优化方法第五章作业

总结不等式约束凸优化问题的求解方法及相互联系

$min \quad f_0(x)$

$s.t. f_i(x) \leq 0, i = 1, 2, \ldots m$

$Ax = b$

假设问题严格可行 (Slater 约束品行成立)：存在 $x \in D$ 满足 $Ax = b, f_i(x) < 0$

最优解的充要条件: 存在最优解 $x^* \in \mathbb{R}^n$ 和最优对偶 $\lambda^* \in \mathbb{R}^m, v^* \in \mathbb{R}^p$ 满足 KKT 方程：

$Ax^* = b, f_i(x^*) \leq 0, \lambda_i^* \geq 0, \lambda_i^* f_i(x^*) = 0$

$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + A^T v^* = 0$

可以将等式不等式约束凸优化问题转化为等式约束凸优化问题：定义示性函数 $I_{-}\mathbb{R} \rightarrow \mathbb{R}$

$$I_{-}(i) = \begin{cases} 0 = & u \leq 0 \\ \\ \infty = & u > 0 \end{cases}$$

这样问题转化为：

$min \quad f_0(x) + \sum\limits_{i=1}^m I_{-}(f_i(x))$

$s.t. Ax = b$ 由于示性函数不可微，所以这样替代无法使用牛顿方法。

定义近似示性函数 $\hat{I}_{-}(u) : \mathbb{R} \rightarrow \mathbb{R}, \hat{I}_{-} = -\frac{1}{t} log(-u)$

参数 t 的值越大，近似示性函数越接近示性函数。

对数障碍函数:$\phi : \mathbb{R}^n \rightarrow \mathbb{R}, \phi(x) = -\sum_{i=1}^m log(-f_i(x))$

$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x)$

$\nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x)$

这样问题转化为:$min \quad t f_0(x) + \phi(x)$

$s.t. Ax = b$

对于不同的 t 求解时，可以用上一个 t 值对应问题的最优解作为初始点开始迭代。

中心路径: 对任意 $t > 0$, 近似等式约束优化问题能用牛顿方法求解，且存在唯一解，记作 $x^*(t)$

中心路径 $\{x^*(t) : t > 0\}$ 所有中心路径上的点 $x^*(t)$ 满足 $Ax^*(t) = b, f_i(x^* t(t) < 0)$

存在 $v \in \mathbb{R}^p$, 使得 $t\nabla f_0(x^*(t)) + \nabla\phi(x^*(t)) + A^T v = t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T v = 0$

令 $\lambda_i^*(t) = \frac{1}{-tf_i(x^*(t))}, v^*(t) = \frac{\hat{v}}{t}, then \quad has: \nabla f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t)\nabla f_i(x^*(t)) + A^T v^*(t) = 0$

$\lambda_i^*(t), v^*(t)$ 是对偶可行解, 即 $x^*(t) = argmin_x L(x, \lambda_i^*(t), v^*(t)) = argmin_x \nabla f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t)\nabla f_i(x^*(t)) + v^*(t)^T(Ax - b)$

$x^*(t)$ 和对偶可行解的对偶间隙为 $\frac{m}{t}$

中心路径条件是 KKT 最优性条件的连续变形

障碍方法：顺序求解一系列线性约束的极小问题，由内部迭代得到原问题的可行解，每次所得的可行解作为下一次问题的初始点，逐渐增加 t 的值，直到达到精度要求

等式不等式凸优化问题的障碍方法:

给定严格可行初始点 $x, t = t^0, \mu > 0, \epsilon > 0$, 令 $k = 0$

中心点步骤: 从 x 开始，求解 $x^*(t) = argmint f_0(x) + \phi(x) \quad s.t. Ax = b$

改进:$x = x^*(t)$ 若 $\frac{m}{t} \leq \epsilon$ 退出，否则 $t = \mu t$ 回到中心点步骤

每一次对偶间隙为 $\frac{m}{t}$ 经过 n 次中心点步骤后，对偶间隙为 $\frac{m}{\mu^k t^0}$，故最多经过 $\lceil \frac{log(\frac{m}{\epsilon t^0})}{log\mu} \rceil$ 次外层迭代，算法达到 ϵ 精度要求。

故 μ 的选择很重要，μ 较小，外层迭代次数多，内层迭代较少，μ 较大，外层迭代较少，内层迭代增多。

需要进行阶段一来计算一个严格可行初始点, 在确定严格可行初始点之后，使用障碍方法求解

阶段一方法:

情况 1:

使用障碍方法求解

$x^0 \in dom f_0 \cap \dots dom f_m$ 满足 $Ax^0 = b$

构造等式不等式限制凸优化问题:

$\min\limits_{s,x} s$

$s.t. f_i(x) \leq s, Ax = b$

s 为不等式约束的最大不可行值的上界，该问题总是可行的 $x^0, s^0 > \max\limits_{i=1,2,...,m} f_i(x^0)$

如果 $s^* < 0$, 存在满足 $f_i(x) \leq 0$ 的严格可行解，只要 $s < 0$ 障碍方法迭代就可以停止

$s^* > 0$ 不存在满足 $f_i(x) \leq 0$ 的严格可行解

$s^* = 0$ 且最小值在 $x^*h, s^* = 0$ 处达到，那么不存在严格可行解

$s^* = 0$ 且最小值不可达到, 则不等式组是不可行的

情况 2:

$x^0 \in dom f_0 \cap \ldots dom f_m$ 不满足 $Ax^0 = b$

用不可行初始点牛顿方法求解初始中心点:

$\min\limits_{s,x} t^0 f_0(x) - \sum_1^m log(s - f_i(x))$

$s.t. Ax = b, s = 0$

初始点选择任意 $x \in D, s > \max\limits_{i=1,2,...,m} f_i(x^0)$

情况 3:

不能在函数公共定义域内确定一点

用不可行初始点牛顿方法求解初始中心点:

$\min\limits_{s,x} t^0 f_0(x + z_0) - \sum_1^m log(s - f_i(x + z_i))$

$s.t. Ax = b, s = 0, z_0 = 0, \ldots, z_m = 0$

初始点选择任意 $x \in \mathbb{R}^n, x + z_i \in dom f_i, s > \max\limits_{i=1,2,...,m} f_i(x^0)$

故采用障碍函数法求解等式不等式约束问题:

初始化步骤: 阶段一，采用阶段一方法来解决

确定 x 满足 $f_i(x) < 0, Ax = b$, 设定 t, μ, ϵ 中心点步骤:

从 x 开始，对当前 t 求接近近似约束优化问题 $x^*(t) = argmint f_0(x) + \phi s.t. Ax = b$

停止准则:

若 $\frac{m}{t} \leq \epsilon$ 退出，否则 $t = \mu t$ 回到中心点步骤

原对偶中心残差:

$$t>0, r_t(x,\lambda,v)=\begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix}=\begin{bmatrix} \nabla f_0(x)+Df(x)^T\lambda+A^T \\ -diag(\lambda)f(x)-\frac{1}{t}\mathbf{1} \\ Ax-b \end{bmatrix}$$

其中

$$f(x)=\begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \quad Df(x)=\begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_m(x)^T \end{bmatrix}$$

固定 t, 从满足 $f(x)<0,\lambda>0$ 的点 $y=(x,\lambda,v)$ 开始求解非线性方程 $r_t(x,\lambda,v)$ 的牛顿步进 $d_y=(d_x,d_\lambda,d_v)$

$$r_t(y+d_y)\approx r_t(y)+Dr_t(y)d_y=0\Leftrightarrow\begin{bmatrix} \nabla^2 f_0(x)+\sum_{i=1}^m\lambda_i\nabla^2 f_i(x) & Df(x)^T & A^T \\ -diag(\lambda)Df(x) & -diag(f(x)) & 0 \\ A & 0 & 0 \end{bmatrix}\begin{bmatrix} d_x \\ d_\lambda \\ d_v \end{bmatrix}$$

$$=\quad -\begin{bmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{bmatrix}$$

在 $y^k=(x^k,\lambda^k,v^k)$ 收敛到极限值之前,r_t^k 不一定是可行的，所以不能估计出对偶间隙

定义代理对偶间隙:$\hat{\eta}(x,\lambda)=-f(x)^T\lambda$

如果 x 是原可行的,λ,v 是对偶可行点，那么代理对偶间隙就是对偶间隙，有:

$f_0(x)-g(\lambda,v)=-f(x)^T\lambda=\hat{\eta}(x,\lambda)$

此时对应的 $t=\frac{m}{-f(x)^T\lambda}=\frac{m}{\hat{\eta}}$

原对偶内点法:

初始化步骤: 确定 x 满足 $f_i(x)<0$ 设定 t,μ,ϵ(无等实现制满足要求)

重复基本步骤:

1. 确定 t:$t=\mu\frac{m}{\hat{\eta}(x,\lambda)}=\mu\frac{m}{-f(x)^T\lambda}$

2. 计算原对偶搜索方向 $\nabla y_{pd}=(\nabla x_{pd},\nabla\lambda_{pd},\nabla v_{pd})$

3. 以减少 $\|r_t(y+s\nabla t_{pd})\|_2$ 为目标进行直线搜索，确定步长 $s>0$，令 $y=y+s\nabla y_{pd}$

停止准则：$\|r_{pri}\|_2<\epsilon_{feas},\|r_{dual}\|_2<\epsilon_{feas},\hat{\eta}(x,\lambda)=-f(x)^T\lambda<\epsilon$

原对偶方法仅有一轮迭代，每次迭代更新原对偶变量，且原对偶迭代的值不需要是可行的，一般更有效，展现超线性收敛性质