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：

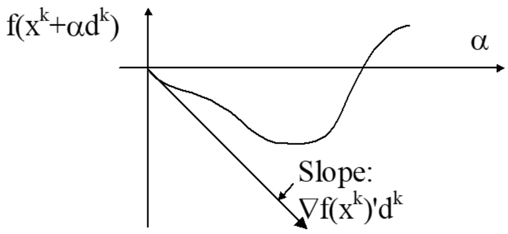

Figure 1: 一维搜索函数图

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. $\begin{cases} Ax \leq 0 \\ c^T x \geq 0 \end{cases}$ 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 $\forall d \in M = \{d \in R_n \mid d^T \nabla c_i(x^*) = 0, i \in E \cup I^*\}$ 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：$x^T A x > 0$,A 为正定矩阵；

$x^T A x \geq 0$,A 为半正定矩阵;

$x^T A x < 0$,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

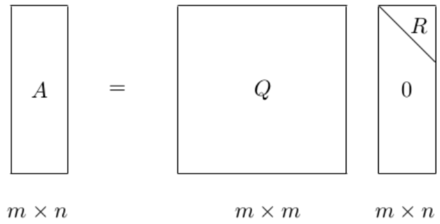

Figure 2: QR 分解示意图

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

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

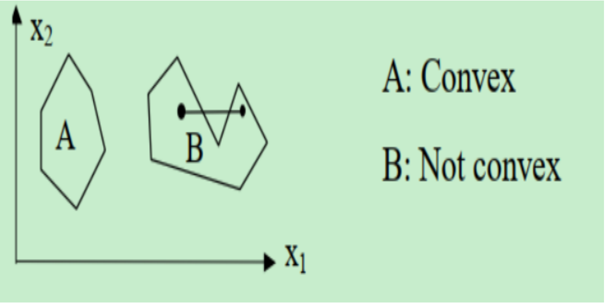

Figure 4: convex set

1.4.2 凸函数

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

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

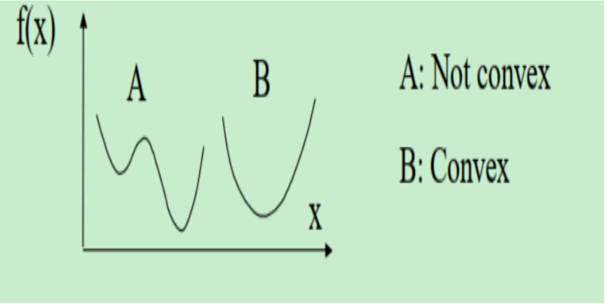

Figure 5: Convex function

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

假设x^k 收敛与

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

3 Acquirement and Suggestion

3.1 Experience

3.2 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.