3 数学知识-2

3.1 本集内容简介

最优化中的基本概念
梯度下降法
牛顿法
坐标下降法
数值优化算法面临的问题
拉格朗日乘数法
凸优化问题
- 凸集
- 凸函数
- 凸优化

拉格朗日对偶
KKT条件

3.2 最优化中的基本概念

最优化问题
目标函数
优化变量
可行域
等式约束
不等式约束
局部极小值
全局极小值

最优化问题，通常说的基本上是求极小值问题。

3.3 为什么要用迭代法

基本思路

为什么要用迭代法？

$$f(x,y)=x^3-2x^2+e^{xy}-y^3+10y^2+100\sin(xy)$$

$$\begin{cases}3x^2-4x+ye^{xy}+100y\cos(xy)=0\\ xe^{xy}-3y^2+20y+x\cos(xy)\quad=0\end{cases}$$

$$\lim_{k\to+\infty}\nabla f(\mathrm{x}_k)=0$$

$$\mathrm{x}_{k+1}=h(\mathrm{x}_k)$$

3.4 梯度下降法

数值优化算法：求近似解。

$$\mathrm{x}_{k+1}=\mathrm{x}_k-\gamma\nabla f(\mathrm{x}_k)$$

3.5 牛顿法

$$\mathrm{x}_{k+1}=\mathrm{x}_k-\mathrm{H}_k^{-1}\mathrm{g}_k$$

3.6 坐标下降法

分治法思想

$$\min f(\mathrm{x}),\mathrm{x}=(x_1,x_2,\ldots,x_n)$$

$$\min_{x_i} f(\mathrm{x})$$

3.7 优化算法面临的问题

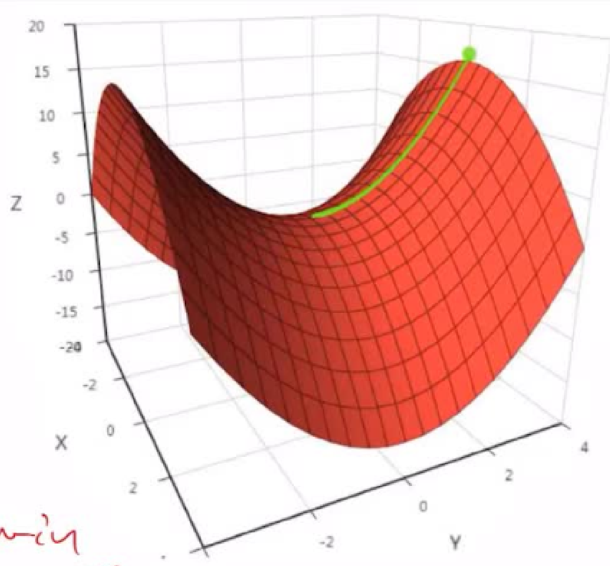
局部极值问题
鞍点问题

$\nabla f(x_0)=0$

必要

充分

$\nabla f(x_0)=0 \nRightarrow x_0 \min$



3.8 拉格朗日乘数法

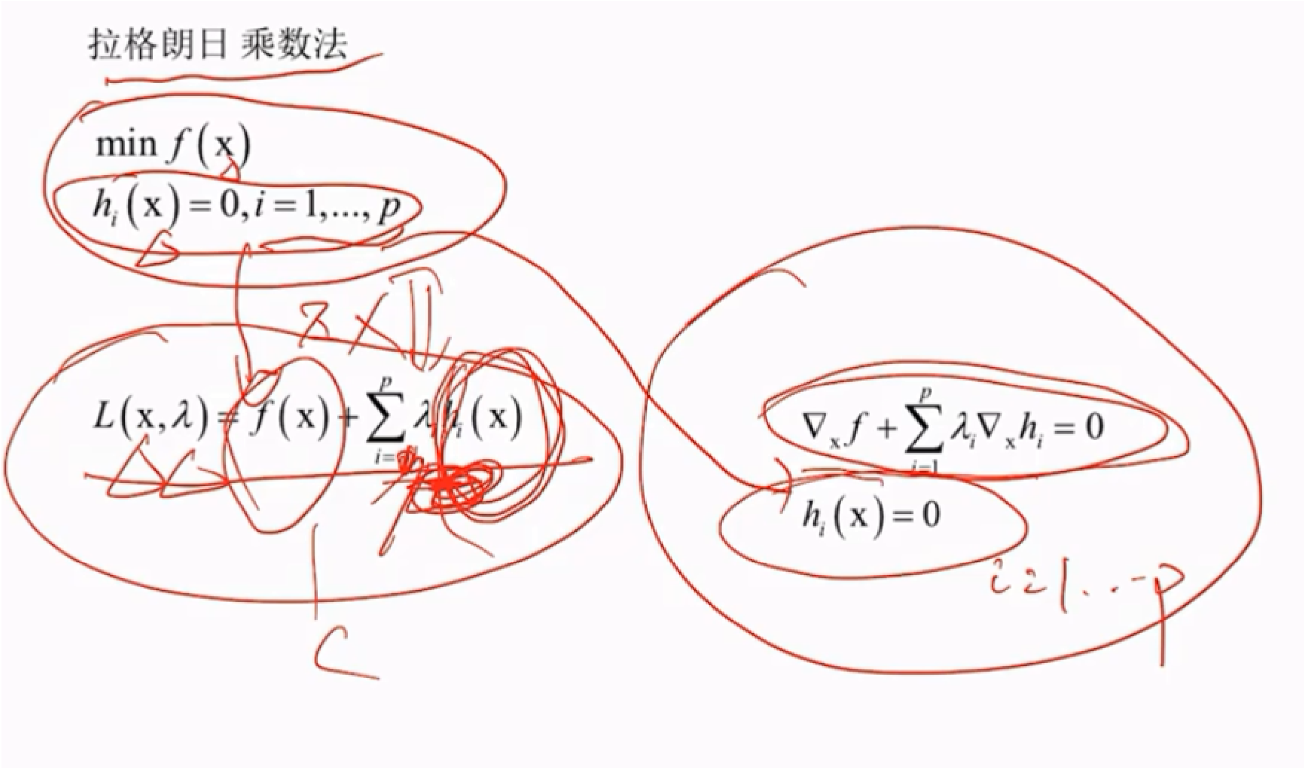
拉格朗日 乘数法

$$\min f(\mathrm{x})$$
$$h_i(\mathrm{x})=0, i=1,\dots,p$$

$$L(\mathrm{x},\lambda)=f(\mathrm{x})+\sum_{i=}^{p}\lambda_i h_i(\mathrm{x})$$

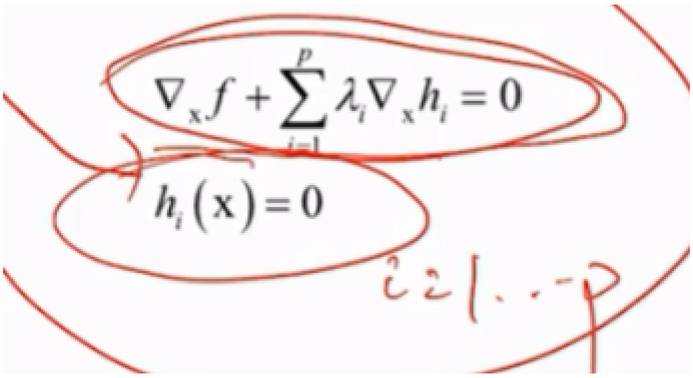
$$\nabla_{\mathrm{x}} f+\sum_{i=1}^{p}\lambda_i \nabla_{\mathrm{x}} h_i=0$$
$$h_i(\mathrm{x})=0$$

这里



优化问题，lambda称为乘子变量

- 带约束条件的等价于不带约束条件的。
- 对x求梯度。
- 对lambda求梯度，前面一项没有lambda,所以是个常数项。
- 最后求解方程组，得到极大值，极小值。

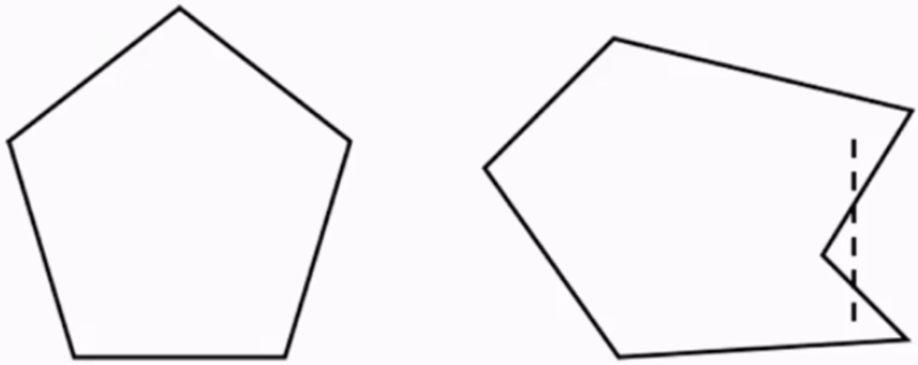


3.9 凸优化简介

- 凸优化
    - 凸集
    - 凸函数

凸优化
凸集
凸函数

$$\theta x+(1-\theta)y \in C$$



3.10 凸集

限定

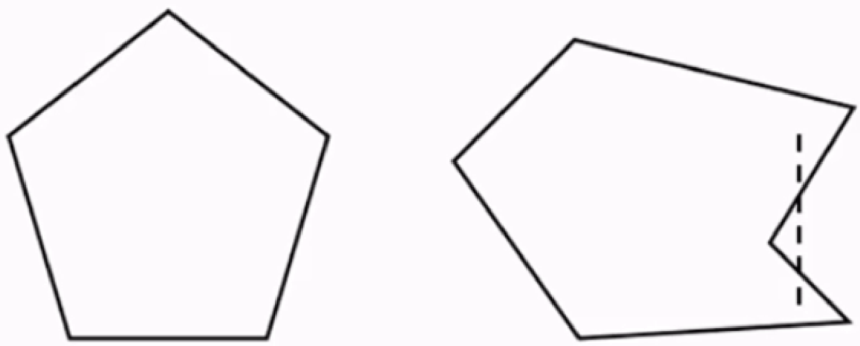
凸优化

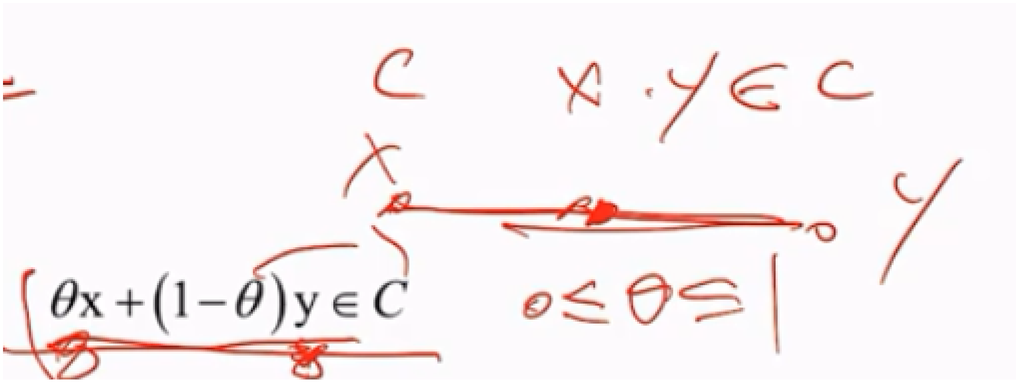
1 凸集

2 凸函数

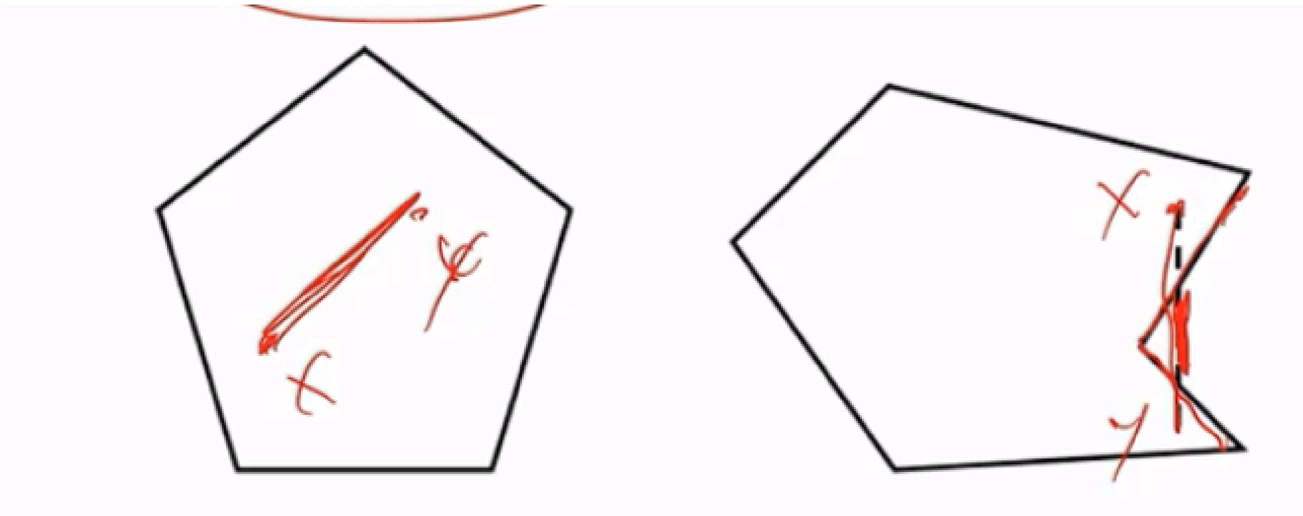
f(x)

$$\theta x+(1-\theta)y \in C$$

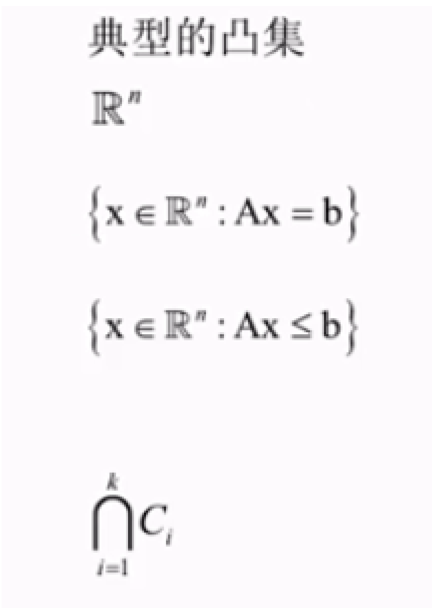




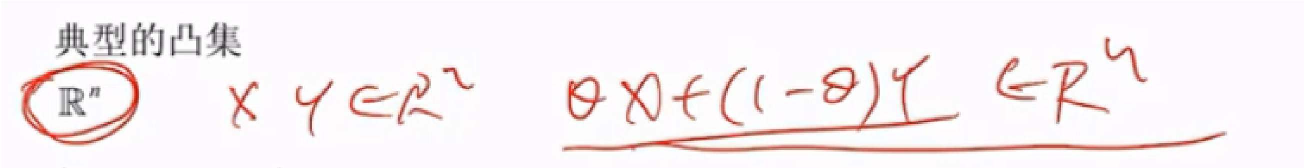
凸集的直观理解



那么我们怎么定义凸集呢？

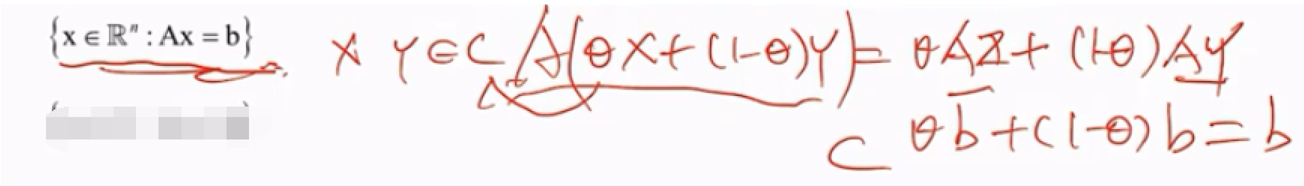


凸集的性质

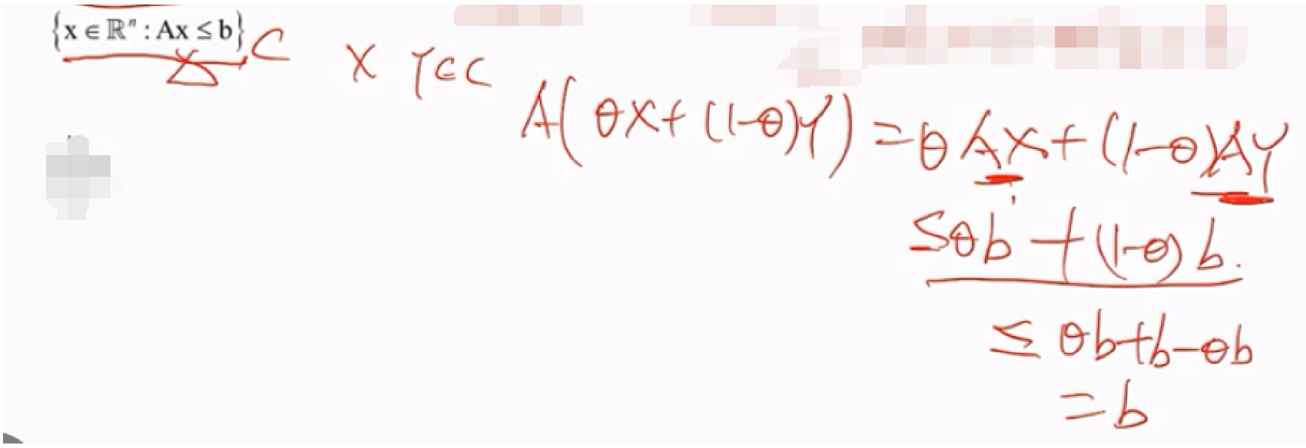


根据这个性质推导下面的公式

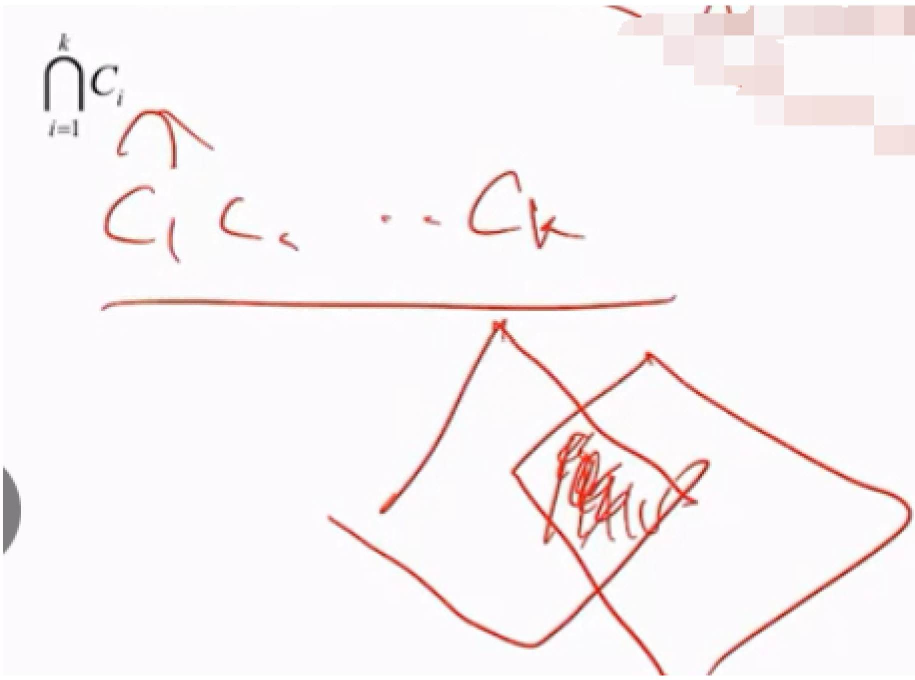
- C 代表集合



不等式约束的凸集



如果k个集合是凸集，那么他们的交集也是凸集。



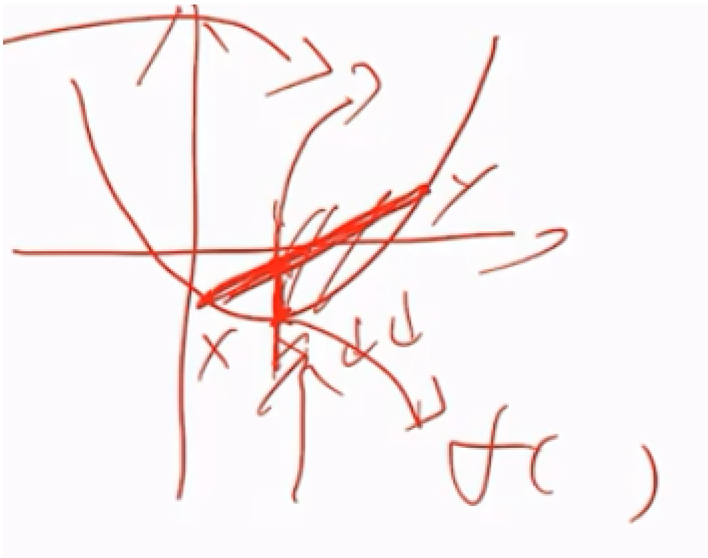
3.11 凸函数

凸函数的定义

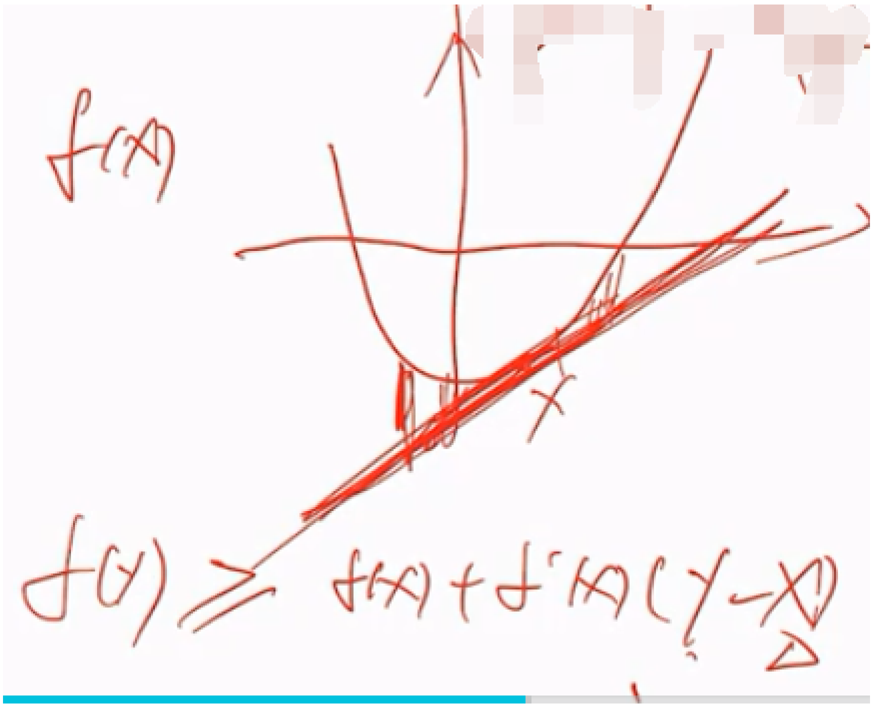
$$f\left(\theta \mathrm{x}+(1-\theta)\mathrm{y}\right)<\theta f\left(\mathrm{x}\right)+(1-\theta)f\left(\mathrm{y}\right)$$

一阶判别法
二阶判别法

$$f\left(\mathrm{x}\right)=\sum_{i=1}^{k}w_i f_i\left(\mathrm{x}\right)$$



- 一阶判别法



- 二阶判别法

  二阶导数大于等于零

  H矩阵半正定

3.12 凸优化的性质

给出了我们证明凸优化的思路。

局部最优解一定是全局最优解

$$z=\theta y+(1-\theta)x \qquad \theta=\frac{\delta}{2\|x-y\|_2}$$

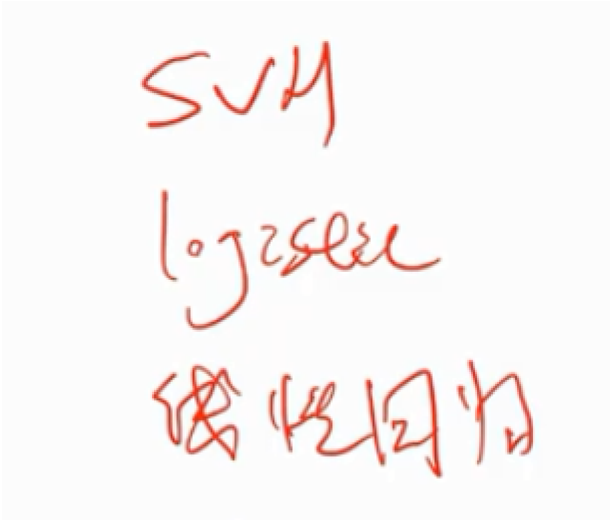
$$\begin{aligned}\|x-z\|_2&=\left\|x-\left(\frac{\delta}{2\|x-y\|_2}y+\left(1-\frac{\delta}{2\|x-y\|_2}\right)x\right)\right\|_2\\&=\left\|\frac{\delta}{2\|x-y\|_2}(x-y)\right\|_2\\&=\frac{\delta}{2}\le\delta\end{aligned}$$

$$f(z)=f(\theta y+(1-\theta)x)\le\theta f(y)+(1-\theta)f(x)<f(x)$$

3.13 凸优化的一般性质

$$\begin{aligned}&\min f(x)\\&x\in C\end{aligned}$$

$$\begin{aligned}&\min f(x)\\&g_i(x)\le 0, i=1,\dots,m\\&h_i(x)=0, i=1,\dots,p\end{aligned}$$

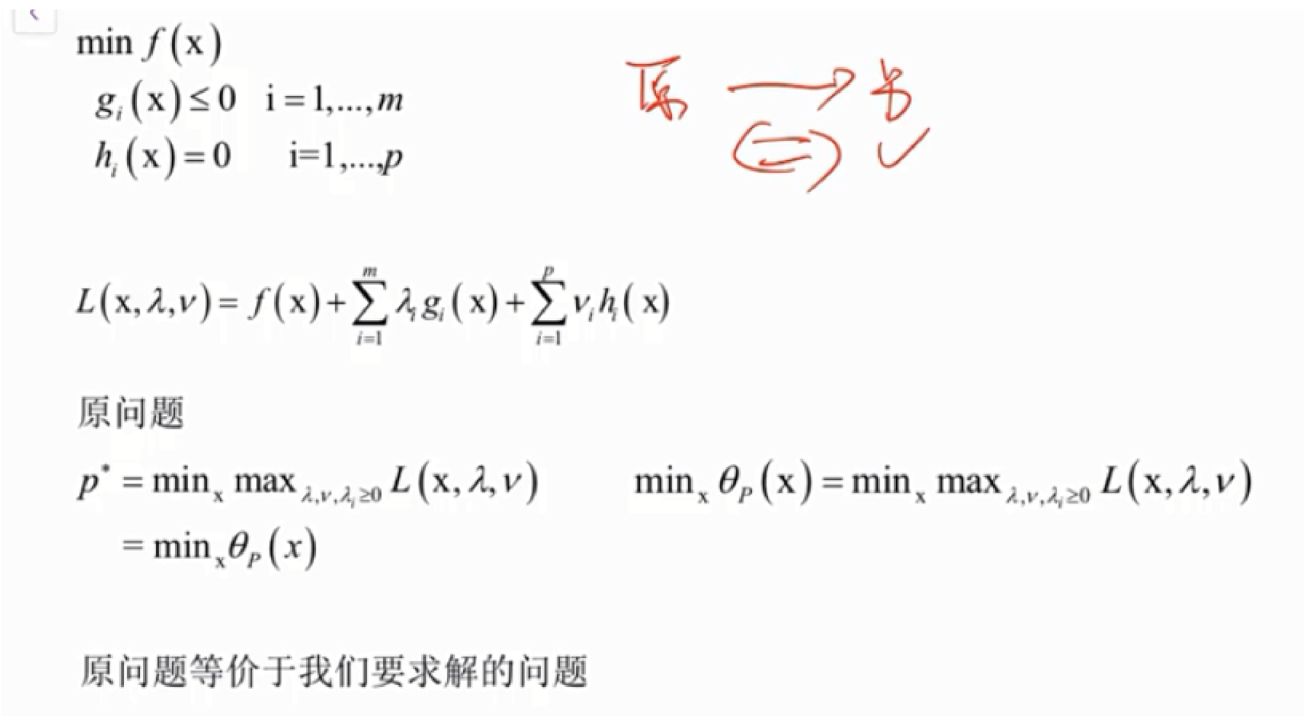


- SVM支持向量机
- logistics回归
- 线性回归
- 鞍点问题

以后在这些问题中会用到凸优化的性质

3.14 拉格朗日对偶

这个是比较难理解的。



$$\min f(x)$$
$$g_i(x) \le 0 \quad i=1,\dots,m$$
$$h_i(x) = 0 \quad i=1,\dots,p$$

$$L(x,\lambda,v) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

原问题

$$p^* = \min_x \max_{\lambda,v,\lambda_i \ge 0} L(x,\lambda,v) \qquad \min_x \theta_P(x) = \min_x \max_{\lambda,v,\lambda_i \ge 0} L(x,\lambda,v)$$
$$= \min_x \theta_P(x)$$

原问题等价于我们要求解的问题

- ## 原问题

  分两步求，接下来操作x

下面来看对偶问题

相对于原问题，从公式上看，就是把 max 和 min 的位置调换。

对偶问题

$$d^* = \max_{\lambda,\nu,\lambda_i \ge 0} \min_x L(x,\lambda,\nu) = \max_{\lambda,\nu,\lambda_i \ge 0} \theta_D(\lambda,\nu)$$

弱对偶

$$d^* = \max_{\lambda,\nu,\lambda_i \ge 0} \min_x L(x,\lambda,\nu) \le \min_x \max_{\lambda,\nu,\lambda_i \ge 0} L(x,\lambda,\nu) = p^*$$

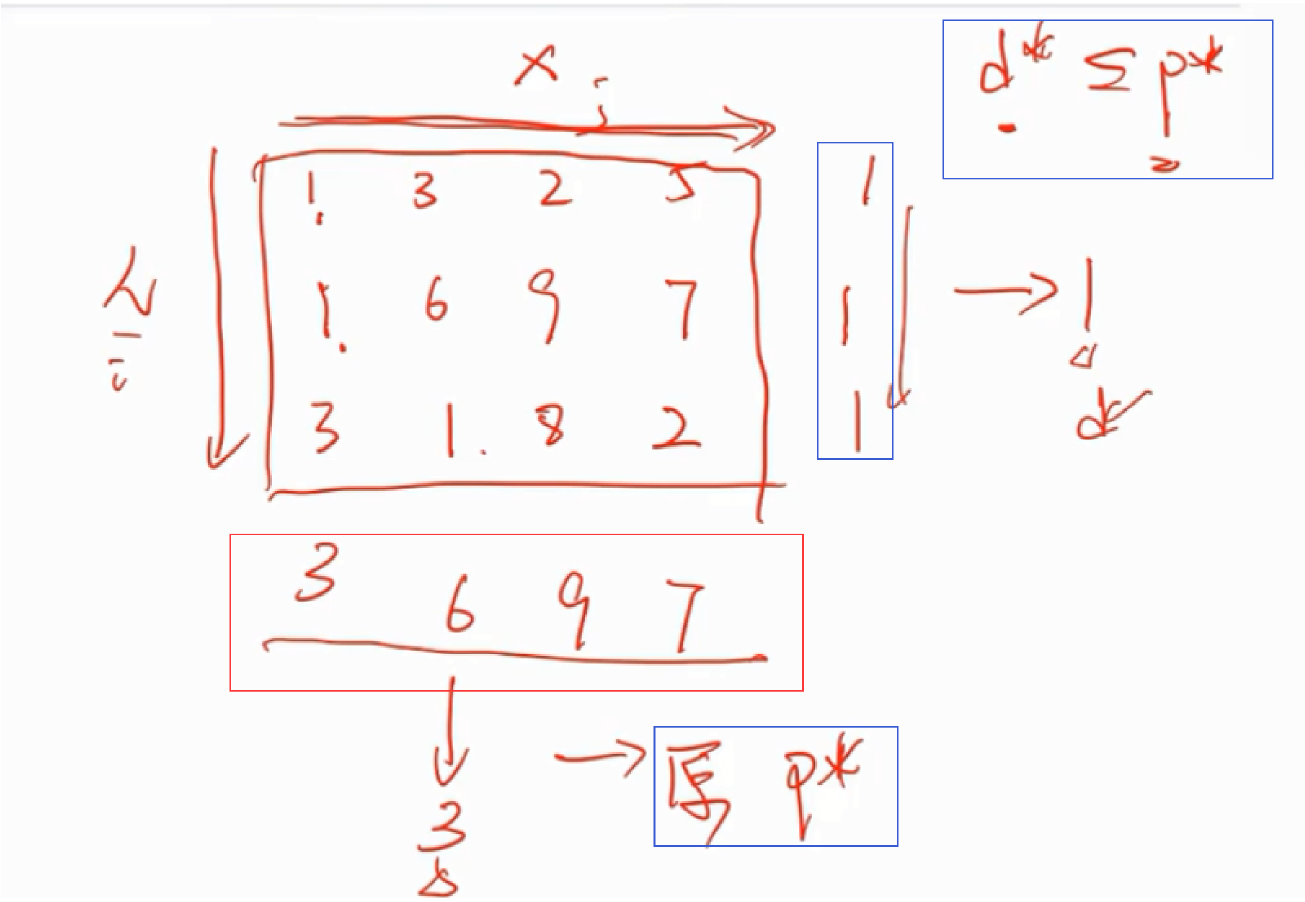
对偶问题的最优解。

弱对偶，有点抽象。我们来看看一个矩阵。

先求出每一列的最大值，然后从最大值里面求出最小值。这里得出的最小值是3。这是原问题。

然后对行标进行操作，对每一行求出极小值。然后从这些极小值里求出最大值。

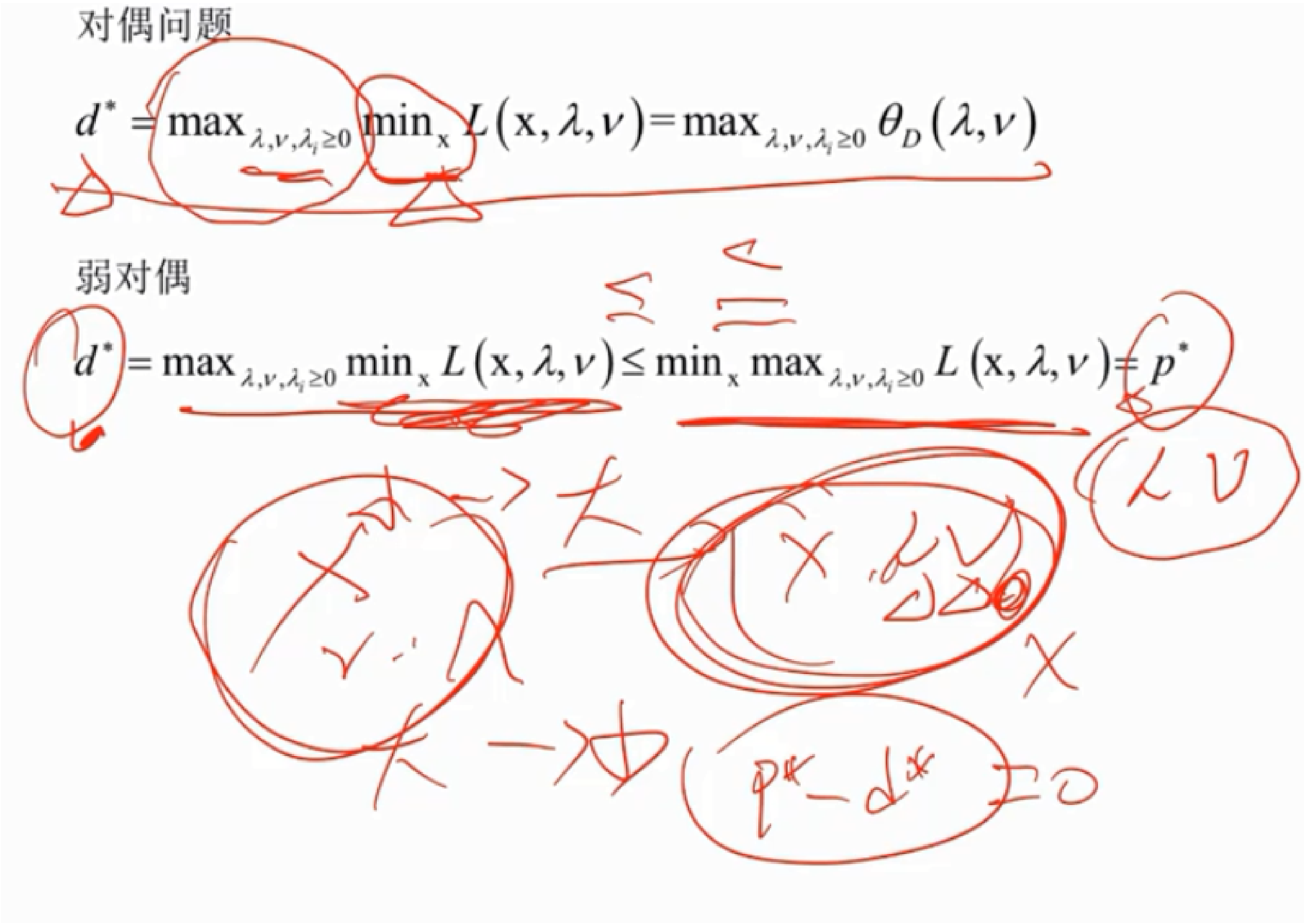
先挑小的问题，在挑大的问题。



这里让我想起以前老师讲过，最小值大于最大值，最大值小于其他的最小值。

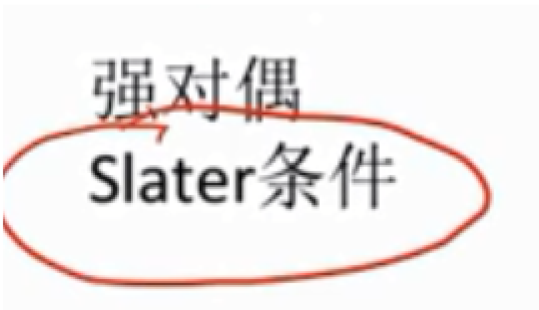
也就是，小于小，大于大。

而且老师是专门研究最优化问题的，有必要请教，说不定老师肯指点方向。



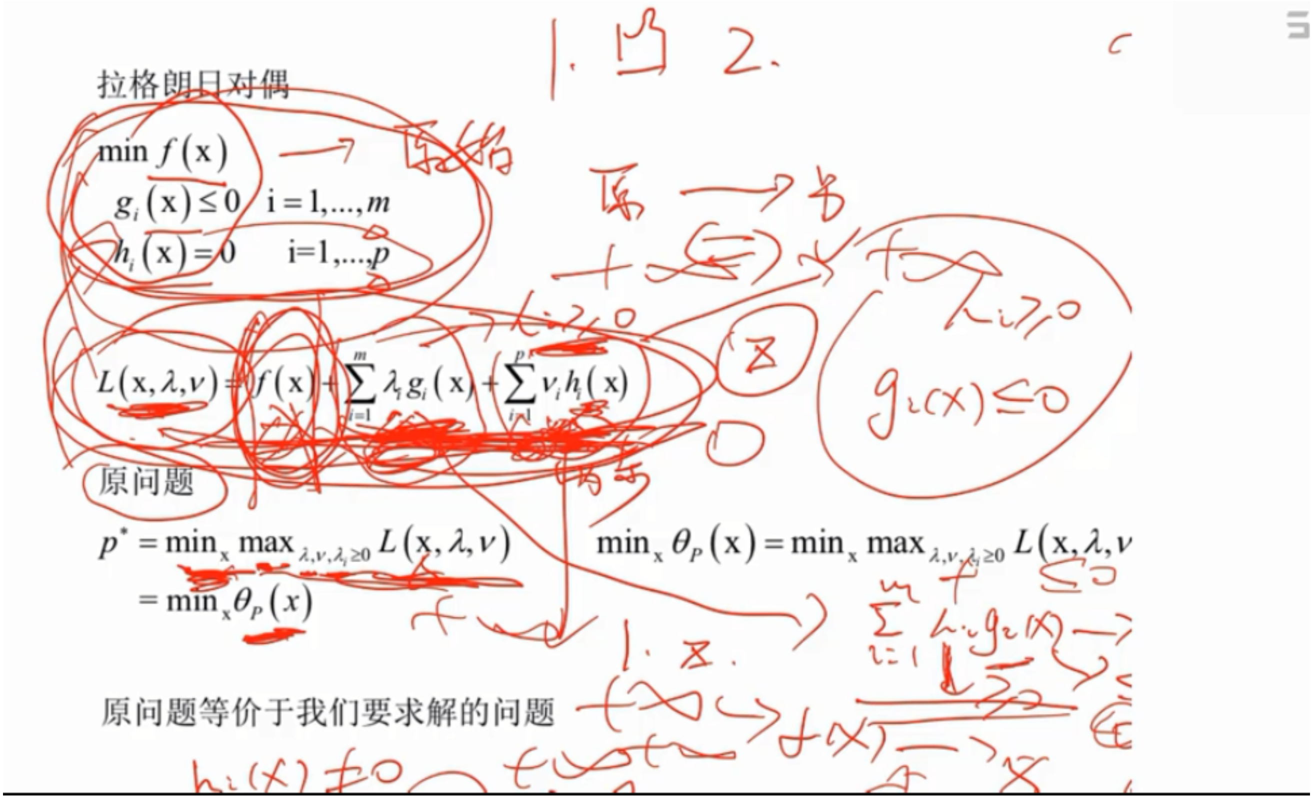
如果两个问题有相等的最优解，那么

强对偶

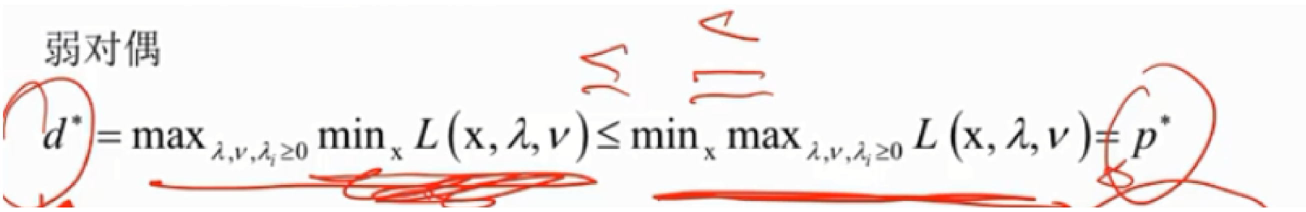


假设我们有这么一个最优化问题。

满足两个条件



这里最好可以取等号



如果能相等

那么怎么保证强对偶呢？

其中有一个slater条件。

那么它是怎么做的呢？

假设有一个最优化问题

$$
\begin{aligned}
&\min f(\mathrm{x}) \\
&\quad g_i(\mathrm{x}) \le 0 \quad \mathrm{i}=1,\dots,m \\
&\quad h_i(\mathrm{x}) = 0 \quad \mathrm{i}=1,\dots,p
\end{aligned}
$$

1. 他是一个凸优化问题，
2. 他至少存在一个可行解

slater 条件是强对偶成立的充分条件，不是必要条件。

Slater $\Longrightarrow$ 强对偶

$\not\Longleftarrow$

总结：

弱对偶是对所有的优化问题成立的。强对偶是有条件的。

3.15 KKT条件

如下图，假设有一个最优化问题，有不等式约束和等式约束，然后构造一个拉格朗日乘子函数。

对偶就是把原问题看成另外一个问题。

拉格朗日乘数法的推广。

$$
\begin{aligned}
&\min f(\mathrm{x}) \\
&\quad g_i(\mathrm{x}) \le 0 \quad \mathrm{i}=1,\dots,m \\
&\quad h_i(\mathrm{x}) = 0 \quad \mathrm{i}=1,\dots,p
\end{aligned}
$$

原 $\longrightarrow$ 对

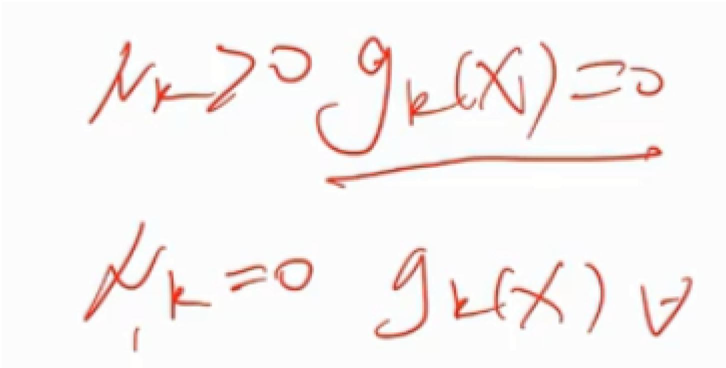
$\Longleftarrow$ ✓

$$
L(\mathrm{x},\lambda,v) = f(\mathrm{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathrm{x}) + \sum_{i=1}^{p} v_i h_i(\mathrm{x})
$$

要满足如下条件

$$
\begin{aligned}
&\nabla_x L(\mathrm{x}^*) = 0 \\
&\mu_k \ge 0 \\
&\boxed{\mu_k g_k(\mathrm{x}^*) = 0} \\
&h_j(\mathrm{x}^*) = 0 \\
&g_k(\mathrm{x}^*) \le 0
\end{aligned}
$$

如果mu k 大于零，那么



这是最核心的。

最终汇总如下图：

KKT条件

$$\min f(x)$$
$$g_i(x) \le 0 \quad i = 1, \ldots, q$$
$$h_i(x) = 0 \quad i = 1, \ldots, p$$

$$L(x, \lambda, \mu) = f(x) + \sum_{j=1}^{p} \lambda_j h_j(x) + \sum_{k=1}^{q} \mu_k g_k(x)$$

$$\nabla_x L(x^*) = 0$$
$$\mu_k \ge 0$$
$$\mu_k g_k(x^*) = 0$$
$$h_j(x^*) = 0$$
$$g_k(x^*) \le 0$$