Finally Report

name: Ensheng Shi (石 恩 升)
student ID number: 4119105089

2019/ 11/ 18/

1 Mathematics foundation and Key technique

优化问题的基本形式为:

$$\begin{aligned} &min \quad f(x) \\ st.\ &h_i(x) = 0, i \in E = 1, 2, ...m_e \\ &c_i(x) \geq 0, i \in I = m_e + 1, , ...m \end{aligned} \tag{1}$$

$x \in R_n$,h_i, c_i $R_n -> R_1$
f(x) 为优化目标函数，h(x) 为等式约束，c(x) 为不等式约束。为了解决这个问题，我们需要知道一些关于函数，变量等的数学基础:
1. 函数的一阶导数与二阶导数
2. 范数
3. 泰勒展开
因为求解的目的是达到 f(x) 的极值点，我们需要知道函数极值以及达到极值的条件。
1. 多元函数的极值
2. KKT 条件 (考虑不等式约束)
因为是多元函数，其中涉及矩阵的概念和矩阵的运算
1. 矩阵的正定性
2. 向量的正交化
3. 矩阵的分解
由于本课程多以凸优化为例，还需要介绍
1. 凸集
2. 凸函数
3. 凸优化。
最后由于算法需要进行收敛性分析，简单介绍收敛速度的判断

1.1 数学基础

1.1.1 函数的一阶导数与二阶导数

对于 n 元函数 f(x),$x = (x1, x2, x3, ...xn)_T \in R^n$, 梯度向量为:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ ... \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} \tag{2}$$

有时记做 g(x)
二阶偏导数矩阵，也称 Hesse 阵。

$$\nabla f(x)=\begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} \tag{3}$$

有时记作 G(x)

1.1.2 范数

范数用来衡量 n 维空间中两点的距离，这里指的是向量的范数。满足：1. 非负性：对于任意的 $x \in R^n, \|x\| \geq 0$, 当且仅当 x = 0 时，$\|x\|= 0$; 2. 齐次性：对于任意的 $x \in R^n, \alpha \in R$, $\|\alpha * x\| = |\alpha| * \|x\|$ 3. 三角不等式: 对于任意的 $x, y \in R^n$, $\|x+y\| \leq \|x\| + \|y\|$

1.1.3 泰勒展开

一元函数的 Tayor 展开式为：

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

多元函数的 Tayor 展开式为：

$$f(\boldsymbol{x})=f(\boldsymbol{x}_0)+\nabla f(\boldsymbol{x}_0)......T(\boldsymbol{x}-\boldsymbol{x}_0)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T\nabla^2 f(\boldsymbol{x}_0)^T(\boldsymbol{x}-x_0)+...+o(\|\boldsymbol{x}-\boldsymbol{x}_0\|^2)$$

在一维搜索中，我们需要构造一元函数：

$$\begin{aligned} &\Phi(t)=f(\boldsymbol{x}_0+t\boldsymbol{d})=f(x1_0+td_1,\cdots,x_n^0+td_n) \\ &\Phi(t)'=\nabla f(\boldsymbol{x}_0+t\boldsymbol{d})^T\boldsymbol{d} \\ &\Phi(t)'=d^T\nabla^2 f(\boldsymbol{x}_0+t\boldsymbol{d})^T\boldsymbol{d} \end{aligned} \tag{4}$$

因为在迭代过程中 $\Phi(t)'=\nabla f(\boldsymbol{x}_0+t\boldsymbol{d})^T\boldsymbol{d}<0$ 所以以为搜索的函数大致形状如图1所示：

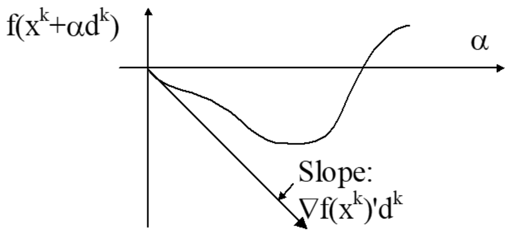


Figure 1: 一维搜索函数图

1.2 极值条件

1.2.1 多元函数的极值

多元函数中，当 $\nabla f(x) = 0$ 时
(1) 若 $nabla^2 f(x_0)$ 负定, 则 f(x) 在 x0 处取得极大值;
(2) 若 $nabla^2 f(x_0)$ 正定, 则 f(x) 在 x0 处取得极小值;
(3) 若 $nabla^2 f(x_0)$ 不定, 则无法判断。

1.2.2 优化问题的一阶必要条件–KKT 条件

定理 (kuHn - Tucker 定理) 设 x* 是约束优化问题的局部极小解，f(x) 在 x* 处可微，当约束规范条件

$$SFD(x*, D) = LFD(x*, D)$$

成立时，存在实数 $\lambda_i^*(i = 1, 2, \cdots, m)$ S.t.

$$\begin{cases} \nabla f(x*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0 \\ \lambda_i^* \nabla c_i(x^*) = 0, \quad \lambda_i^* \geq 0, i \in I \end{cases} \tag{5}$$

互补松弛条件 $\lambda_i^* \nabla c_i(x^*) = 0$ 是为了让拉格朗日乘子法成立, 对于 λ 的非负性要求是在 farkas 引理里引入的。
证明的思路为：farkas lemma：

$$1. \begin{cases} Ax \leq 0 \\ c^T x \geq 0 \end{cases} \qquad 2. \begin{cases} A^T y \leq c \\ y > 0 \end{cases}$$

公式一和公式二 iif 一个有解。约束规范条件 $SFD(x*, D) = LFD(x*, D)$ 可得

$$\nabla f(x_*)^T d \geq 0 \quad \forall d \in LFD(x^*, D)$$

根据 LFD(x*,D) 定义可以得到方程组：

$$\begin{cases} \nabla c_i(x^*)^T d = 0 & i \in E \\ \nabla c_i(x^*)^T d = 0 & i \in I^* \\ \nabla f(x^*)^T d < 0 \end{cases} \text{无解，也就是} \begin{cases} Ad \leq 0 \\ \nabla f(x^*)^T d > 0 \end{cases} \quad A = [\nabla c_i(x^*)^T, -\nabla c_i(x^*)^T, \quad i \in E \quad \nabla - c \tag{6}$$

根据 farkas lemma,

$$\begin{cases} A^T y \leq f(x^*) \\ y > 0 \end{cases} \tag{7}$$

有解，其中 $y = (\mu_{*-T}, \mu_{*+T}, \omega_{*-T})^T$ 也就是部分 λ, 加上互补松弛条件刚好使得拉格朗日乘子法的梯度等于零。构成了一阶必要条件。

定理 2(二阶充分条件)

(1) 当 x 为 KKT 点，

(2) 对于 $\forall d \in M = \{d \in R_n \quad | \quad d^T \nabla c_i(x^*) = 0, i \in E \cup I^*\}$ 都有

$$d^T \nabla_x^2 L(x^*, \lambda^*) d > 0$$

则 x* 为严格局部最优解。

1.3 线性代数基础

1.3.1 矩阵的正定性

矩阵的正定义的定义：$x^T Ax > 0$,A 为正定矩阵；

$x^T Ax \geq 0$,A 为半正定矩阵;

$x^T Ax < 0$,A 为负定矩阵;

1.3.2 向量的正交化

设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 n 维欧式空间 V 的一个基，若令：

$$\beta_1 = \alpha_1,$$

$$\beta_2 = \alpha - \frac{<\alpha_2, \beta_1>}{<\beta_1, \beta_1>}\beta_1 \quad \cdots\cdots \beta_n - \frac{<\alpha_n, \beta_1>}{<\beta_1, \beta_1>}\beta_1 - \frac{<\alpha_n, \beta_2>}{<\beta_2\beta_2>}\beta_2 - \cdots - \frac{<\alpha_n, \beta_{n-1}>}{<\beta_{n-1}, \beta_{n-1}>}\beta_{n-1} \tag{8}$$

则 $\beta_1, \beta_2, .., \beta_n$ 就是 V 的一个正交基，若在令

$$e_i = \frac{\beta_i}{\|\beta_i\|}$$

就得到了 V 的标准正交基 $e_1, e_2, .., e_n$.

1.3.3 矩阵的分解

1. LU 分解

如果顺序主子式不等于 0，再算 uii 时，其分子是主子式，算 lji 时分母为 uii。

对于矩阵

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = LU \tag{9}$$

其中 LU 分别为:

$$L=\begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} U=\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & \ddots & & \vdots \\ & & & u_{nn} \end{pmatrix} \tag{10}$$

s.t $A=LU$ Then Ax= b => Ly = b, Ux = y.

2. 半正定矩阵的 LDL^T 分解

对于矩阵 A

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1i} \\ a_{21} & a_{22} & \cdots & a_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = LU = \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & \ddots & & \vdots \\ & & & u_{nn} \end{pmatrix} \tag{11}$$

$$= \begin{pmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} m_{11} & & \\ & m_{22} & \\ & \ddots & \\ & & m_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & \cdots & u_{2n} \\ & \ddots & & \vdots \\ & & & 1 \end{pmatrix}$$

if $A^T=A$ then, M = L, that is $A=LDL^T$

如果 A 时半正定的

$$A=LDL^T=LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T=GG^T$$

3. QR 分解 QR 分解是把矩阵分解成一个正交矩阵与一个上三角矩阵的积。QR 分解经常用来解线性最小二乘法问题. 课上的,QR 分解Householder

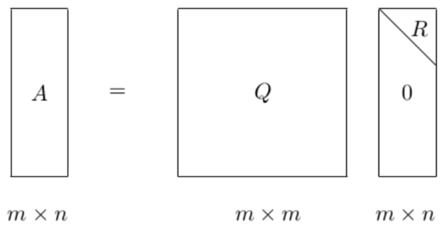


Figure 2: QR 分解示意图

变换，其次还有 Gram-Schmidt 正交化等方法。其中，Q 是一个标准正交方阵，R 是上三角矩阵。

假设 A 是一个 5×4 的矩阵，用 × 号表示本次变换未变化的元素，用 + 号表示本次发生变换的元素，H 矩阵等效于对右侧的 A 矩阵进行行操作 四次

$$H_1 A = H_1 \begin{pmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{pmatrix} = \begin{pmatrix} + & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \\ 0 & + & + & + \end{pmatrix}$$

$$H_2 \begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{pmatrix} = \begin{pmatrix} \times & \times & \times & \times \\ 0 & + & + & + \\ 0 & 0 & + & + \\ 0 & 0 & + & + \\ 0 & 0 & + & + \end{pmatrix}.$$

$$H_3 \begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{pmatrix} = \begin{pmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & + & + \\ 0 & 0 & 0 & + \\ 0 & 0 & 0 & + \end{pmatrix}$$

Figure 3: QR 分解过程示意图

变换之后, A 就转化成一个上三角矩阵。并且如果 A 是列向量不相关，则 R 矩阵是非奇异矩阵。$Q^T A = (R, 0)^T, \quad Q^T = H_4 H_3 H_2 H_1$

1.4 凸优化

1.4.1 凸集

集合 $X, x \in R^n$ is convex iff

$$\alpha x_1 + (1-\alpha) x_2 \in X, \quad \forall x_1, x_2 \in X, \quad \forall \alpha \in [0, 1]$$

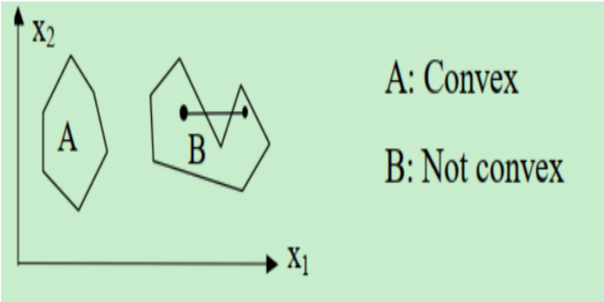


Figure 4: convex set

1.4.2 凸函数

X a convex set in r^n. f(x): X -> R is convex iff

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \quad , \quad \forall x_1, x_2 \in X, \forall \alpha \in [0,1]$$

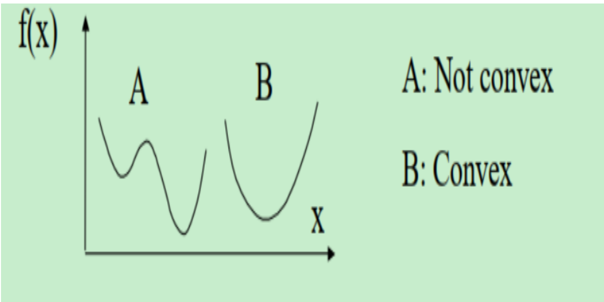


Figure 5: Convex function

1.4.3 凸优化

For a convex function f(x) over a convex set X is convex programming.

1.5 收敛速度

1. 假设x^k 收敛于 x*
2. error function $= \|x - x^*\|$ or $f(x) - f(x^*)$
3. 以二次型为例，因为二次型的结果已知，而且模型复杂度适中。
此时 X* =0 $e(x^k) \leq q\beta^k$ A geometric sequence $\frac{e(x^{k+1})}{e(x^k)} = \frac{q\beta^{k+1}}{q\beta^k} = \beta$
Linear: $\lim\limits_{k\to\infty} sup \frac{e(x^{k+1})}{e(x^k)} \leq \beta$

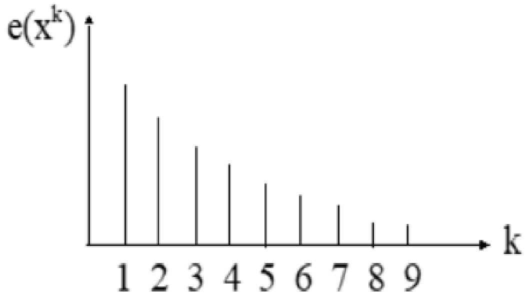


Figure 6: Linear or geometrical convergence

Super Linear: $\lim\limits_{k\to\infty} sup\frac{e(x^{k+1})}{e(x^k)} = 0$

Quadratic convergence: $e(x^k) \leq q(\beta)^{2^k} \quad \lim\limits_{k\to\infty} sup\frac{e(x^{k+1})}{e(x^k)^2} < \infty$

2 Method

一维搜索中：精确搜索和非精确搜索
无约束优化问题：最速下降法，牛顿法和拟牛顿法
线性约束优化问题：有效集法
非线性优化问题：罚函数法
现代优化算法：遗传算法等

2.1 无约束优化问题

梯度下降法的框架 1. $x^{k+1} = x^k + \alpha^k d^k$ as long as $\nabla f(x)^T d < 0$
2. A specific form $d^k = -D^k \nabla f(x^k) with D^k > 0$ => Gradient
How far to go along the direction d, or how to select the step size? What are the advantages and disadvantages of each method?

2.1.1 共轭下降法

设目标函数 $f(x) = \frac{1}{2}x^t Gx + bx$
step1 取初始点 x_0 及初始下降方向 d_0 令 k =0。
step2 一维搜索求步长 α_k
step3 令 $x_{k+1} = x_k + \alpha_k d_k$
step4 若 $\|g_k\| \leq \varepsilon$, 则 $x^* = x_{k+1}$, 停止计算；否则转 step5
step5 取共轭方向 d_{k+1}，s.t.

$$d_{k+1}{}^T G d_i = 0, \quad i = 0, 1, 2, ..., k$$

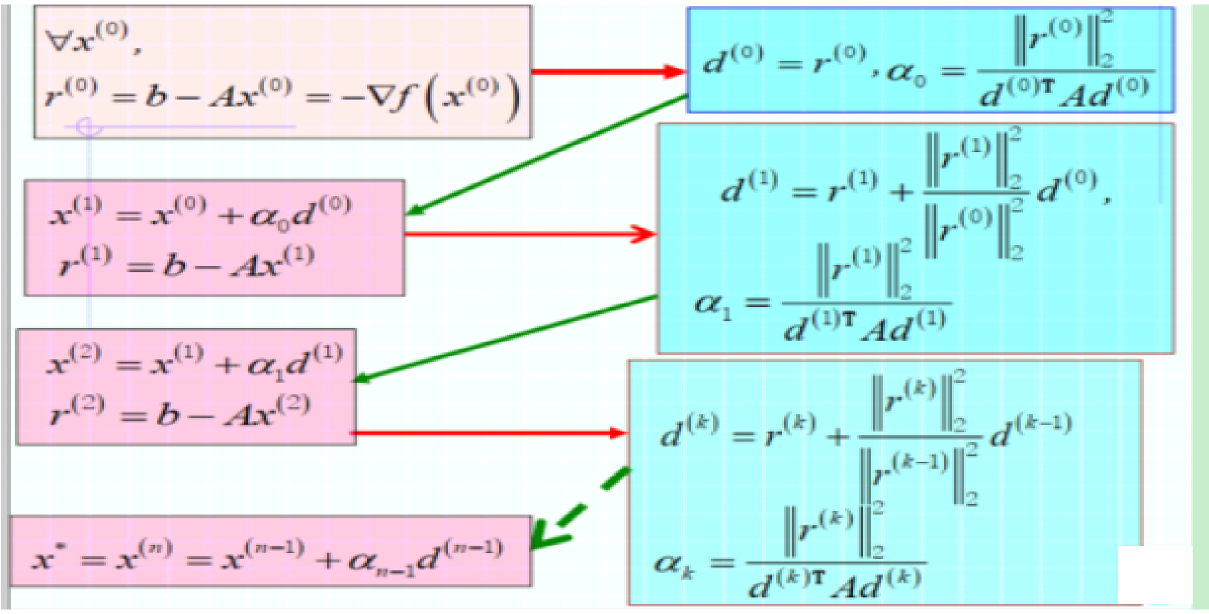


Figure 7: the processing of CGM

step6　若 k= k+1, goto step2.
由于共轭性质，可以推导出更简单的形式如图7所示

2.1.2　最速下降法

给定控制误差 $\varepsilon > 0$
step1　取初始点 x_0, 令 k =0。
step2　计算梯度 $g_k = \nabla f(x_k)$
step3　若 $\|g_k\|) \leq \epsilon$, 则 $x^* = x^k$, 停止计算；否则，令 $d_k = -g_k$, 由一维搜索求步长 α_k
step 4　令 $x_{k+1} = x_k + \alpha_k d_k$　k= k+1,goto step2.

2.1.3　牛顿法

给定控制误差 $\varepsilon > 0$
step1　取初始点 x_0, 令 k =0。
step2　计算梯度 $g_k = \nabla f(x_k)$
step3　若 $\|g_k\|) \leq \epsilon$, 则 $x^* = x^k$, 停止计算；否则，计算 $G_k = \nabla^2 f(x_k)$, 并由 $G_k d_k = -g_k$ 解出 d_k
step 4　令 $x_{k+1} = x_k + d_k$　k= k+1,goto step2.

2.1.4 拟牛顿法

DFP 修正公式

给定控制误差 $\varepsilon > 0$

step1　取初始点 x_0, 和初始矩阵 H_0, H 是正定对称阵，计算 $g_0 = \nabla f(x_0)$ 令 k =0。

step2　令 $d_k = -H_k g_k$

step3　由一维搜索求步长 α_k

step4　令 $x_{k+1} = x_k + \alpha_k d_k$

step5　若 $\|g_k\|) \le \epsilon$, 则 $x^* = x^k$, 停止计算; 否则, 计算 $s_k = x_{k+1} - x_k, y_k = g_{k+1} - gk$

step6 由 DFP 修正公式得

$$H_{k+1} = H_k - \frac{H_k y_k {y_k}^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}$$

BFGS 修正公式：step6:

$$H_{k+1} = H_k - \frac{H_k y_k {y_k}^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + \varphi \omega_k \omega_K^T$$

其中 φ 为参数，可以取任何实数，而

$$\omega_k = (y_k^T H_k y_k)^{0.5} (\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k})$$

其余和 DFP 没有区别，当 $\varphi = 0$ 时，BFGS 即就是 DFP。

2.1.5 四者的优缺点以及如何使用

二次终止性：求解严格凸二次函数极小化问题，计算经至多 n 步终止的性质.

1. 共轭梯度，对二次函数的第一个搜索方向为负梯度方向的前提下具有二次终止性，可以重启共轭梯度法，进一步提高算法的效率。从以前的作业看，速度快于最速下降法。

2. 最速下降法具有全局收敛性，但收敛速度慢，不实用，但是可以和其他算法相结合，比如先用最速达到最优解附近，再用其他算法。本质使用线性函数取近似目标函数，tayor 只展开到一阶。

3. 牛顿法，当海森阵正定且初始点合适，算法二阶收敛，对于正定二次函

数，迭代一次就可以得到极小值。缺点是对多数问题整体不收敛，并且在每次迭代时需要计算 hesse 阵 G_k，以及求逆，得到的 d_k 可能不是下降方向，导致算法收敛于其他非极小值点. 4. 拟牛顿法

(1) 对于正定二次函数

a. 二次收敛性，并且 $H_n = G^{-1}$

b. 保持满足拟牛顿方程

$$s_j = H_i y_j, \quad j = 0, 1, 2, ..., i-1$$

c. 产生的方向是共轭方向

(1) 对于一般函数

a. 保证 H_k 的正定性，从而保证算法的下降性

b. 超线性收敛速度

c. 对于凸函数整体收敛

2.2 线性约束优化问题

2.2.1 有效集法

线性约束满足规范条件，只需要达到 KKT 点，就满足一阶必要条件，再根据二阶充分条件，hesse 阵大于零即可。其基本思想是利用 active 约束将线性约束优化问题为一系列无约束优化问题来求解。The algorithm flow graph is showed in figure8. where

$$\alpha = min\left(\min_{j\in I/S_k \ a_j^T(\delta)_k<0} \frac{b_j - a_j^T * x_k}{a_j^T(\delta)_k}, 1\right) \delta = x^{k+1} - x^k \tag{12}$$

2.3 非线性优化问题

非线性约束满足规范条件，很难达到 KKT 点，只能将约束优化问题为一系列无约束优化问题来求解近似解。

2.3.1 罚函数法

step1 初始化 x_0, 惩罚因子 σ_1, 放大系数 c 和迭代次数 k = 0.
step2 约束优化问题

$$minP(x, \sigma) = f(x) + \sigma_k \overline{P}(x), \quad \sigma > 0$$

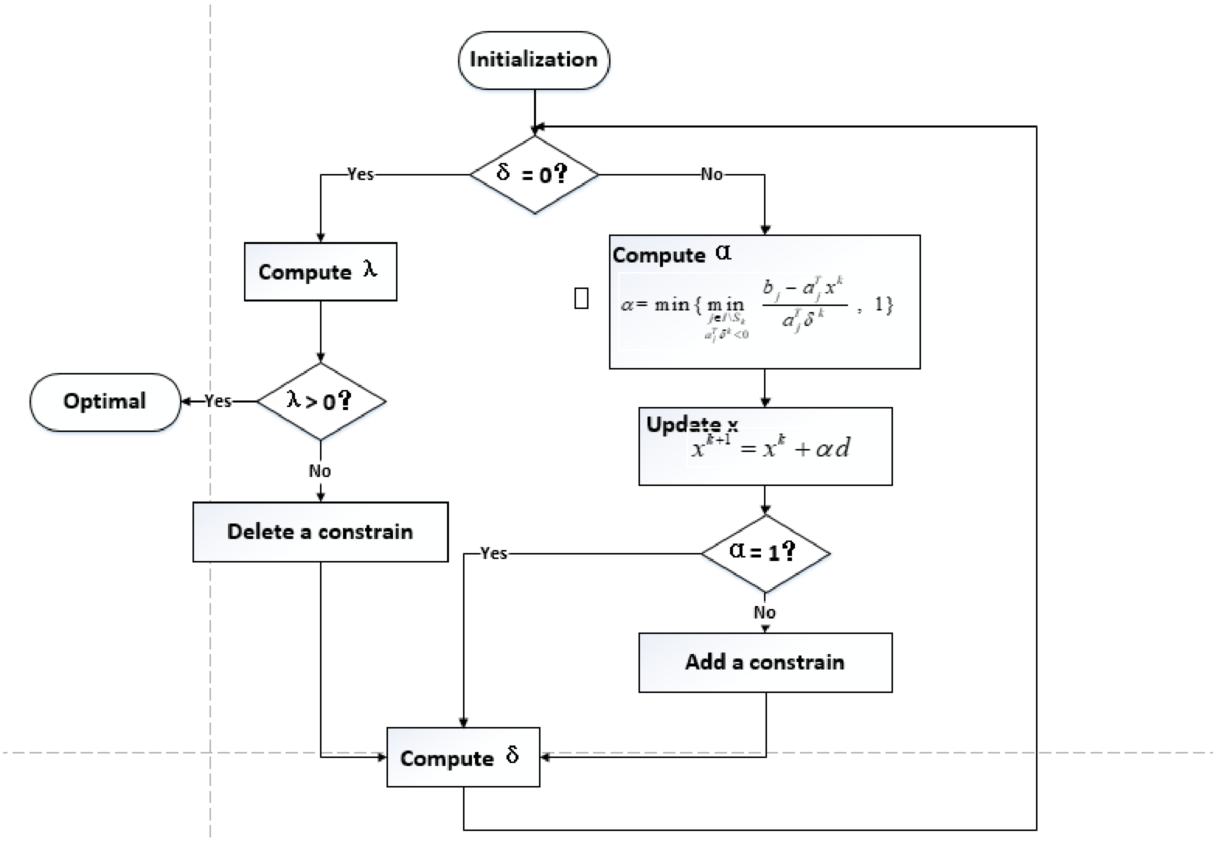


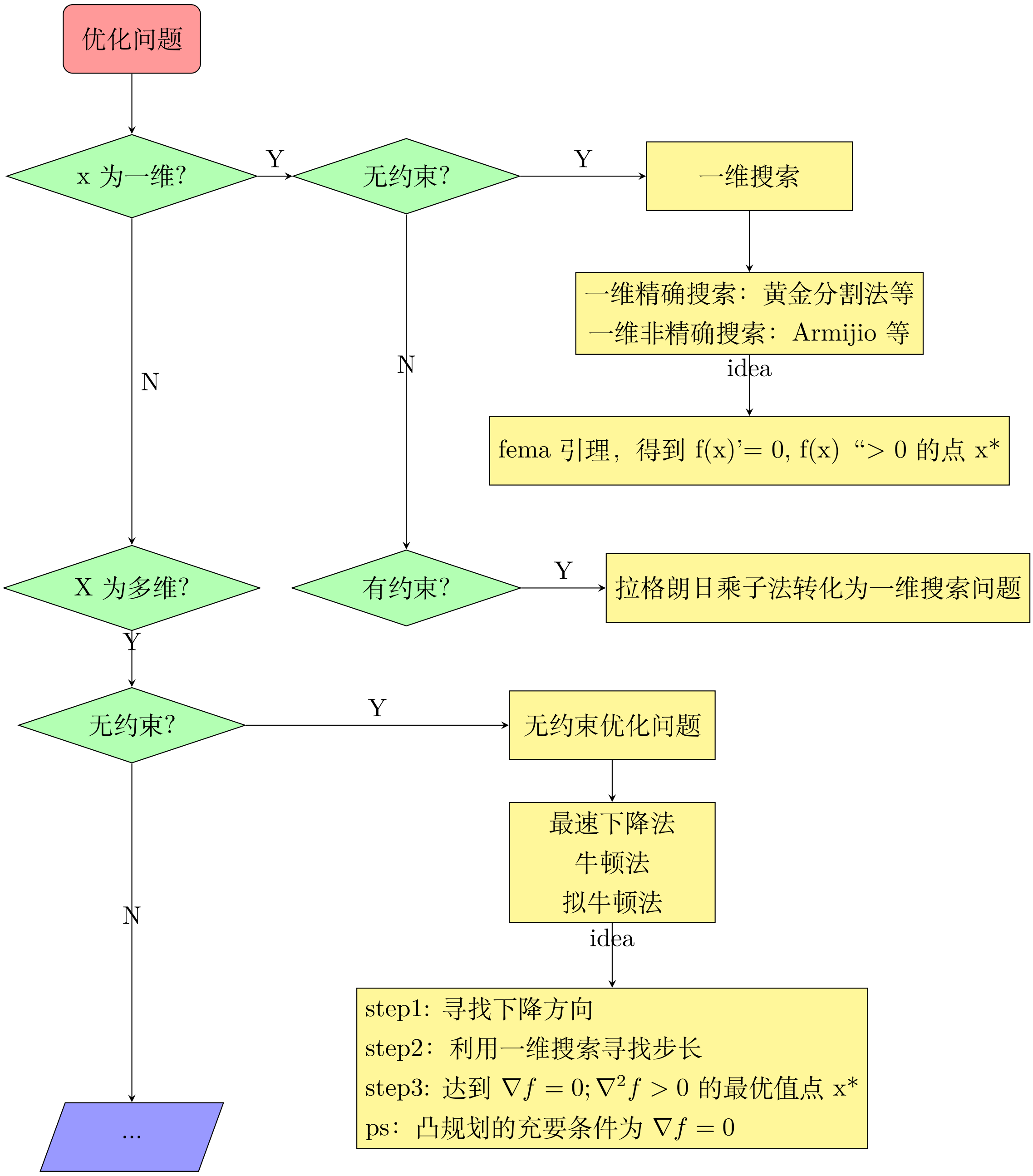
Figure 8: Algorithm flow graph of ASM

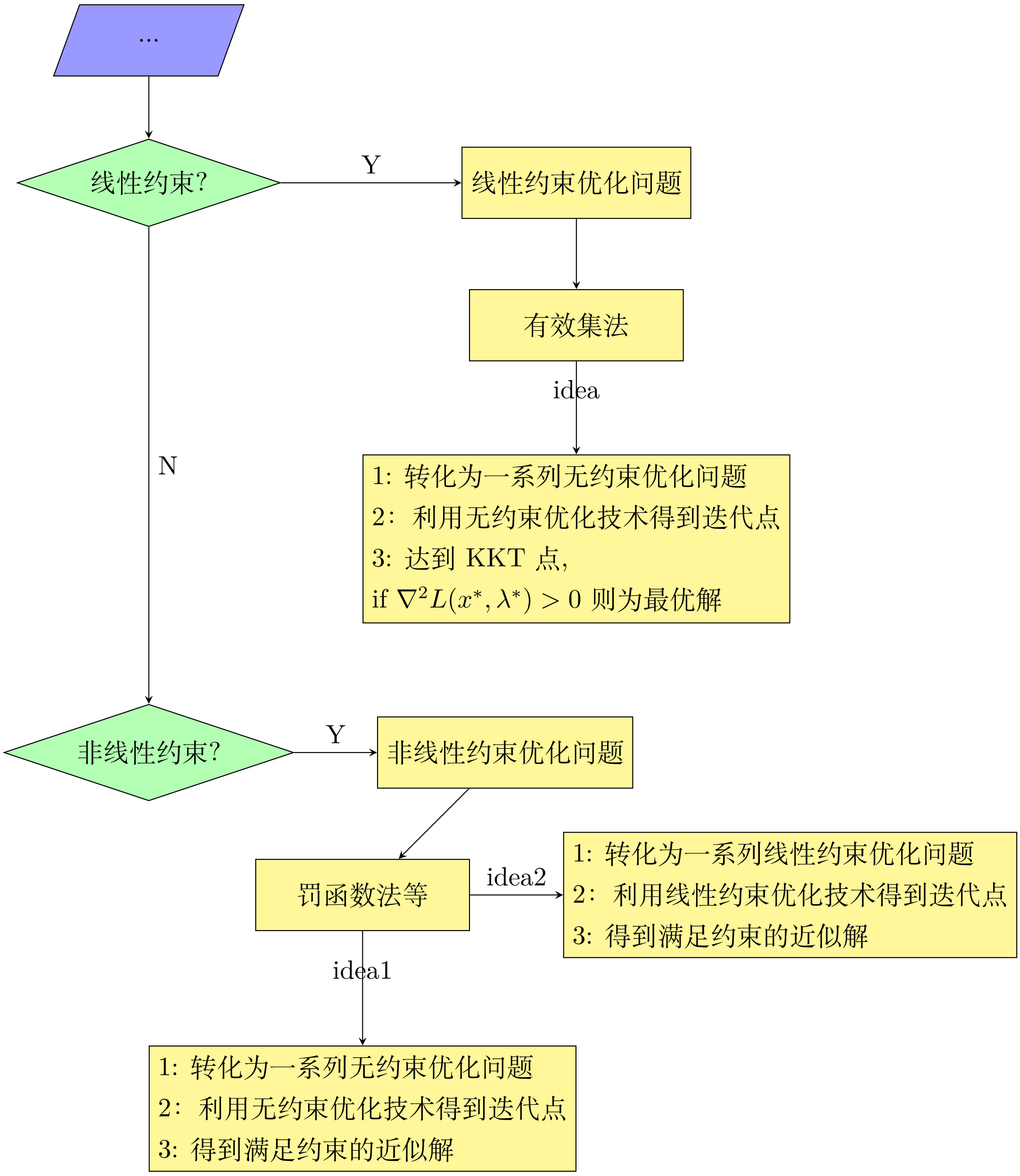
where $\overline{p}(x) = \sum_{i=1}^{m_e} |c_i(x)|^2 + \sum_{i=m_e+1}^{m} |min(0, c_i(x))|^2 yields x_k = x(\sigma_k)$
step3 If $\sigma_k \overline{P}(x) < \epsilon$, x_k 是近似最优解，否则 $\sigma_{k+1} = c\sigma_k$， $k = k + 1$ and goto step2.

3 Experience

解优化问题的基本框架：

1. 当无约束时, 需要得到 f(x)'= 0, f(x) “> 0 的点 x*;
2. 当有约束的时候，得到 KKT 点，if $\nabla^2 L(x^*, \lambda^*) > 0$ 则为最优解。
3. 对于有约束的处理找到 active 或者放缩成等式约束


优化问题
x 为一维?
Y
无约束?
Y
一维搜索
一维精确搜索：黄金分割法等
一维非精确搜索：Armijio 等
idea
fema 引理，得到 f(x)'= 0, f(x) “> 0 的点 x*
N
N
X 为多维?
有约束?
Y
拉格朗日乘子法转化为一维搜索问题
Y
无约束?
Y
无约束优化问题
最速下降法
牛顿法
拟牛顿法
idea
step1：寻找下降方向
step2：利用一维搜索寻找步长
step3：达到 $\nabla f = 0; \nabla^2 f > 0$ 的最优值点 x*
ps：凸规划的充要条件为 $\nabla f = 0$
N
...



课中所讲的最小二乘方法与整数规划由于目标函数 (前者) 和变量约束 (变量) 的特殊性可以独立对待。当我们需要求解一个数学模型，可以按照无约束规划直接求解，当解满足约束的时候，我们便得到了问题的最优解，如果不可以我们按照上面的流程图所画的一步一步求解。

4 Acquirement and Suggestion

4.1 Acquirement

1. 解决一个有问题的基本思路和一些方法, 运用最优化方法对数学模型进行求解
2. 比较不同算法的性质，从初值，收敛性，收敛速度等方向, 了解其性能优劣
3. 对于拉格朗日乘子法和 KKT 点有了更深刻的认识.
4. 在编程与教材结合，提高了编程能力。

4.2 Suggestion

1. 建议小组进行作业，可以进行讨论
2. 班级人数较少的情况下可以进行小组展示，做 final report 的 presentation.