一、数学基础

范数：满足正定性、齐次性、三角不等式的向量到数的映射

如 p 范数$\|X\|_p$

椭球范数：$\|X\|_A \triangleq (X^{\mathrm{T}}AX)^{\frac{1}{2}}$，其中$A$半正定

0 范数：非零元素的个数（不是范数，更不是$p=0$ 的 p 范数）
矩阵范数：满足正定性、齐次性、三角不等式、相容性的矩阵到数的映射
如矩阵 p 范数$\|A\|_{mp}$

算子 p 范数，即从属于向量 p 范数的矩阵范数：$\|A\|_p \triangleq \max\limits_{X\neq 0}\frac{\|AX\|_p}{\|X\|_p} = \max\limits_{\|X\|_p=1}\|AX\|_p$

特例：算子 1 范数为最大列和，算子∞范数为最大行和
算子 2 范数为谱范数$\|A\|_2 = \sqrt{\lambda_{max}(A^{\mathrm{T}}A)}$

正定矩阵

充要条件：

特征值大于 0
合同于单位矩阵 I，即存在$C^{\mathrm{T}}AC = I$，其中C可逆。或者说$A = C^{\mathrm{T}}C$
A的顺序主子式大于 0
A的正惯性指数等于 n

必要条件：

A可逆
A^{-1}正定
A的主对角线元素$a_{ii} > 0$
A的所有主子式大于 0
A^m正定

柯西不等式

内积小于等于模的积
即$|X^TY| \leq \|X\|\|Y\|$，当X和Y线性相关时取等号

可微

存在 n 维向量 L 使得$\lim\limits_{\|P\|\to 0}\frac{f(X^0+P)-f(X^0)-L^TP}{\|P\|} = 0$ 对于定义域 $\mathrm{D} \subset \mathrm{R}^{\mathrm{n}}$内的任何$X^0$均成立

梯度$\nabla f(X^0) = \mathrm{L} = (\frac{\partial f(X^0)}{\partial x_1}, \frac{\partial f(X^0)}{\partial x_2}, \dots, \frac{\partial f(X^0)}{\partial x_n})^{\mathrm{T}}$

设$\|P\| = 1$，则方向导数$\frac{\partial f(X)}{\partial P} = (\nabla f(X), P) = \|\nabla f(X)\| cos(\nabla f(X), P)$

Hesse矩阵

向量值函数

设$g: D \subset R^n \to R^m$，即$g(X) = (g_1(X), g_2(X), \dots, g_m(X))^{\mathrm{T}}$，$X = (x_1, x_2, \dots, x_n)^{\mathrm{T}}$
若$g_1(X), g_2(X), \dots, g_m(X)$在$X^0$处均可微，则称$g(X)$在$X^0$处可微

Jacobi矩阵

若$g(X)$在X^0处可微，则称$g'(X^0) = J_g(X^0) = \begin{pmatrix} \frac{\partial g_1(X^0)}{\partial x_1} & \frac{\partial g_1(X^0)}{\partial x_2} & \cdots & \frac{\partial g_1(X^0)}{\partial x_n} \\ \frac{\partial g_2(X^0)}{\partial x_1} & \frac{\partial g_2(X^0)}{\partial x_2} & \cdots & \frac{\partial g_2(X^0)}{\partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g_m(X^0)}{\partial x_1} & \frac{\partial g_m(X^0)}{\partial x_2} & \cdots & \frac{\partial g_m(X^0)}{\partial x_n} \end{pmatrix}$为$g(X)$在$X^0$处的导数或 Jacobi 矩阵（也就是对$g_i(X^0)$求 n 维导数，分别写在第$i$行）
称$g'(X^0)^{\mathrm{T}}$为$g(X)$在X^0处的梯度，记作$\nabla g(X^0)$（梯度就转置一下）

Hesse矩阵

设 $f: D \subset R^n \to R$ 具有二阶连续偏导数，则$\nabla^2 f(X) = \nabla(\nabla f(X)) = \begin{pmatrix} \frac{\partial^2 f(X)}{\partial {x_1}^2} & \frac{\partial^2 f(X)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(X)}{\partial {x_2}^2} & \cdots & \frac{\partial^2 f(X)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(X)}{\partial {x_n}^2} \end{pmatrix}$为 f(X)的二阶导数，称之为 f(X)的 Hesse 矩阵

常用梯度公式

$\nabla(\mathrm{b}^{\mathrm{T}}\mathrm{X}) = \mathrm{b}$
$\nabla(\mathrm{AX}) = \mathrm{A}^{\mathrm{T}}\mathrm{X}$
$\nabla(\mathrm{X}^{\mathrm{T}}\mathrm{AX}) = \mathrm{AX} + \mathrm{A}^{\mathrm{T}}\mathrm{X}$
梯度计算同样适用链式规则

Taylor公式

$$f(X) = f(X^0) + \nabla f(X)^{\mathrm{T}}(X - X^0) + \frac{1}{2}(X - X^0)^{\mathrm{T}}\nabla^2 f(X^0)(X - X^0) + \mathrm{o}\left(\|X - X^0\|^2\right)$$

~~内点和外点~~

内点：存在X^0的邻域$N(X^0, \delta) \subset D$
边界点：对于X^0的任意邻域，既存在$X^1 \in D$，又存在$X^2 \notin D$
外点：不是内点或者外点的点
极限点：X^0的邻域内包含 D 中其他点

~~开集和闭集~~

开集：D 的每一个点都是内点
闭集：D 的每个极限点都属于 D
（严格）局部/全局极小点

驻点

X^*是 D 的内点，且$\nabla f(X^*) = 0$，则X^*是$f(X)$的驻点
一阶必要条件：若X^*是$f(X)$的局部极小点，则$\nabla f(X^*) = 0$ 也即X^*是$f(X)$的驻点
二阶必要条件：若X^*是$f(X)$的局部极小点，则$\nabla^2 f(X^*)$也即 Hesse 矩阵半正定
二阶充分条件：如果X^*是$f(X)$的驻点，且 Hesse 矩阵正定，则X^*是$f(X)$的严格局部极小点

凸集

任取$X^1, X^2 \in D$，均有$\alpha X^1 + (1 - \alpha)X^2 \in D$，其中$\alpha \in [0,1]$
（也就是 D 中任意两点的连线都在 D 中）
凸组合：设$X^1, X^2, ..., X^m \in R^n$，$\alpha_1, \alpha_2, ..., \alpha_m \geq 0$，且$\sum_{i=1}^m \alpha_i = 1$，则$\sum_{i=1}^m \alpha_i X^i$称为$X^1, X^2, ..., X^m$的一个凸组合
D 为凸集的充要条件是 D 种任意个点的凸组合仍在 D 中

凸函数

任取$X^1, X^2 \in D$，$\alpha \in [0,1]$，都有$f(\alpha X^1 + (1 - \alpha)X^2) \leq \alpha f(X^1) + (1 - \alpha)f(X^2)$
也即函数值小于等于连线上的值，弦在弧之上

充要条件

任取$X^1, X^2, ..., X^m \in D$，$\alpha_1, \alpha_2, ..., \alpha_m \geq 0$，且$\sum_{i=1}^m \alpha_i = 1$，则$f(\sum_{i=1}^m \alpha_i X^i) \leq \sum_{i=1}^m \alpha_i f(X^i)$
凸组合的函数值小于等于函数值的凸组合，也即平面在曲面之上

$f(X)$是凸函数的充要条件是 Hesse 矩阵半正定
$f(X)$是严格凸函数的充分条件是 Hesse 矩阵正定

凸规划

目标函数是凸函数，可行集是凸集

凸规划任何局部最优解为全局最优解，且最优解集也是凸集

如果目标函数$f(X)$是严格凸函数，如果有最优解，则最优解唯一

最优性条件：对于任意$X \in S$，都有$(X - X^*)^{\mathrm{T}} \nabla f(X^*) \geq 0$

（从X^*朝 S 内任何方向走都不会减小）

如果 $\mathrm{S} = \mathrm{R}^{\mathrm{n}}$，则有$\nabla f(X^*) = 0$

二、单纯形方法

标准形式：

$\min f(X)=C^{\mathrm{T}}X$
s.t.
$\mathrm{AX}=\mathrm{b}$ （$\mathrm{b}\geq 0$）
$\mathrm{X}\geq 0$

线性规划化为标准型：

如果目标函数是 max 就变成 min，C取相反数
如果约束条件是≤，就在左端+一个松弛变量，改为=
如果约束条件是≥，就在多端−一个剩余变量，改为=
如果有一个$x\leq 0$，就改成$x\geq 0$，在其他式子里这个变量都填上负号
如果有一个$x=0$，就新增两个变量，将其写作x_i-x_j，带入其余式子中

单纯形算法步骤

1.初始单纯形表：$\begin{pmatrix} A & b \\ \sigma & f_0 \end{pmatrix}$
其中判别数$\sigma_i=C_l^{\mathrm{T}}P_j-c_j$，基变量判别数始终为 0。初始值$f_0=C_l^{\mathrm{T}}b$
2.找到最大的判别数σ_l
3.若$\sigma_l\leq 0$，则 X 是最优解。
4.若$P_l\leq 0$（即第l列没有正数），则无最优解。
5.找到最小的$\frac{b_k}{a_{kl}}$（最后一列除以这一列，找最小值）
6.以a_{kl}为主元做换基运算，转 2

两阶段方法步骤

引入人工变量$y_1,y_2,\ldots,y_m$，添加到每个等式约束中，人为地造出来 m 个基变量
辅助线性规划：
$\min g(Y)=y_1+y_2+\ldots+y_m$
$s.t.$
$y_1+a_{11}x_1+\ldots+a_{1n}x_n=b_1$
$y_2+a_{21}x_1+\ldots+a_{2n}x_n=b_2$
$\ldots$
$y_m+a_{m1}x_1+\ldots+a_{mn}x_n=b_m$
求解出最优值g^*。
若$g^*>0$，则原规划无可行解。
若$g^*=0$，则$y_i^*=0$
如果此时基变量全在x_i，则将x_i拿出来其作为第二个阶段（也即是原规划）的初始单纯形表
如果存在基变量为y_i
1.如果该行的x的系数全为 0，则直接将这一行以及y_i所在列删掉。
2.如果存在一个x_i的系数不等于 0，则进行一次换基运算，x_i进基，y_i出基
最后基变量全在x_i，将x_i拿出来做第二个阶段

大 M 法步骤

同样引入人工变量y_i，只不过是利用一个充分大的常数迫使y_i出基。

构造规划（M 为充分大常数）

$\min g(Y)=c_1x_1+c_2x_2+\ldots+c_nx_n+M(y_1+y_2+\ldots+y_m)$

$s.t.$

$y_1+a_{11}x_1+\ldots+a_{1n}x_n=b_1$

$y_2+a_{21}x_1+\ldots+a_{2n}x_n=b_2$

$\ldots$

$y_m+a_{m1}x_1+\ldots+a_{mn}x_n=b_m$

若求解得$Y=0$，则X为原规划最优解

若求解得$Y\neq 0$或无最优解，则原规划无可行解

修正单纯形算法步骤

1.计算B^{-1}，$\pi=C_B^{\mathrm{T}}B^{-1}$（最开始$B^{-1}=I$单位矩阵）

2.计算$\sigma_j=\pi P_j-c_j$，如果所有$\sigma_j\leq 0$，则当前基可行解即为最优解

3.l为最左边的$\sigma_l>0$所在列，计算这一列$(a_{1l},\ldots,a_{ml})^{\mathrm{T}}=B^{-1}P_l$，如果没有$a_{il}>0$，则原规划无最优解

4.求出最后一列除以这一列$\frac{b_i}{a_{il}}$，找到$\frac{b_i}{a_{il}}>0$的且最小的行k（若有多个，取最上边的）（最后一列b是$B^{-1}b$）

5.形成初等阵$E_{kl}=\begin{pmatrix}1 & & & -\frac{a_{1l}}{a_{kl}} & & & \\ & \ddots & & \vdots & & & \\ & & 1 & -\frac{a_{k-1,l}}{a_{kl}} & & & \\ & & & \frac{1}{a_{kl}} & & & \\ & & & -\frac{a_{k+1,l}}{a_{kl}} & 1 & & \\ & & & \vdots & & \ddots & \\ & & & -\frac{a_{ml}}{a_{kl}} & & & 1\end{pmatrix}$

6.B^{-1}更新为$E_{kl}B^{-1}$，转到 1

三、对偶线性规划

对称形式

原规划：	对偶线性规划：
$\min C^{\mathrm{T}}X$	$\max b^{\mathrm{T}}W$
s.t.	s.t.
$AX \geq b$	$A^{\mathrm{T}}W \leq C$
$X \geq 0$	$W \geq 0$
（n 个变量，m 个约束）	（m 个变量，n 个约束）

混合形式

原规划（对偶规划）	对偶规划（原规划）
min 目标中系数 约束条件右端项	max 约束条件右端项 目标中系数
约束条件≥ 约束条件≤ 约束条件=	变量≥ 0 变量≤ 0 变量无约束
变量≥ 0 变量≤ 0 变量无约束	约束条件≤ 约束条件≥ 约束条件=

min 的约束条件和 max 的变量符号相同
min 的变量和 max 的约束条件符号相反

对偶定理

弱对偶性：$C^{\mathrm{T}}X \geq b^{\mathrm{T}}W$
推论（无界性）：原规划有无界解，则对偶规划无可行解
推论（最优性）：若$C^{\mathrm{T}}X = b^{\mathrm{T}}W$，则$X$和$W$分别为原规划和目的规划的最优解
即原规划与对偶规划的目标函数若能取相同值，则该值为最优值
强对偶性：原规划和目的规划其中一个有最优解，另一个必有最优解，且最优值相等
若B是原规划的最优基，那么$(C_B^{\mathrm{T}}B^{-1})^{\mathrm{T}}$就是对偶规划的最优解
对偶规划的最优解的第l个分量就是原规划的最终单纯形表中剩余变量$x_{n+l}(l = 1 \sim m)$的判别数σ_{n+l}的相反数
对偶约束：原规划的非平凡约束$A_iX \geq b_i$和对偶规划的平凡约束$y_i \geq 0$，或者原规划的平凡约束$x_j \geq 0$ 和对偶规划的非平凡约束$W^{\mathrm{T}}P_j \leq c_j$
松紧约束：如果线性规划的每一个最优解都使得某个约束取等号，就称这个约束是紧约束，否则称为松约束。
松约束的对偶约束是紧约束。
互补松弛条件：如果严格$A_iX > b_i$，则$y_i = 0$；如果严格$x_j > 0$，则$W^{\mathrm{T}}P_j = c_j$。

~~对偶单纯形方法~~

正则解：满足所有判别数非正的基本解（不要求可行）。其所对应的基称为正则基。
对偶单纯形方法思想：保持解的正则性，逐步将解移动到可行性区间内。满足可行性的正则解即为最优解。

1.构造初始对偶单纯形表$\begin{pmatrix} P_1 & P_2 & \cdots & P_n & b \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n & f_0 \end{pmatrix}$

2.若$b \geq 0$，则当前正则解就为最优解。否则，取k为第一个使得$b_k < 0$ 的行数。

3.若该行均有$a_{kj} \geq 0$，则原规划无可行解。

4.求出在所有$a_{kj} < 0$ 的列j中，使得$\frac{\sigma_j}{a_{kj}}$最小的列l。

5.换基运算：第l列主元a_{kl}进基。

四、无约束最优化计算方法

~~下降迭代算法~~

一般格式：

确定初始迭代点X^0，$k=0$

确定下山方向P^k，使得$\nabla f(X^0)^T P^k < 0$

确定最优步长t_k，使得 $f(X^k + t_k P^k) < f(X^k)$

计算$X^{k+1} = X^k + t_k P^k$

判断能否满足终止准则，满足则终止，否则转第二步

一维搜索：

求解 $\min \varphi(t) = f(X^k + tP^k)$

收敛速度：

$\lim\limits_{k\to\infty} \frac{\|X^{k+1}-X^*\|}{\|X^k-X^*\|} = \beta$，当 $0<\beta<1$ 时称$\{X^k\}$为线性收敛，$\beta=0$ 称为超线性收敛，$\beta=1$ 称为次线性收敛

$\lim\limits_{k\to\infty} \frac{\|X^{k+1}-X^*\|}{\|X^k-X^*\|^p} = \beta < +\infty$称为 p 阶收敛，$p \ge 1$

二次终止性：

算法在求解正定二次函数时可以在有限步内找到极小点。

精确一维搜索：

黄金分割法：

1.确定初始搜索区间$[a,b]$

2.计算$t_2 = a + \beta(b-a)$，$\varphi_2 = \varphi(t_2)$

3.计算$t_1 = a + \alpha(b-a)$，$\varphi_1 = \varphi(t_1)$

4.若$t_1 - t_2 < \varepsilon$，算法终止

5.若$\varphi(t_1) \le \varphi(t_2)$，则 $b = t_2$，$t_2 = t_1$，$t_1 = a + \alpha(b-a)$，$\varphi_1 = \varphi(t_1)$

6.否则$a = t_1$，$t_1 = t_2$，$t_2 = a + \beta(b-a)$，$\varphi_2 = \varphi(t_2)$，转步骤 4

Fibonacci 法：

$t_1 = a + \frac{L_{n-2}}{L_n}(b-a)$，$t_2 = a + \frac{L_{n-1}}{L_n}(b-a)$（其中$L_0 = L_1 = 1$），后续过程同黄金分割法

~~三点二次插值法：~~

选取三个初始点t_1、t_2、t_3，要求三点处的函数值φ_1、φ_2、φ_3必须φ_2最小

通过三个点作一条抛物线 P(t)，近似曲线φ(t)

找到 P(t)的极小点μ，在四个点中再次找出相邻且两边高的三点，如此反复。

~~两点三次插值法：~~

选取两个初始点t_1、t_2，要求导数值$\varphi'_1 < 0 < \varphi'_2$。

用两个函数值、两个导数值，算出待定三次多项式的系数

找三次近似的极小点μ

非精确一维搜索：

~~Goldstein 准则：~~

$\begin{cases} f(X^k + t_k P^k) \le f(X^k) + \rho t_k \nabla f(X^k)^T P^k \\ f(X^k + t_k P^k) \ge f(X^k) + (1-\rho) t_k \nabla f(X^k)^T P^k \end{cases}$，$0 < \rho < \frac{1}{2}$

（把下个迭代点限定在两条直线的夹角范围内）

~~Wolfe 准则：~~

$\begin{cases} f(X^k + t_k P^k) \le f(X^k) + \rho t_k \nabla f(X^k)^T P^k \\ \nabla f(X^k + t_k P^k)^T P^k \ge \sigma \nabla f(X^k)^T P^k \end{cases}$，即$\varphi'(t_k) \ge \sigma\varphi'(0)$

（在下一个迭代点的切线斜率大于当前斜率的σ倍，因为当前斜率$\varphi'(0) < 0$，$\rho < \sigma < 1$，故下个迭代点的切线更平坦）

Armijo 准则：

$f(X^k + \beta^{m_k}\tau P^k) \le f(X^k) + \rho\beta^{m_k}\tau\nabla f(X^k)^T P^k$，$m_k$是满足其成立的最小非负整数

非精确一维搜索一般算法：

1.给出初始点X^0，允许误差 $0 \le \varepsilon < 1$，$k = 0$
2.若$\|\nabla f(X^k)^T\| \le \varepsilon$，算法停止，否则求出下降方向$P^k$
3.求出满足准则的步长t_k
4.$X^{k+1} = X^k + t_k P^k$，$++k$，转步骤 2

最速下降法：

沿$P^k = -\nabla f(X^k)$方向搜索
在有限步终止到驻点，或得到的点列的极限点为驻点
算法：
1.确定初始迭代点X^0，$g^0 = \nabla f(X^0)$，k = 0
2.一维搜索确定最优步长$t_k = arg\min f(X^k - t\nabla f(X^k))$，$X^{k+1} = X^k - t_k\nabla f(X^k)$，计算$f^{k+1}$，$g^{k+1} = f(X^{k+1})$
3.判断能否满足$\|g^{k+1}\| \le \varepsilon$，满足则终止，否则转 2

当目标函数是正定二次函数时，接近极小点时会呈锯齿状搜索，影响收敛速度

牛顿法：

对$f(X)$进行二阶近似，在X^k二阶展开

$$f(X) \approx q(X) = f(X^k) + g(X^k)^T(X - X^k) + \frac{1}{2}(X - X^k)^T H(X^k)(X - X^k)$$

$g(X^k)$和$H(X^k)$分别是X^k处的梯度和 Hesse 矩阵
算法：
1.给定目标函数$f(X)$，梯度$g(X)$，Hesse 矩阵$H(X)$，精度ε
2.选定初始点X^0，计算$f^0 = f(X^0)$，$g^0 = g(X^0)$，置$k = 0$
3.计算$H^k = H(X^k)$
4.搜索方向$P^k = -(H^k)^{-1}g^k$（考虑到求逆的计算量大，可以用$H^kP^k = -g^k$求解）
5.计算$X^{k+1} = X^k + P^k$，$f^{k+1} = f(X^{k+1})$，$g^{k+1} = g(X^{k+1})$
6.判断是否满足终止条件，若不满足，转 3

若初始迭代点位于最优解的邻域内，则牛顿法具有二阶收敛速度
牛顿法可视为椭球范数下的最速下降算法

共轭方向法：

共轭：$X^TQY = 0$，称为 X 和 Y 关于 Q 共轭（正交的推广）
如果非零向量组两两共轭，称为一组向量共轭
扩张子空间定理：
目标函数为正定二次函数，其 Hesse 矩阵为 Q，如果搜索方向P^0，P^1，...，P^{n-1}为 Q 共轭，那么从任意一点X^0出发，至多经过 n 次迭代，必然收敛至最优解X^*。

Gram-Schmidt 共轭化方法：

设v^i，$i = 0 \sim n-1$ 是无关向量组，首先取$P^0 = v^0$。

设已求得P^i，$i = 0 \sim k$，则$P^{k+1} = v^{k+1} - \sum_{j=0}^{k} \frac{(P^j)^TQv^{k+1}}{(P^j)^TQP^j} P^j$

共轭梯度法（FR）：

首先构造$P^0 = -g^0$

对$k = 1 \sim n-1$，令$P^{k+1} = -g^{k+1} + \alpha_k P^k$。可以取$\alpha_k^{FR} = \frac{\|g^{k+1}\|^2}{\|g^k\|^2}$

算法：

1.目标函数$f(X)$，初始点X^0，梯度g^0，精度ε，置$k=0$
2.若满足精度，输出X^k
3.计算$P^k=\begin{cases} -g^k & ，k=0 \\ -g^k+\frac{\|g^k\|^2}{\|g^{k-1}\|^2}P^{k-1} & ，k>0 \end{cases}$
4.计算$X^{k+1}=X^k+t_kP^k$，g^{k+1}，$++k$，转步骤 2

~~拟牛顿法：~~

求 Hesse 矩阵的逆计算量很大，所以用变尺度矩阵近似 Hesse 矩阵的逆

对称秩一算法：

1.给定目标函数$f(X)$，初始点X^0，变尺度矩阵$H^0=I$，梯度g^0，搜索方向$P^0=-g^0$，精度ε，置$k=0$
2.计算$X^{k+1}=X^k+t^kP^k$，g^{k+1}
3.判断是否满足终止条件，若是，输出X^{k+1}
4.计算$S^k=X^{k+1}-X^k$，$Y^k=g^{k+1}-g^k$，$Z^k=S^k-H^kY^k$
5.计算$H^{k+1}=H^k+\frac{Z^k(Z^k)^T}{(Y^k)^TZ^k}$
6.计算$P^{k+1}=-H^{k+1}g^{k+1}$，$++k$，转步骤 2

优点：
具有简单、对称迭代形式；
满足拟牛顿条件。
缺点：可能出现$(Y^k)^TZ^k=0$，数值不稳定；
不能保证$P^{k+1}=-H^{k+1}g^{k+1}$是下山方向，H^k不正定。

对称秩二算法（DFP 算法）：

把对称秩一算法步骤 5 的公式改为$H^{k+1}=H^k+\frac{S^k(S^k)^T}{(S^k)^TY^k}-\frac{H^kY^k(H^kY^k)^T}{(H^kY^k)^TY^k}$

BFGS 算法：

数值稳定性更好

信赖域方法：

1.给出初始点X^0，初始半径h^0，置$k=0$
2.给出X^k，h^k，计算梯度g^k和 Hesse 矩阵G^k
3.解信赖域模型（$\min q^k(S)=f(X^k)+(g^k)^TS+\frac{1}{2}S^TG^kS$，$s.t.\|S\|\le h^k$，即$f(X)$在$X^k$的二阶泰勒展开的前两项，约束条件为半径小于等于$h^k$），求出$S^k$
4.求$f(X^k+S^k)$和$r^k=\frac{\Delta f^k}{\Delta q^k}=\frac{f(X^k)-f(X^k+S^k)}{f(X^k)\text{（或}q^k(0)\text{）}-q^k(S^k)}=\frac{\text{目标函数在两个迭代点的差值}}{\text{二阶近似在两个迭代点的差值}}$（$r^k$衡量二次模型近似的程度，越接近 1 说明近似程度越好）
5.(1)$r^k<0.25$（近似效果不好）：$h^{k+1}=\frac{\|S^k\|}{4}$（信赖域过大，往往有$S^k<h^k$）
(2)$r^k>0.75$ && $S^k=h^k$（近似效果很好，并且信赖域不是很大）：$h^{k+1}=2h^k$（可以再贪一点）
(3)例外情况（一般情况）：$h^{k+1}=h^k$（保持不变）
6.若$r^k\le 0$，则$X^{k+1}=X^k$（$q^k(0)>q^k(S^k)$一定成立，说明$f(X^k)\le f(X^k+S^k)$，那就维持当前点）；否则$X^{k+1}=X^k+S^k$

五、约束最优化方法

可行方向：在这个方向上走任意小的一步，不会走出可行区域。
下降方向：在这个方向上走任意小的一步，能保证目标函数下降。
积极约束：不等式约束取等号
线性化可行方向$\mathcal{LF}$：与所有积极约束的梯度成锐角或直角，与所有等式约束的梯度成直角
序列可行方向$\mathcal{SF}$：在 P 方向上走 t 长度，不会走出可行区域

$$\{P_{可行}\} \subset \mathcal{LF} \subset \mathcal{SF}$$

稳定点：可行区域为凸集，从当前点指向所有可行区域内的点，其方向都与这一点的梯度成锐角或直角（都不是下降方向），这个点就是局部最优解。
正则点：所有积极约束和等式约束的梯度线性无关。

一阶必要条件（KKT 条件）：

如果X^*是局部极小点且为正则点，则存在下方成立：

$$\begin{cases} \nabla f(X^*) = \sum\limits_{i\in I(X^*)} \lambda_i^* \nabla g_i(X^*) + \sum\limits_{j=1}^{l} \mu_j^* \nabla h_j(X^*) & \text{（线性）} \\ \lambda_i^* \geq 0，i = 1,2,\ldots,m & \text{（非负性）} \\ \lambda_i^* g_i(X^*) = 0，i = 1,2,\ldots,m & \text{（互补性）} \\ g_i(X^*) \geq 0，i = 1,2,\ldots,m & \text{（可行性）} \\ h_j(X^*) = 0，j = 1,2,\ldots,l & \end{cases}$$

线性：X^*处$f(X)$的梯度可以表示为所有积极约束和等式约束梯度的线性组合
非负性：所有不等式约束的拉格朗日因子均≥ 0
互补性：$\lambda_i^* g_i(X^*) = 0$，也即是说对于所有非积极约束（$g_i(X^*) \neq 0$），必有$\lambda_i^* = 0$
可行性：满足可行区域

KKT 点（KT 点）：满足 KKT 条件的点

拉格朗日（Lagrange）函数：$L(X,\lambda,\mu) = f(X) - \sum_{\forall i} \lambda_i^* g_i(X) - \sum_{\forall j} \mu_j^* h_j(X)$

二阶充分条件：

若对$\forall Z(\neq 0) \in \mathcal{F}$，有$Z^T(\nabla^2 f(X^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla^2 h_j(X^*))Z > 0$ 成立（$Z^T \nabla_*^2 L(X,\lambda,\mu) Z > 0$），则$X^*$为严格局部极小点。其中

$$\mathcal{F} \overset{\text{def}}{=} \left\{ Z \in R^n \middle| \begin{array}{ll} Z^T \nabla g_i(X^*) = 0 & 积极约束且\lambda_i^* > 0 \\ Z^T \nabla g_i(X^*) \geq 0 & 积极约束且\lambda_i^* = 0 \\ Z^T \nabla h_j(X^*) = 0 & j = 1,2,\ldots,l \end{array} \right\}$$

表示所有与$\lambda_i^* > 0$ 的积极约束和等式约束的梯度垂直，与所有$\lambda_i^* = 0$ 积极约束的梯度成锐角或直角的向量。

一阶必要条件+二阶充分条件求约束优化问题最优解：

①设X^*为 KKT 点，利用 KKT 条件求出X^*，λ^*，μ^*
如果是凸优化，KKT 点即为最优解（不用二阶充分条件判定）
②求解$L(X,\lambda,\mu)$在X^*的 Hesse 矩阵$\nabla_*^2 L(X^*,\lambda^*,\mu^*)$。如果正定，则满足二阶充分条件。
③如果非正定，则求$\mathcal{F}$，判断$Z^T \nabla_*^2 L(X^*,\lambda^*,\mu^*) Z > 0$ 是否成立。若成立，则为最优解。

外点罚函数法：

对于 $\min f(X)$，$\text{s.t.}\ g_i(X) \geq 0$，$h_j(X) = 0$，构造惩罚函数

$$P(X, m_k) = f(X) + m_k \left(\sum_{i=1}^{m} (\min\{g_i(X), 0\})^2 + \sum_{j=1}^{l} (h_j(X))^2 \right)$$

令惩罚项$\alpha(X) = \sum_{i=1}^{m} (\min\{g_i(X), 0\})^2 + \sum_{j=1}^{l} (h_j(X))^2$，可写成$P(X, m_k) = f(X) + m_k \alpha(X)$

算法：

1.选定初始点X^0，初始惩罚因子$m_1 > 0(m_1 = 1)$，惩罚因子增长率$C > 1(C = 10)$，置$k = 1$
2.以X^{k-1}为初始点，求解无约束优化$P(X, m_k)$的最优解X^k
3.若$m_k \alpha(X^k) < \varepsilon$（满足条件），则$X^k$为约束优化问题的最优解，否则转 4
4.令$m_{k+1} = Cm_k$，$++k$，转 2（惩罚力度越来越大）

解析算法：把m当做常数，求解无约束优化$P(X, m)$的最优解X^m（驻点），令$m \to \infty$，得到X^*

收敛性定理：
随着迭代次数增加，无约束优化问题的最优值上升，惩罚函数下降，原目标上升
求出的迭代点列收敛至约束优化问题的极小点

内点罚函数法：

对于 $\min f(X)$，$\text{s.t.}\ g_i(X) \geq 0$（只能处理不等式约束），构造障碍函数

$$P(X, r_k) = f(X) + r_k \sum_{i=1}^{m} \frac{1}{g_i(X)}$$

令惩罚项$\beta(X) = \sum_{i=1}^{m} \frac{1}{g_i(X)}$。也可以构造对数罚函数$P(X, r_k) = f(X) - r_k \sum_{i=1}^{m} \ln g_i(X)$。

算法：

1.选定初始点$X^0 \in S^0$（初始点必须在可行区域内），初始惩罚因子$r_1 > 0(r_1 = 10)$，惩罚因子减小率$C < 1(C = 0.1)$，置$k = 1$
2.以X^{k-1}为初始点，求解障碍函数$P(X, r_k)$的最优解X^k。
3.若$r_k \beta(X^k) < \varepsilon$（满足条件），则$X^k$为约束优化问题的最优解，否则转 4
4.令$r_{k+1} = Cr_k$，$++k$，转 2（惩罚力度越来越小，使得X^k更容易收敛到边界）

解析算法：把r当做常数，求解无约束优化$P(X, r)$的最优解X^r（驻点），令$r \to 0$，得到X^*

收敛性定理：随着迭代次数增加，无约束优化（障碍函数）的最优解下降。求出的迭代点列收敛至约束优化问题的极小点。

优缺点：外点法能处理所有约束，但在有限步结束时得到的解往往不满足可行性。
内点法得到的点都在可行区域内，但只能处理不等式约束。

~~乘子罚函数法：~~

（**如果只有等式约束**）对于 $\min f(X)$，$\text{s.t.}\ h_j(X) = 0$，构造乘子罚函数（增广 Lagrange 函数）

$$\phi(X, V, \sigma) = f(X) - \sum_{j=1}^{l} v_j h_j(X) + \frac{\sigma}{2} \sum_{j=1}^{l} h_j^2(X)$$

解析算法：

给定σ满足充分大，求解无约束优化$\phi(X, V^k, \sigma)$的最优解X^k（含有参数$V^k = (v_1^k, v_2^k, \ldots, v_n^k)^T$）

修正公式$v_j^{k+1} = v_j^k - \sigma h_j(X^k)$，$j = 1,2, \ldots, l$，求出$v_j^* = \lim\limits_{k \to \infty} v_j^k$，代入求出$X^*$

（**考虑不等式约束**）对于 $\min f(X)$，$\text{s.t.}\ g_i(X) \ge 0$
引入变量y_i，问题转变为 $\min f(X)$，$\text{s.t.}\ g_i(X) - y_i^2 = 0$
依照等式约束处理，构造乘子罚函数

$$\phi(X,Y,W,\sigma) = f(X) - \sum_{i=1}^{m} w_i\left(g_i(X) - y_i^2\right) + \frac{\sigma}{2}\sum_{i=1}^{m}\left(g_i(X) - y_i^2\right)^2 = f(X) + \sum_{i=1}^{m}\left(\frac{\sigma}{2}\left(g_i(X) - y_i^2\right)^2 - w_i\left(g_i(X) - y_i^2\right)\right)$$

$$= f(X) + \sum_{i=1}^{m}\left(\frac{\sigma}{2}\left(y_i^2 - \frac{1}{\sigma}(\sigma g_i(X) - w_i)\right)^2 - \frac{w_i^2}{2\sigma}\right)\text{（配方）} = f(X) + \sum_{i=1}^{m}\left(\frac{\sigma}{2}\left(y_i^2 - \frac{1}{\sigma}(\sigma g_i(X) - w_i)\right)^2 - \frac{w_i^2}{2\sigma}\right)$$

因为 $\min\left(y_i^2 - \frac{1}{\sigma}(\sigma g_i(X) - w_i)\right)^2 = \frac{1}{\sigma^2}(\max\{0, w_i - \sigma g_i(X)\})^2$（当$\frac{1}{\sigma}(\sigma g_i(X) - w_i) > 0$ 时取$y_i^2 = \frac{1}{\sigma}(\sigma g_i(X) - w_i)$，有上式$= 0$；否则取$y_i^2 = 0$，有上式$= (w_i - \sigma g_i(X))^2$），所以上式可将$y_i^2$消掉，得

$$\phi(X,W,\sigma) = f(X) + \sum_{i=1}^{m}\left(\frac{\sigma}{2}\frac{1}{\sigma^2}(\max\{0, w_i - \sigma g_i(X)\})^2 - \frac{w_i^2}{2\sigma}\right) = f(X) + \frac{1}{2\sigma}\sum_{i=1}^{m}\left((\max\{0, w_i - \sigma g_i(X)\})^2 - w_i^2\right)$$

解析算法：
给定σ满足充分大，求解无约束优化$\phi(X, W^k, \sigma)$的最优解X^k（含有参数$W^k = (w_1^k, w_2^k, \dots, w_n^k)^T$）
修正公式$w_i^{k+1} = \max\left\{0, w_i^k - \sigma g_i(X^k)\right\}$，$i = 1,2,\dots,m$，求出$w_i^* = \lim\limits_{k\to\infty} w_i^k$，代入求出$X^*$

~~一般情形（有等式和不等式约束）：~~

对于 $\min f(X)$，$\text{s.t.}\ g_i(X) \ge 0$， $h_j(X) = 0$，构造乘子罚函数

$$\phi(X,W,V,\sigma) = f(X) + \frac{1}{2\sigma}\sum_{i=1}^{m}\left((\max\{0, w_i - \sigma g_i(X)\})^2 - w_i^2\right) - \sum_{j=1}^{l} v_j h_j(X) + \frac{\sigma}{2}\sum_{j=1}^{l} h_j^2(X)$$

求解无约束优化$\phi(X,W,V,\sigma)$的最优解X^k
（含有参数$W^k = (w_1^k, w_2^k, \dots, w_n^k)^T$，$V^k = (v_1^k, v_2^k, \dots, v_n^k)^T$）

修正公式有$\begin{cases} w_i^{k+1} = \max\left\{0, w_i^k - \sigma g_i(X^k)\right\} \\ v_j^{k+1} = v_j^k - \sigma h_j(X^k) \end{cases}$，求出$w_i^* = \lim\limits_{k\to\infty} w_i^k$，$v_j^* = \lim\limits_{k\to\infty} v_j^k$，代入求出$X^*$

算法：
1.给定初始点X^0，乘子向量估计W^1，V^1，参数σ，允许误差$\varepsilon > 0$，参数增长率$C > 1$，判断因子 $0 < \beta < 1$，置$k = 1$
2.求解 $\min \phi(X, W^k, V^k, \sigma)$，得到$X^k$。
3.如果惩罚项$\left\|\alpha(X^k)\right\| < \varepsilon$，则停机，输出$X^k$；
如果两次迭代惩罚项的比值$\frac{\|\alpha(X^k)\|}{\|\alpha(X^{k-1})\|} \ge \beta$，说明$\sigma$不满足充分大，$\sigma = C\sigma$
4.通过公式修正W^k和V^k，$++k$，转 2

Rosen 梯度投影法

求解线性约束的非线性规划
基本思想：若当前迭代点的负梯度方向不是可行方向，则将负梯度方向投影到积极约束的法向量张成的子空间的正交补子空间中（即积极约束的交线上），使下降方向满足可行性。
投影矩阵定义：满足对称性、幂等性的矩阵。
性质：
半正定
$Q = I - P$也是投影矩阵（且$PQ = 0$）
任意向量可以表示为$X = p + q$，其中$p \in \{PX | X \in R^n\}$（P的列空间），$q \in \{QX | X \in R^n\}$（Q的列空间）

算法：

1.选定X^0，置$k=0$

2.令$N_k=\begin{pmatrix}A'_k\\C_k\end{pmatrix}$（其中$A'_k$是积极约束的梯度（也即线性方程的系数），$C_k$是等式约束的梯度）

3.记$r=rand(N_k)$，若$r>0$，计算$Q_k=I-N_k^T(N_kN_k^T)^{-1}N_k$，转 5。

4.若$r>0$，计算$Q_k=I$。

5.计算$P^k=-Q_k\nabla f(X^k)$

6.若$\|P^k\|<\varepsilon$（数值计算时$P^k=0$），转 7；否则转 9

7.若$r=0$，则X^k是最优解；否则：

计算$q_k=\begin{pmatrix}y\\z\end{pmatrix}=(N_kN_k^T)^{-1}N_kg_k$

8.若$y\geq 0$，则X^k是最优解（KKT 点）；若$\exists y_i<0$，则去掉N_k中相应的行向量，转 2

9.根据非积极约束确定搜索步长上限：令$\begin{matrix}u=A''_kX^k-b''_k\\v=A''_kP^k\end{matrix}$，则

$\bar{t_k}=\begin{cases}+\infty, & v\geq 0\\ \min\limits_{1\leq i\leq\theta}\left\{-\frac{u_i}{v_i}\middle|v_i<0\right\}, & \exists v_i<0\end{cases}$，求解一维搜索$\min\limits_{0\leq t\leq\bar{t_k}}f(X^k+tP^k)$，$X^{k+1}=X^k+tP^k$，$++k$，转 2