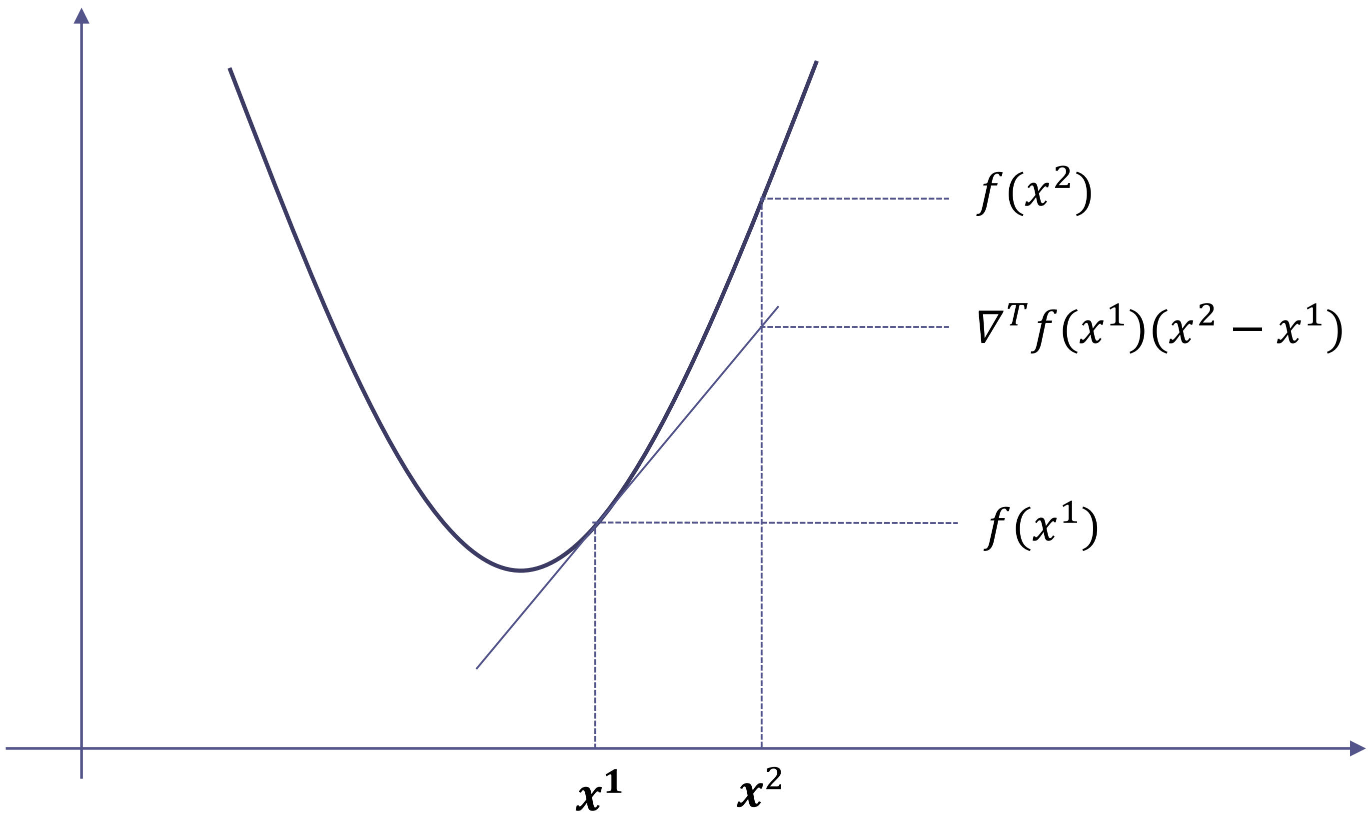
$$= f(\boldsymbol{z}) + \nabla^{\mathrm{T}} f(\boldsymbol{z})\left[\lambda \boldsymbol{x}^2 - \lambda \boldsymbol{z} + (1-\lambda)\boldsymbol{x}^1 - \boldsymbol{z} + \lambda \boldsymbol{z}\right]$$

$$= f(\boldsymbol{z}) = f\left[\lambda \boldsymbol{x}^2 + (1-\lambda)\boldsymbol{x}^1\right]$$

即：$f\left[\lambda \boldsymbol{x}^2 + (1-\lambda)\boldsymbol{x}^1\right] \le \lambda f(\boldsymbol{x}^2) + (1-\lambda) f(\boldsymbol{x}^1)$

故 $f(x)$为凸函数，严格凸函数情况类此。证毕。

- **几何意义**
  - x^2与x^1点的函数值差总大于等于从x^1点做切线的纵坐标值。



- **凸函数的性质**
  - 性质4
  - 设$f(x)$是二阶可微的，则$f(x)$在开凸集R上为凸函数的充分必要条件是：对一切$x \in R$，$Hesse$矩阵$H(x)$为半正定的。若$H(x)$为正定的，则$f(x)$为严格凸函数。
  - 性质4充要性证明

- **凸函数的性质**
  - *Hesse*矩阵： $f(x)$的$Hesse$矩阵是函数$f(x)$对各x分量求二阶偏导组成的矩阵。

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

• 凸函数的性质

- 其二是矩阵的正定与半正定，由线性代数学可知，矩阵**H**正定的充要条件是ZHZ二次型正的，半正定是ZHZ $<= 0$，负定是$ZHZ < 0$；二次型ZHZ为正定的充要条件是它的各阶子式都大于零，半正定是各阶主子式均大于等于零，负定是各阶主子式正负相间。

- 即：正定 $h_{11} > 0, \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} > 0, \cdots\cdots, \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} > 0$
- 半正定：将上面的>号改≥号

- 负定：$h_{11} < 0, \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} > 0, \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} < 0, \cdots, (-1)^n \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} > 0$

- **泰勒定理**
  - 若函数$f(x)$在点x_0的某一领域内具有直到$(n+1)$阶的导数，则有：

$$
\begin{aligned}
&\boldsymbol{f(x)} \\
&= \boldsymbol{f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^n(x_0)}{n!}(x - x_0)^n + R_n(x)}
\end{aligned}
$$

$\boldsymbol{R_n(x)}$: 拉格朗日余项，高阶无穷小

- **性质4必要性证明**
  - 设$f(x)$为R上的凸函数，任取$x \in R, z \in E^n$
  - $\because R$为凸开集，故存在领域σ, 当$\lambda \in [-\sigma, +\sigma]$时

$$x + \lambda z \in R$$

由性质3可知：$f(x+\lambda z) \geq f(x) + \nabla^{\mathrm{T}} f(x)(x + \lambda z - x)$

$$= f(x) + \lambda \nabla^{\mathrm{T}} f(x) z$$

$$\therefore f(x+\lambda z) - f(x) \geq \lambda \nabla^{\mathrm{T}} f(x) z$$

  - 又$\because f(x)$在x处是二阶可微的，由Taylor 定理将$f(x+\lambda z)$在x点处展开：

$$f(x+\lambda z) = f(x) + \lambda \nabla^{\mathrm{T}} f(x) z + \frac{\lambda^2}{2} z^T H z + o(\lambda^2)$$

$$f(x+\lambda z) - f(x) - \lambda \nabla^{\mathrm{T}} f(x) z = \frac{\lambda^2}{2} z^T H z + o(\lambda^2)$$

- **性质4证明**

由（2）式可知：$\frac{\lambda^2}{2} z^T H z + o(\lambda^2) \geq 0$

$$\therefore z^T H z \geq 0$$

故矩阵$H(x)$半正定

- **性质4证明**
  - 充分性
  - 设$H(x)$在R上为半正定，任取$\boldsymbol{x}^1, \boldsymbol{x}^2 \in \boldsymbol{R}, \boldsymbol{x}^1 \neq \boldsymbol{x}^2$

则有 $\boldsymbol{\alpha} \in (0,1), \boldsymbol{x} = \boldsymbol{\alpha}\boldsymbol{x}^1 + (1-\boldsymbol{\alpha})\boldsymbol{x}^2$

由$Taylor$式知：

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^1) + \nabla^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^1)(\boldsymbol{x}-\boldsymbol{x}^1) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^1)^T\boldsymbol{H}(\boldsymbol{x}^1)(\boldsymbol{x}-\boldsymbol{x}^1) + \boldsymbol{o}(\boldsymbol{\alpha})$$

$\because \boldsymbol{H}(\boldsymbol{x}^1)$是半正定的，则有：

$$\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^1)^T\boldsymbol{H}(\boldsymbol{x}^1)(\boldsymbol{x}-\boldsymbol{x}^1) \geq \boldsymbol{0}$$

$$\therefore \boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\boldsymbol{x}^1) + \nabla^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^1)(\boldsymbol{x}-\boldsymbol{x}^1)$$

由性质3知 $f(x)$在R上为凸函数。

- **例**

判断函数$f(x_1,x_2)=2x_1^2+x_2^2-2x_1x_2+x_1+1,x\in R^2$的凸性。

3 凸规划

设NLP为：$(p) \quad min_{x \in R} f(x)$

$$R = \{x | g_i(x) \geq 0, i = 1, \dots, m\}$$

其中：$f(x), g_i(x), i = 1, \dots, m$为凸函数，则$p$为凸规划。

- 定理：若某问题为凸规划，则：

  1）p的可行解集合R为凸/凹集；

  2）p的最优解集合R^*为凸/凹集；

  3）p的任何局部最优解都是全局最优解；

  4）若$f(x)$为严格凸函数，且p的最优解集合$R^* \neq \varphi$，则最优解必唯一。

- **证明定理1**
  - **证: p的可行解集合R为凸集.**

不妨设$R \neq \varphi$，因为空集也是凸集,

则对于任意的两点：$x^1, x^2 \in R$,及$\lambda \in (0,1)$有:

$$\boldsymbol{g_i\left(x^1\right) \geq 0, g_i\left(x^2\right) \geq 0}$$

$\because \boldsymbol{g_i(x)}$是凹函数

$$\therefore \boldsymbol{g_i\left(\lambda x^1 + (1-\lambda)x^2\right) \geq \lambda g_i\left(x^1\right) + (1-\lambda) g_i\left(x^2\right) \geq 0}$$

$$\therefore \boldsymbol{g_i\left(\lambda x^1 + (1-\lambda)x^2\right) \geq 0 \quad \forall i = 1,2,\cdots\cdots m}$$

即：$\boldsymbol{\lambda x^1 + (1-\lambda)x^2 \in R}$　　$\boldsymbol{R}$ 为凸集

- **证明定理2**
  - 证：p的最优解集合R^*为凸集

$\boldsymbol{R}^* \neq \boldsymbol{\phi}$,取$\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}} \in \boldsymbol{R}^*$

由(1)知，对于任意的$\lambda \in (0,1)$有：$\lambda\overline{\boldsymbol{x}} + (1-\lambda)\overline{\boldsymbol{y}} \in \boldsymbol{R}$

又$\because \boldsymbol{f}(\overline{\boldsymbol{x}}), \boldsymbol{f}(\overline{\boldsymbol{y}})$为最优值，则：$\boldsymbol{f}(\overline{\boldsymbol{x}}) = \boldsymbol{f}(\overline{\boldsymbol{y}}) = \min\limits_{x \in R} \boldsymbol{f}(\boldsymbol{x})$

$\therefore \boldsymbol{f}(\lambda\overline{\boldsymbol{x}} + (1-\lambda)\overline{\boldsymbol{y}}) \leq \lambda\boldsymbol{f}(\overline{\boldsymbol{x}}) + (1-\lambda)\boldsymbol{f}(\overline{\boldsymbol{y}}) = \min\limits_{x \in R} \boldsymbol{f}(\boldsymbol{x})$

上式只能取等号，否则与假设矛盾，即：

$\boldsymbol{f}(\lambda\overline{\boldsymbol{x}} + (1-\lambda)\overline{\boldsymbol{y}}) = \min\limits_{x \in R} \boldsymbol{f}(\boldsymbol{x})$

$\lambda\overline{\boldsymbol{x}} + (1-\lambda)\overline{\boldsymbol{y}} \in \boldsymbol{R}^*$

- **证明定理3**
  - 证 P的任何局部最优解为全局最优解。
  - 反证法：设$\bar{x}$为P的局部最优解，那么存在$\bar{x}$的某领域$N_\varepsilon(\bar{x})$
  对于任意$x \in N_\varepsilon(\bar{x}) \cap R$,都有$f(x) \geq f(\bar{x})$
  假设$\bar{x}$不是P的全局最优解，则必存在$\bar{y} \in R$有$f(\bar{y}) < f(\bar{x})$
  $\because f(x)$是凸函数，故对$\lambda \in (0,1)$有：
  $f(\lambda\bar{x} + (1-\lambda)\bar{y}) \leq \lambda f(\bar{x}) + (1-\lambda)f(\bar{y})$
  $< \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x})$
  即：$f(\lambda\bar{x} + (1-\lambda)\bar{y}) < f(\bar{x}) \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (1)$
  但，当λ足够接近1时，$\lambda\bar{x} + (1-\lambda)\bar{y} \in N_\varepsilon(\bar{x}) \cap R$
  又由局部最优解的定义知：
  $$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \geq f(\bar{x}) \cdots\cdots\cdots\cdots\cdots\cdots (2)$$

故与（1）式矛盾，所以$\bar{x}$是P的全局最优解。

- **证明定理4**
  - 证 (4)若$f(x)$为严格凸函数，且$R^* \neq \varphi$，则最优解唯一。
  - 反证法： 若最优解不唯一，即存在$\overline{\boldsymbol{x}} \in \boldsymbol{R}^*, \overline{\boldsymbol{y}} \in \boldsymbol{R}^*, \overline{\boldsymbol{x}} \neq \overline{\boldsymbol{y}}$,

取$\boldsymbol{\lambda} \in (0,1)$有：$\boldsymbol{\lambda}\overline{\boldsymbol{x}} + (1-\boldsymbol{\lambda})\overline{\boldsymbol{y}} \in \boldsymbol{R}^*$

$\therefore \boldsymbol{f}(\boldsymbol{\lambda}\overline{\boldsymbol{x}} + (1-\boldsymbol{\lambda})\overline{\boldsymbol{y}}) = \boldsymbol{f}(\overline{\boldsymbol{x}}) = \boldsymbol{f}(\overline{\boldsymbol{y}}) = \boldsymbol{minf}(\boldsymbol{x}) \cdots\cdots (3)$

又$\because \boldsymbol{f}(\boldsymbol{x})$严格凸，

$$\therefore \boldsymbol{f}(\boldsymbol{\lambda}\overline{\boldsymbol{x}} + (1-\boldsymbol{\lambda})\overline{\boldsymbol{y}}) < \boldsymbol{\lambda}\boldsymbol{f}(\overline{\boldsymbol{x}}) + (1-\boldsymbol{\lambda})\boldsymbol{f}(\overline{\boldsymbol{y}}) = \boldsymbol{f}(\overline{\boldsymbol{x}})$$

此与 $\boldsymbol{\lambda}\overline{\boldsymbol{x}} + (1-\boldsymbol{\lambda})\overline{\boldsymbol{y}} \in \boldsymbol{R}^*$ 相矛盾，或与（3）式相矛盾，

故得证。

- **定理5**
  - 证 (5)设P为凸规划，$f(x)$可微，则$\boldsymbol{x}^*$是P的最优解的充分必要条件是对任意 $\boldsymbol{x} \in \boldsymbol{R}$ 有：$\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \geq \mathbf{0}$
  - 证明

- **定理5**
  - 证 (5)设P为凸规划，$f(x)$可微，则x^*是P的最优解的充分必要条件是对任意 $\boldsymbol{x} \in \boldsymbol{R}$ 有：$\nabla^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0$
  - 证明：充分性

由凸函数性质3知，对于任意$\boldsymbol{x} \in \boldsymbol{R}$有：

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

$$\because \nabla^{\mathrm{T}} \boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \geq 0$$

$\therefore \boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\boldsymbol{x}^*)$ 故$\boldsymbol{x}^*$为$\boldsymbol{P}$的最优解。

- **定理5**
  - 证 (5)设P为凸规划，$f(x)$可微，则x^*是P的最优解的充分必要条件是对任意 $\boldsymbol{x} \in \boldsymbol{R}$ 有：$\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*) \geq \mathbf{0}$
  - 证明：必要性
  - 用反证法

若$\boldsymbol{x}^*$是P的最优解，但存在$x^0 \in R$，使：$\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{f}(\boldsymbol{x}^*)\left(\boldsymbol{x}^0-\boldsymbol{x}^*\right) < \mathbf{0}$

则由：$\boldsymbol{f}\left(\boldsymbol{x}^*+\boldsymbol{\lambda}\left(\boldsymbol{x}^0-\boldsymbol{x}^*\right)\right)=\boldsymbol{f}(\boldsymbol{x}^*)+\boldsymbol{\lambda}\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{f}\left(\boldsymbol{x}^0-\boldsymbol{x}^*\right)+\boldsymbol{o}(\boldsymbol{\lambda})$

当取充分小的λ时，总可以使$\boldsymbol{\lambda}\boldsymbol{\nabla}^{\mathrm{T}}\boldsymbol{f}\left(\boldsymbol{x}^0-\boldsymbol{x}^*\right)+\boldsymbol{o}(\boldsymbol{\lambda}) < \mathbf{0}$

故有：$\boldsymbol{f}\left(\boldsymbol{x}^*+\boldsymbol{\lambda}\left(\boldsymbol{x}^0-\boldsymbol{x}^*\right)\right) \leq \boldsymbol{f}(\boldsymbol{x}^*)$

与假设$\boldsymbol{x}^*$是最优解矛盾，故得证。