常用知识点回顾

汇 报 人 | XXX

指导老师 | XXX

目 录 CONTENTS

1、次导数、次微分

1.次导数：函数f(x)在某点(x0,f(x0))不可导时，引入次导数，
通过该不可导点(x0,f(x0))的一条直线的斜率（直线要在函数图像之下或与函数图像重合）

2.次微分：所有次导数的集合称为f在x0的次微分

3.性质：

1.凸函数f:I→R在x0可导，当且仅当次微分只由一个点组成，这个点就是函数在x_0的导数。
2.点x_0是凸函数f的最小值，当且仅当次微分中包含零，也就是说，在上面的图中，我们可以作一条水平的"次切线"。这个性质是"可导函数在极小值的导数是零"的事实的推广。

4.多元空间中次导数称为次梯度，所有次梯度的集合称为次微分，

次导数和次微分的概念可以推广到多元函数。如果f:U→ R是一个实变量凸函数，定义在欧几里得空间Rn内的凸集，则该空间内的向量v称为函数在点x0的次梯度，如果对于所有U内的x，都有：

$$f(x) - f(x_0) \geq v \cdot (x - x_0)$$

所有次梯度的集合称为次微分，记为$\partial f(x_0)$。次微分总是非空的凸紧集。

2.1、泰勒公式应用——近似计算

1.1 带有拉格朗日型余项的泰勒公式

假设函数 $y=f(x)$ 在 x_0 点的某个邻域内 $n+1$ 阶可微，则在此邻域内

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\Lambda+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+R_n(x).$$

其中 $R_n(x)$ 为拉格朗日余项，

$$R_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}\quad(x_0<\xi<x\ ).$$

特别的，若 $x_0=0$，则

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\Lambda+\frac{f^{(n)}(0)}{n!}x^n+\frac{f^{(n+1)}(\theta x)}{(n+1)!}x^{n+1}.\quad 0<\theta<1$$

1.2 带有皮亚诺型余项的泰勒公式

假设函数 $y=f(x)$ 在 x_0 点 n 阶可微，则在 x_0 附近有

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\Lambda+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+o(x-x_0)^n.$$

特别的，若 $x_0=0$ 则

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\Lambda+\frac{f^{(n)}(0)}{n!}x^n+o(x^n).$$

由以上可知泰勒公式的实质是使用一个 n 次多项式 $P_n(x)$ 去逼近一个已知的函数 $f(x)$，而且这种逼近有很好的性质：$P_n(x)$ 与 $f(x)$ 在 x_0 点具有相同的直到 n 阶的导数。

(1) $e^x=1+x+\frac{x^2}{2!}+\cdots+\frac{x^n}{n!}+o(x^n)$；

(2) $\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}+\cdots+(-1)^{m-1}\frac{x^{2m-1}}{(2m-1)!}+o(x^{2m})$；

(3) $\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}+\cdots+(-1)^m\frac{x^{2m}}{(2m)!}+o(x^{2m+1})$；

(4) $\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}+\cdots+(-1)^{n-1}\frac{x^n}{n}+o(x^n)$；

例 1 求 $\ln 1.2$ 的近似值 (精确到 10^{-4})

解：由于 $\ln 1.2=\ln(1+0.2)$，设 $f(x)=\ln(1+x)$，将其在 $x_0=0$ 处展成带拉格朗日型余项的泰勒公式

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\Lambda+(-1)^{n-1}\frac{x^n}{n}+R_n(x)$$，其中

$$R_n(x)=\frac{(-1\ )^n}{(n+1)(1+\xi)^{n+1}}x^{n+1}\quad(0<\xi<x\ )$$，令 $x=0.2$，

则 $0\le\xi\le0.2$ 要使

$$|R_n(x)|=\frac{(0.2)^{n+1}}{(n+1)(1+\xi)^{n+1}}<(0.2)^{n+1}\le10^{-4}$$，只需取 $n=5$ 则

$$\ln 1.2\approx0.2-\frac{1}{2}(0.2)^2+\frac{1}{3}0.2^3-\frac{1}{4}(0.2)^4+\frac{1}{5}(0.2)^5$$
$$=0.20000-0.02000+0.00267-0.00040+0.00006=0.1823$$

其误差 $|R_5|<10^{-4}$

2.2、泰勒公式应用——函数逼近

由于多项式函数性质较好：任意次可导，可微，凸函数，易于计算

可用泰勒公式将一复杂函数在x0点展开成多项式函数，

以此研究原函数的解析性质

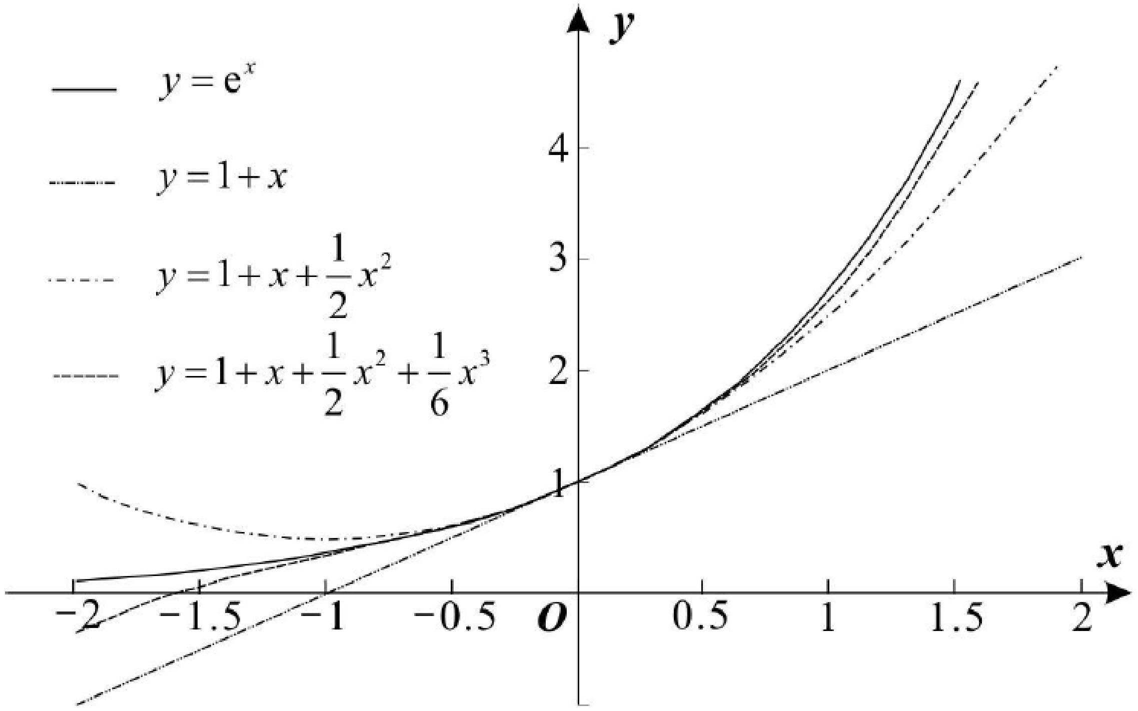


图 1　指数函数的 Taylor 多项式逼近示意图

3.1、梯度下降法

1. **梯度下降法**（**Gradient descent**）是一个**寻找极小值的优化算法**，通常也称为最速下降法。

思想：要使用梯度下降法找到一个函数的局部极小值，必须向函数上当前点对应梯度（或者是近似梯度）的反方向的规定步长距离点进行迭代搜索。如果相反地向梯度正方向迭代进行搜索，则会接近函数的局部极大值点；这个过程则被称为**梯度上升法**。

2. 迭代过程及数学原理推导

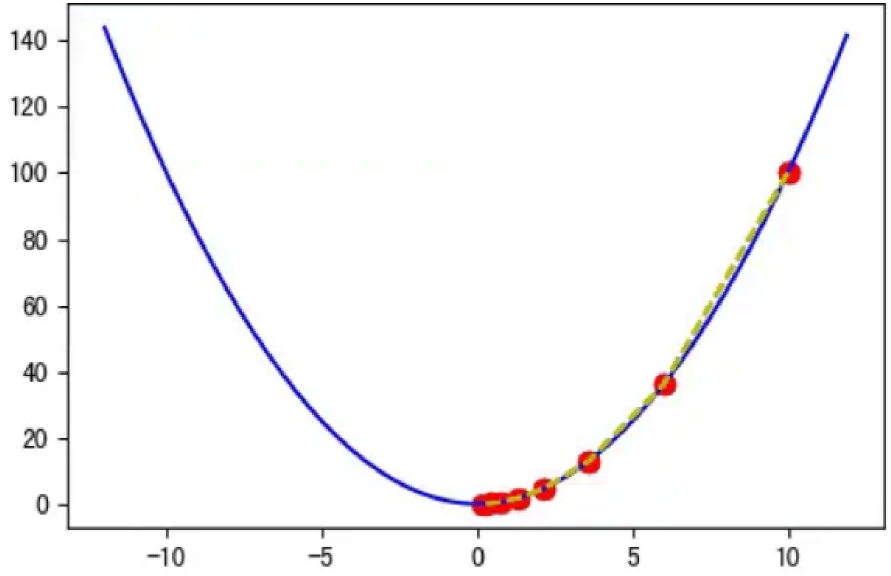
通过不断迭代得到函数极小值，保证f(x+Δx) 比f(x)小

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

原理

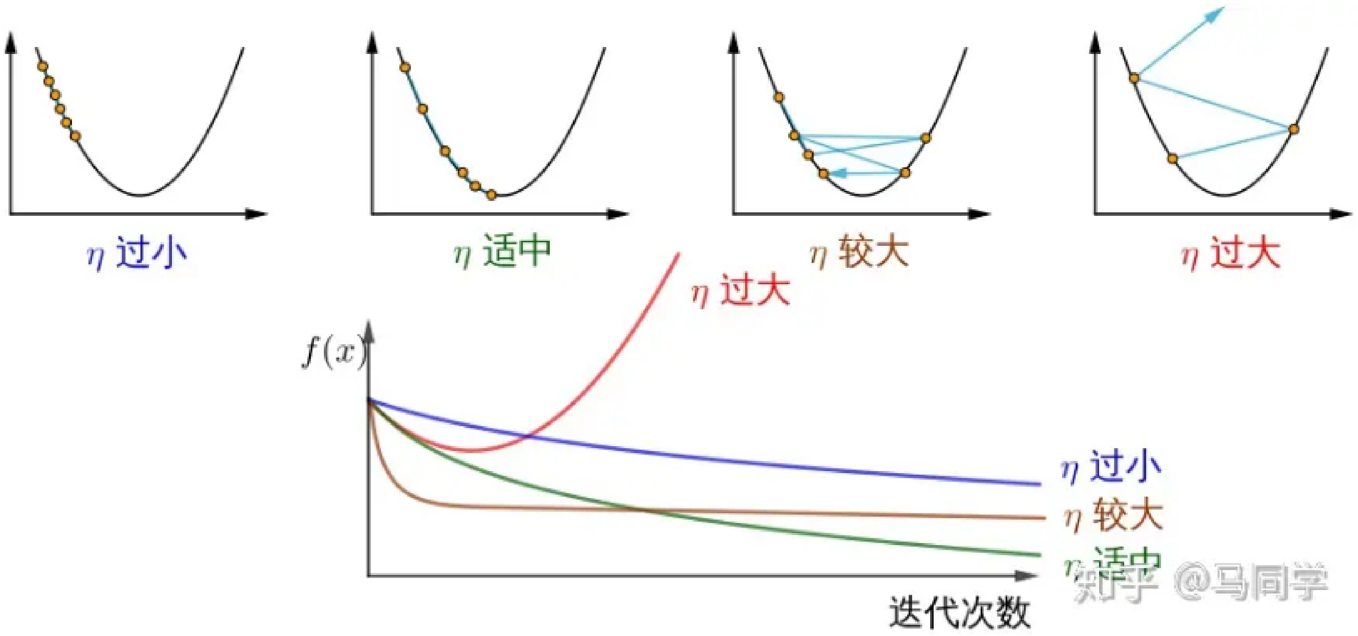
$$f(x+\Delta x) - f(x) = (\nabla f(x))^T \Delta x + o(\Delta x)$$

3.2.例： min f（x）=x^2



3.3.步长选取

总结下，不同的步长η，随着迭代次数的增加，会导致被优化函数$f(x)$的值有不同的变化：



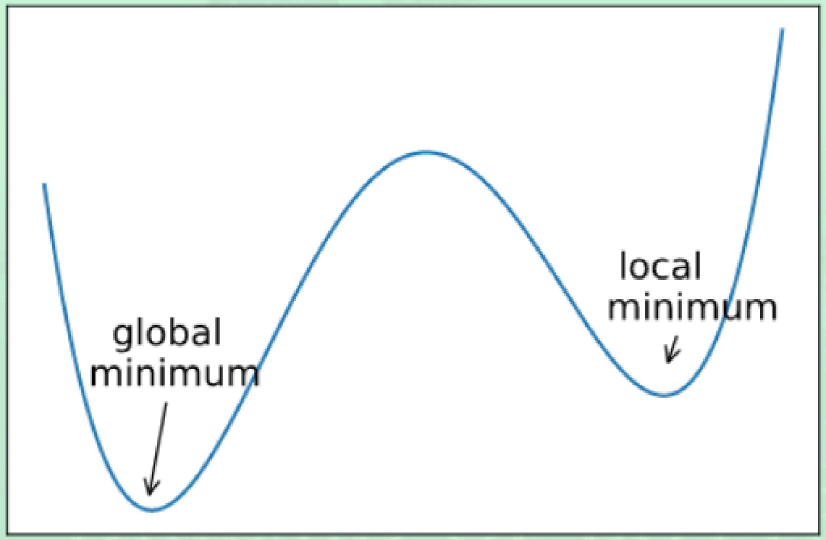
寻找合适的步长η是个手艺活，在工程中可以将上图画出来，根据图像来手动调整：

- $f(x)$往上走（红线），自然是η过大，需要调低
- $f(x)$一开始下降特别急，然后就几乎没有变化（棕线），可能是η较大，需要调低
- $f(x)$几乎是线性变化（蓝线），可能是η过小，需要调高

3.4、常见算法及存在问题

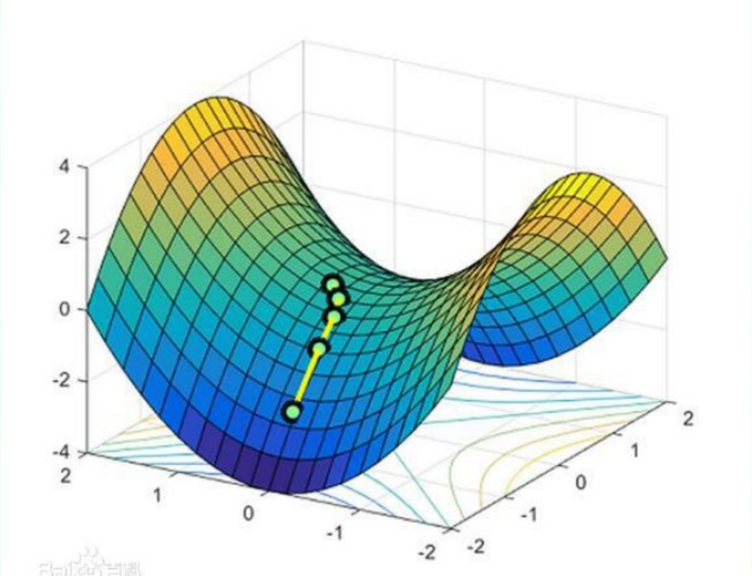
- **批量梯度下降法**：Batch Gradient Descent，简称BGD。求解梯度的过程中用了全量数据。
  - 全局最优解；易于并行实现。
  - 计算代价大，数据量大时，训练过程慢。
- **随机梯度下降法**：Stochastic Gradient Descent，简称SGD。依次选择单个样本计算梯度。
  - 优点：训练速度快；
  - 缺点：准确度下降，并不是全局最优；不易于并行实现。
- **小批量梯度下降法**：Mini-batch Gradient Descent，简称MBGD。每次更新参数时使用b个样本。（b一般为10）。
  - 两种方法的性能之间取得一个折中。

- **局部极小值**
  - 梯度下降可能在局部最小的点收敛。



鞍点

- 鞍点是指梯度为0，Hessian矩阵既不是正定也不是负定，即不定的点。如函数$x^2 - y^2$在$(0, 0)$点梯度为0，点。



4. 1、范数与稀疏性、正则化

1.稀疏性概念

2.向量的L0，L1，L2范数

3.矩阵的F范数，L21范数，

4.正则化：对损失函数中某些参数做限制

常用：L1正则，L2正则

- L1正则化可以产生稀疏权值矩阵，即产生一个稀疏模型，可以用于特征选择
- L2正则化可以防止模型过拟合（overfitting）；一定程度上，L1也可以防止过拟合

矩阵的F范数定义为：矩阵元素绝对值的平方和再开方

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j}|^2 \right)^{\frac{1}{2}}$$

矩阵的L2,1范数定义为：矩阵A的每一行的L2范数之和

$$\|A\|_{2,1} = \sum_{i=1}^{m} \sqrt{\sum_{j=1}^{n} a_{i,j}^2} = \sum_{i=1}^{m} \|a_{i,:}\|_2$$

4.2、L1与L2区别

L1和L2的区别

稀疏性：

- L1范数正则化能够产生稀疏的权重矩阵，即其解的许多元素为0。这一特性使得L1范数正则化成为一种有效的特征选择方法。
- L2范数正则化则不具备这种性质，它会尽可能地缩小所有权重的大小，但不会将它们压缩到零。

解的稳定性：

- L1范数正则化可能产生不稳定的解，即微小的数据变动可能导致解的大幅度变化。
- 相比之下，L2范数正则化会产生更稳定的解，对数据的微小变动更加鲁棒。

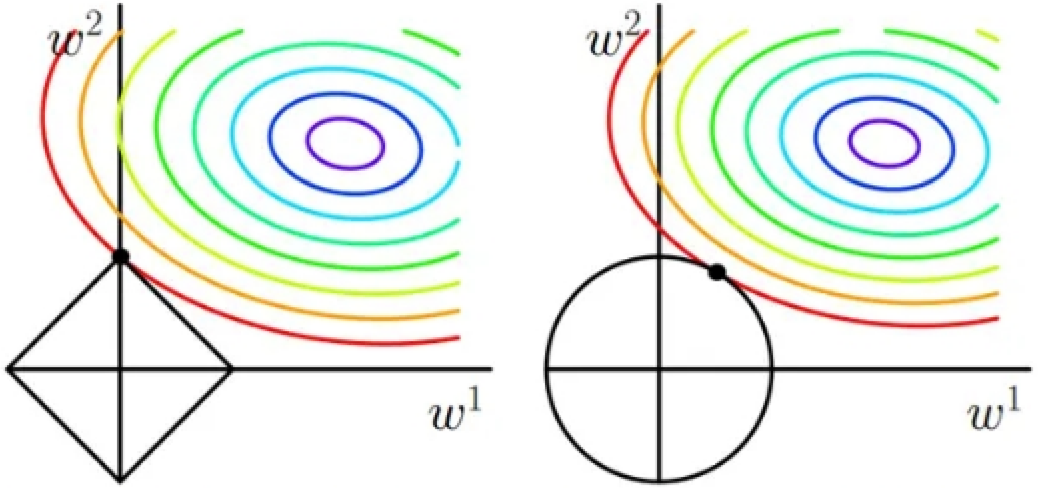
解的唯一性：

- L1范数正则化可能存在多个解，因为L1范数的等高线在高维空间中形状为菱形，这可能会与损失函数的等高线在某些顶点相交，形成多个最优解。
- 相比之下，L2范数正则化总是有唯一解，因为L2范数的等高线在高维空间中形状为球形，与损失函数的等高线只有一点交点。

4.3、几何解释

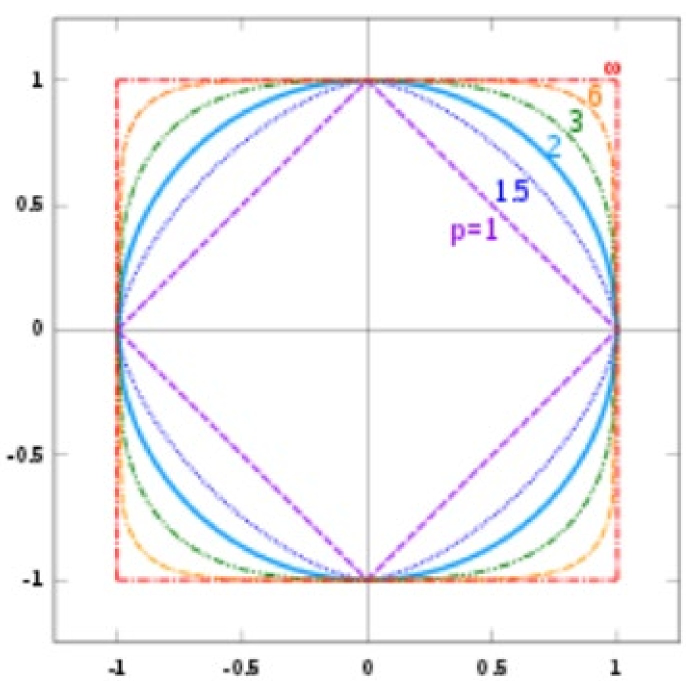
在深度学习/机器学习中，Lp范数往往是作为一个正则化项加在损失函数的后面用以优化参数，L1（黑色菱形）与参数权重（彩色等高线）相交之处多是在坐标轴上，所以多产生稀疏矩阵；L2范数（黑色圆形）与参数矩阵多相交在较低值参数区域，所以能防止一些较大参数的产生，平滑的优化参数。

对于一个二维的范数图像来看，p越大，范数



(a) ℓ_1-ball meets quadratic function. ℓ_1-ball has corners. It's very likely that the meet-point is at one of the corners.

(b) ℓ_2-ball meets quadratic function. ℓ_2-ball has no corner. It is very unlikely that the meet-point is on any of axes.



4.4、2010ADMM中正则项

be considered. One of the most basic and successful image regularization models is the ROF model [34], which reads

$$\min_{u \in V} \left\{ F_{\mathrm{rof}}(u) = R_{\mathrm{rof}}(\nabla u) + \frac{\alpha}{2}\|Ku - f\|_V^2 \right\}, \tag{3.2}$$

where

$$R_{\mathrm{rof}}(\nabla u) = \mathrm{TV}(u) = \sum_{1 \le i,j \le N} |(\nabla u)_{i,j}| \tag{3.3}$$

5.1、线性规划对偶问题

引例：甲方现有两种用来制作圆桌与衣柜的木料。第一种木料72立方米；第二种木料56立方米。已知每个圆桌与衣柜所需的木料数量及利润如表：

表 3-1

单位消耗 木料 产品	第一种木料	第二种木料	利润（元）
圆桌	0.18	0.08	60
衣柜	0.09	0.28	100
现有木料	72 $米^3$	56 $米^3$	

甲方对现有木料的处理方法有两种：

（Ⅰ）将木料卖给乙方，由乙方用来加工成圆桌与衣柜，并销售。问：甲方如何确定这两种木料的售价，才能保证这一方法得以实现？

（Ⅱ）如果上一种方法不能实现，甲方就自己加工成圆桌与衣柜，并销售。试确定圆桌与衣柜各生产多少个，方能获取最大的利润？

分析：

- 甲把原材料卖给已，甲所出售的价格必须大于或等于自己加工所得的利润值。
- 乙买甲的原材料进行加工与销售，其利润值越大越好，所以乙方要求购买原材料的成本越低越好。
- 故方案（Ⅰ）的实现，必须是双方的利益达到一个平衡点。

（I）设第一种木料的售出单价为每立方米x_1元，第二种木料的售出单价为每立方米x_2元。则x_1与x_2应满足甲、乙的双方要求，故（Ⅰ）的数学模型为：

$$Min \quad Z = 72x_1 + 56x_2$$

$$s.t. \quad \begin{cases} 0.18x_1 + 0.08x_2 \geq 60 \\ 0.09x_1 + 0.28x_2 \geq 100 \\ x_1, x_2 \geq 0 \end{cases}$$

（Ⅱ）设圆桌的生产个数为y_1 ,衣柜的生产个数为y_2 ,则（Ⅱ）的数学模型为:

$$Max \quad W = 60y_1 + 100y_2$$

$$s.t \quad \begin{cases} 0.18y_1 + 0.09y_2 \leq 72 \\ 0.08y_1 + 0.28y_2 \leq 56 \\ y_1, y_2 \geq 0 \end{cases}$$

5.1、对偶问题

线性规划的原问题与对偶问题

$$(P)\quad \begin{array}{ll} Min & Z = cX \\ s.t. & \begin{cases} AX \ge b \\ X \ge 0 \end{cases} \end{array} \qquad (D)\quad \begin{array}{ll} Max & w = Yb \\ s.t. & \begin{cases} YA \le C \\ Y \ge 0 \end{cases} \end{array}$$

原问题（对偶问题）	对偶问题（原问题）
目标函数min z	目标函数max w
约束条件： m个 第i个约束类型为“≤” 第i个约束类型为“≥” 第i个约束类型为“＝”	变量数： m个 第i个变量≤0 第i个变量≥0 第i个变量是自由变量
变量数：n个 第j个变量≥0 第j个变量≤0 第j个变量是自由变量	约束条件：n个 第j个约束类型为“≤” 第j个约束类型为“≥” 第j个约束类型为“＝”
决策系数 c	右端项c
右端项b	决策系数b
约束矩阵的第j列	约束矩阵的第j行

$$\begin{array}{ll} Min & Z = 72x_1 + 56x_2 \\ s.t. & \begin{cases} 0.18x_1 + 0.08x_2 \ge 60 \\ 0.09x_1 + 0.28x_2 \ge 100 \\ x_1, x_2 \ge 0 \end{cases} \end{array} \qquad \begin{array}{ll} Max & W = 60y_1 + 100y_2 \\ s.t & \begin{cases} 0.18y_1 + 0.09y_2 \le 72 \\ 0.08y_1 + 0.28y_2 \le 56 \\ y_1, y_2 \ge 0 \end{cases} \end{array}$$

5.2、对偶理论

$$
\begin{array}{ll}
Min \quad Z = cX & Max \quad w = Yb \\
(P) \quad s.t. \begin{cases} AX \geq b \\ X \geq 0 \end{cases} & (D) \quad s.t. \begin{cases} YA \leq C \\ Y \geq 0 \end{cases}
\end{array}
$$

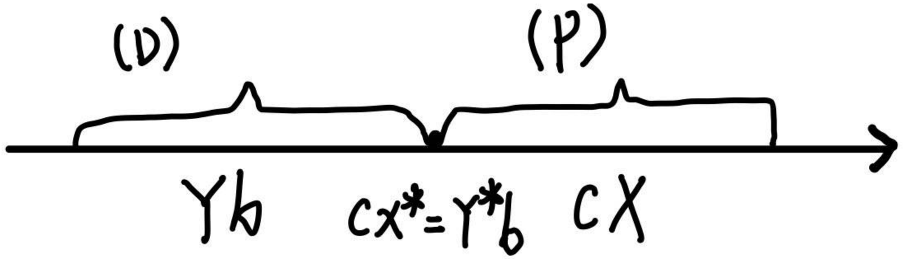
1.弱对偶定理:

若X^0是原问题(P)的可行解，Y^0是对偶问题(D)的可行解，则$cX^0 \geq Y^0$b

表明:

（P）的任一可行解对应目标函数值总是（D）目标函数值的上界,
（D）的任一可行解对应目标函数值总是（P）目标函数值的下界.

2.强对偶定理:



若X^0，Y^0分别是原问题(P)和对偶问题(D)的可行解,

且$cX^0 = Y^0$b，则X^0，Y^0分别是原问题(P)和对偶问题(D)的最优解

（若（P）和（D）中有一个问题有最优解，则另一个问题也有最优解，
且它们的目标函数值相等）

6.1、拉格朗日函数、对偶函数、对偶问题

一般优化问题:

$$\min \ f(x)$$
$$\text{s.t.} \quad g_i(x) \le 0, \ i=1,2,\cdots,s.$$
$$h_j(x) = 0, \ j=1,2,\cdots,t.$$

的Lagrange函数为

$$L(x,\lambda,v) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x) + \sum_{j=1}^{t} v_j h_j(x),$$

Lagrange对偶函数为

$$g(\lambda,v) = \inf_{x\in D} L(x,\lambda,v)$$

Lagrange对偶问题为

$$\max \ g(\lambda,v)$$

6.2、弱对偶性、强对偶性

弱对偶性：

The optimal value of the Lagrange dual problem, which we denote $d^\star$, is, by definition, the best lower bound on $p^\star$ that can be obtained from the Lagrange dual function. In particular, we have the simple but important inequality

$$d^\star \leq p^\star, \tag{5.23}$$

强对偶性

If the equality

$$d^\star = p^\star \tag{5.24}$$

holds, *i.e.*, the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold. But if the primal problem (5.1) is convex, *i.e.*, of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & Ax = b, \end{array} \tag{5.25}$$

with $f_0, \dots, f_m$ convex, we usually (but not always) have strong duality. There are

6.3、拉格朗日对偶问题的鞍点解释

1.鞍点定义

We refer to a pair $\tilde{w} \in W$, $\tilde{z} \in Z$ as a *saddle-point* for f (and W and Z) if

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\tilde{w}$ minimizes $f(w, \tilde{z})$ (over $w \in W$) and $\tilde{z}$ maximizes $f(\tilde{w}, z)$ (over $z \in Z$):

$$f(\tilde{w}, \tilde{z}) = \inf_{w \in W} f(w, \tilde{z}), \qquad f(\tilde{w}, \tilde{z}) = \sup_{z \in Z} f(\tilde{w}, z).$$

2.鞍点解释

若x^*和y^*分别是原问题和对偶问题的最优点，且强对偶性成立，则它们是lagrange函数的一个鞍点；

反之，若（x，λ）是lagrange函数的一个鞍点，那么，x是原问题的最优解，λ是对偶问题的最优解，且 x^*=y^*

$$L(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x)$$

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*)$$

6.3、拉格朗日对偶问题的鞍点解释

3.2010ADMM中优化问题转化为鞍点问题

an auxiliary variable $p \in Q$ is introduced for ∇u. The model (3.2) is thus equivalent to

$$\begin{aligned} &\min_{u \in V, p \in Q} \left\{ G_{\mathrm{rof}}(u,p) = R_{\mathrm{rof}}(p) + \frac{\alpha}{2}\|Ku - f\|_V^2 \right\} \\ &\text{s.t.} \qquad p = \nabla u, \end{aligned} \tag{3.5}$$

which is a constrained optimization problem.

4.1. Augmented Lagrangian method for the ROF model. We first define the augmented Lagrangian functional for the constrained optimization problem (3.5) as follows:

$$\mathscr{L}_{\mathrm{rof}}(v,q;\mu) = R_{\mathrm{rof}}(q) + \frac{\alpha}{2}\|Kv - f\|_V^2 + (\mu, q - \nabla v)_Q + \frac{r}{2}\|q - \nabla v\|_Q^2, \tag{4.1}$$

where $\mu \in Q$ is the Lagrange multiplier and r is a positive constant. For the augmented Lagrangian method for (3.5), we consider the following saddle-point problem:

$$\begin{aligned} &\text{Find} \quad (u,p;\lambda) \in V \times Q \times Q \\ &\text{s.t.} \qquad \mathscr{L}_{\mathrm{rof}}(u,p;\mu) \le \mathscr{L}_{\mathrm{rof}}(u,p;\lambda) \le \mathscr{L}_{\mathrm{rof}}(v,q;\lambda) \ \forall (v,q;\mu) \in V \times Q \times Q. \end{aligned} \tag{4.2}$$

7.1、带约束优化问题的拉格朗日方法

$$\min \quad z = f(x,y)$$

$$s.t. \quad \varphi(x,y) = 0$$

$$L(x,y,\lambda) = f(x,y) + \lambda\varphi(x,y)$$

$$\begin{cases} L_x = f_x(x,y) + \lambda\varphi_x(x,y) = 0 \\ L_y = f_y(x,y) + \lambda\varphi_y(x,y) = 0 \\ L_\lambda = \varphi(x,y) = 0 \end{cases}$$

得稳定点 $P_0(x_0,y_0,\lambda_0)$

例 2 抛物面

$$x^2 + y^2 = z$$

被平面

$$x + y + z = 1$$

截成一个椭圆. 求这个椭圆到原点的最长与最短距离.

解 这个问题实质上就是要求函数

$$f(x,y,z) = x^2 + y^2 + z^2$$

在条件 $x^2+y^2-z=0$ 及 $x+y+z-1=0$ 下的最大、最小值问题. 应用拉格朗日乘数法, 令

$$L(x,y,z,\lambda,\mu) = x^2 + y^2 + z^2 + \lambda(x^2 + y^2 - z) + \mu(x + y + z - 1).$$

对 L 求一阶偏导数, 并令它们都等于 0, 则有

$$\begin{cases} L_x = 2x + 2x\lambda + \mu = 0, \\ L_y = 2y + 2y\lambda + \mu = 0, \\ L_z = 2z - \lambda + \mu = 0, \\ L_\lambda = x^2 + y^2 - z = 0, \\ L_\mu = x + y + z - 1 = 0. \end{cases}$$

7.2、对偶上升法

Consider the equality-constrained convex optimization problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b, \end{aligned} \tag{1}$$

where variable $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $f : \mathbb{R}^n \to \mathbb{R}$ is convex.

The Lagrangian for problem (1) is

$$L(x, y) = f(x) + y^T(Ax - b) \tag{2}$$

and the dual function is

$$g(y) = \min_x L(x, y),$$

where $y \in \mathbb{R}^m$ is the dual variable or Lagrangian multiplier.

The dual problem is

$$\max_y g(y) \tag{3}$$

7.2、对偶上升法

Assuming that strong duality holds, the optimal values $f(x^*)$, $g(y^*)$ of the primal and dual problems are the same. We can recover an optimal point x^* from a dual optimal point y^* as

$$x^* = \underset{x}{\operatorname{argmin}}\ L(x, y^*), \tag{4}$$

provided there is only one minimizer of $L(x, y^*)$. (This is the case if f is strictly convex)

Based on (4), the dual ascent method consists of iterating the updates

$$x^{k+1} := \underset{x}{\operatorname{argmin}}\ L(x, y^k), \tag{5}$$

$$y^{k+1} := y^k + \alpha^k (Ax^{k+1} - b), \tag{6}$$

where $\alpha^k > 0$ is a step size, and k is the iteration number.

Here,

the first step (5) is an x-minimization step;

the second step (6) is a dual variable update (using gradient ascent).

7.2、增广拉格朗日方法

$$L_\rho(x, y) = f(x) + y^T(Ax - b) + \frac{\rho}{2}\|Ax - b\|^2, \tag{7}$$

不同点

1. **形式**：标准拉格朗日函数的一般形式包括目标函数和约束函数的线性组合，而增广拉格朗日函数引入了额外的惩罚项，通常是约束函数的平方范数。这是它们最显著的区别之一。

2. **优化方法**：标准拉格朗日函数通常用于构建拉格朗日对偶问题，然后通过解对偶问题来获得原始问题的解。与此不同，增广拉格朗日函数通常与交替方向乘子法（ADMM）等优化方法结合使用，其中乘子和变量交替更新。这使得增广拉格朗日方法在一些情况下更适用于复杂的非凸优化问题。

3. **惩罚项**：最大的不同点是增广拉格朗日函数引入了一个惩罚项，通常表示为$\frac{\rho}{2}\|\mathbf{g}(\mathbf{x})\|^2$，其中$\rho$是惩罚参数，而标准拉格朗日函数没有这个额外的惩罚项。这个惩罚项的存在使增广拉格朗日函数更加适用于处理约束的违反情况，当约束条件被违反时，它会对目标函数产生额外的惩罚。

7.3、ADMM

考虑一个带有线性约束的凸优化问题：

目标函数： minimize f(x) + g(z)

约束条件： subject to Ax + Bz = c

其中，`x` 和 `z` 是优化变量，`f(x)` 和 `g(z)` 是凸函数，`A` 和 `B` 是系数矩阵，`c` 是常数向量。

ADMM的基本原理包括以下步骤：

1. **构建增广拉格朗日函数（Augmented Lagrangian Function）：**

$$L\rho\ (x,z,u) = f(x) + g(z) + u^T(Ax + Bz - c) + (\rho\ /2)\|Ax + Bz - c\|^2$$

其中，$L\rho\ (x,z,u)$ 是增广拉格朗日函数，`u` 是拉格朗日乘子，`ρ` 是正的惩罚参数。

2. **迭代优化步骤：**

a. 固定 `x` 和 `z`，更新拉格朗日乘子 `u`：

$$u^{k+1} = u^k + \rho\ (Ax^k + Bz^k - c)$$

b. 固定 `u`，分别更新 `x` 和 `z`：

- 更新 `x`（子问题1）：

$$x^{k+1} = argminf(x) + u^{k+1T}Ax + (\rho\ /2)\|Ax + Bz^k - c\|^2$$

- 更新 `z`（子问题2）：

$$z^{k+1} = argming(z) + u^{k+1T}Bz + (\rho\ /2)\|Ax^{k+1} + Bz - c\|^2$$

c. 重复步骤 a 和 b 直到满足收敛条件（例如，目标函数值的变化很小）或达到最大迭代次数。

感 谢 老 师 的 悉 心 指 导

T H A N K　Y O U !

汇 报 人 ｜ XXX　　　　指 导 老 师 ｜ XXX