

优化理论——内点法



星空爱好者
优化理论.机器学习

导语

对于等式约束的优化问题，容易通过牛顿法求解KKT方程组得到最优解，然而对于不等式约束问题KKT方程组则复杂的多。等式不等式联合约束的优化模型如

$$\begin{aligned} \min f(x) \\ \text{s.t. } g(x) = c \\ h(x) \leq t \\ x \in R \end{aligned}$$

其KKT条件为

$$\begin{cases} f'(x) + \lambda g'(x) + \mu h'(x) = 0 \\ g(x) = c \\ h(x) \leq t \\ \mu(h(x) - t) = 0 \\ \mu \geq 0 \end{cases}$$

由于互补松弛条件 $\mu(h(x) - t) = 0$ 的存在，使得KKT方程组不但非线性且常常无法满足正定条件，求解难度大大提升。一个简单的例子

$$\begin{cases} \mu + s = 1 \\ \mu s = 0 \\ \mu, s \geq 0 \end{cases}$$

$us = 0$ 的存在使得方程组不满秩，上述方程的解就难以确定。因此应用牛顿法求解KKT方程组时无法找到有效迭代方向。

虽然LP问题中的图形法、QP问题中的积极集法不需要引入非线性的KKT方程组求解模型，但是仅仅简单线性规划，使用图形法（积极集法属于图形法在二次规划问题中的推广）的复杂度就接近指数级。内点法算是求解优化理论一个非常高效的多项式算法，在LP、NLP问题上都取得了巨大成功，本节主要介绍障碍函数和原始对偶两种内点法。

障碍函数法

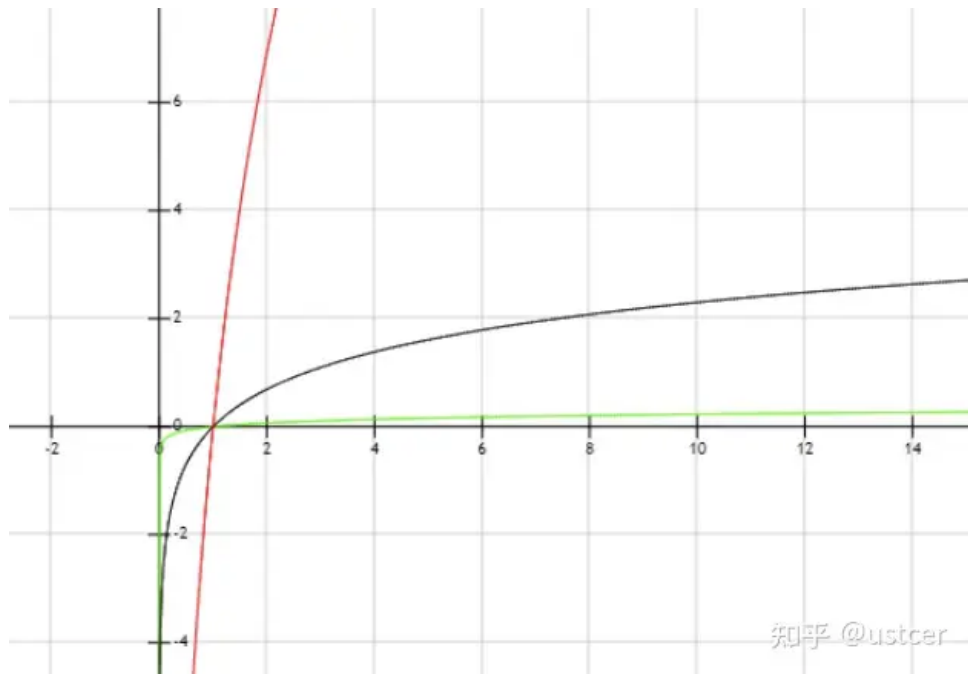
引入障碍函数将约束条件转化为目标函数，生成等价于原模型的优化问题。以引入对数障碍函数为例，原模型近似于

$$\begin{aligned} \min f(x) - \tau \ln(t - h(x)) \\ \text{s.t. } g(x) = c \\ x \in R \end{aligned}$$

τ 只要足够小则下式成立

$$-\tau \ln(s) = \begin{cases} 0, s \geq 0 \\ \infty, s < 0 \end{cases}$$

值得注意的是对数函数 $\ln(s)$ 定义域是 $s > 0$ ，设置成 ∞ 是由于求解KKT方程组需要利用梯度下降法，对于梯度下降法只要初始点在定义域内，迭代无法越过约束边界 $h(x) = t$ ，因为此处函数值无穷大。



$\ln(x)$ 函数，绿线 $\tau = 0.1$ ，黑线 $\tau = 1$ ，红线 $\tau = 10$

我们知道凸优化问题可以转化为更为简洁的KKT方程组进行求解，如果利用KKT条件，要求目标函数必须为凸，很明显 $\frac{\partial^2(-\tau \ln(s))}{\partial^2 s} = \frac{\tau}{s^2} > 0$ ，引入的障碍函数为凸。

内点法模型的拉格朗日函数为

$$L(x, \lambda) = f(x) - \tau \ln(t - h(x)) + \lambda(g(x) - c)$$

对应的KKT条件为

$$\begin{cases} \frac{\partial L}{\partial x} = f'(x) + \frac{\tau h'(x)}{t-h(x)} + \lambda g'(x) = 0 \\ \frac{\partial L}{\partial \lambda} = g(x) - c = 0 \\ h(x) \leq t \end{cases}$$

相比原始的KKT方程，此处没了互补松弛条件，求解难度大幅下降。

接下来首先要证明一件事，引入障碍函数以后的模型最优解接近原模型最优解，尽管直观上我们也会觉得这两个最优解接近。

设 $x^*(\tau), \lambda^*(\tau)$ 为障碍函数法下的最优解，满足

$$L'(x^*(\tau), \lambda^*(\tau)) = f'(x^*(\tau)) + \frac{\tau h'(x^*(\tau))}{t-h(x^*(\tau))} + \lambda^*(\tau) g'(x^*(\tau)) = 0$$

如果指定原模型的拉格朗日乘子，令 $\hat{\mu} = \frac{\tau}{t-h(x^*(\tau))}$ ， $\hat{\lambda} = \lambda^*(\tau)$ ，那么 $x^*(\tau)$ 也是原模型的最优解，因为

$$f'(x^*(\tau)) + \hat{\lambda} g'(x^*(\tau)) + \hat{\mu} h'(x^*(\tau)) = 0$$

此时指定了拉格朗日乘子后的函数值为

$$\begin{aligned} L(x^*(\tau), \hat{\lambda}, \hat{\mu}) &= f(x^*(\tau)) + \lambda^*(\tau)(g(x^*(\tau)) - c) + \frac{\tau}{t-h(x^*(\tau))}(h(x^*(\tau)) - t) \\ &= f(x^*(\tau)) - \tau \end{aligned}$$

至此我们证明障碍函数法模型的最优值与原模型最优值之间的间隔为 τ （被称为对偶间隔），只要 τ 选择的足够小，障碍函数法模型的解就是最优解。

值得注意的并非 τ 越小越好，过小的 τ 可能会导致牛顿法解 KKT 方程组不易收敛，这样做与直接求解原模型 KKT 方程组无异，因此往往对 τ 进行不断的缩小迭代进行最优解的迭代更新。

原始对偶内点法

障碍函数法在 KKT 条件上的表现为对原模型中的互补松弛条件进行放宽，令 $\mu(h(x) - t) = -\tau$ ，得到

$$\begin{cases} f'(x) + \lambda g'(x) + \mu h'(x) = 0 \\ g(x) = c \\ h(x) \leq t \\ \mu(h(x) - t) = -\tau \\ \mu \geq 0 \end{cases}$$

这样的好处是使得方程组易正定。

将其写成矩阵形式

$$F(x, \lambda, \mu) = \begin{bmatrix} f'(x) + \lambda g'(x) + \mu h'(x) \\ \mu(h(x) - t) + \tau \\ g(x) - c \end{bmatrix} = 0$$

对偶内点法相当于假定点在可行域内，不需要考虑不等式约束，因此可用梯度信息获得结果，因此必须控制步长

回顾下牛顿法解方程的迭代步骤

$$F(x_k, \lambda_k, \mu_k) + F'(x_k, \lambda_k, \mu_k)(\Delta x_k, \Delta \lambda_k, \Delta \mu_k) = 0$$

上式方程组是线性的，求解出 $(\Delta x_k, \Delta \lambda_k, \Delta \mu_k)$ 后通过

$$(x_{k+1}, \lambda_{k+1}, \mu_{k+1}) = (x_k, \lambda_k, \mu_k) + \alpha(\Delta x_k, \Delta \lambda_k, \Delta \mu_k) \text{ 进行迭代。}$$

对于障碍函数法而言，牛顿迭代时 τ 不变。

对偶间隔选择：

自然的问题，能否在每步迭代的时候更新 τ ，使得方程组求解更稳定更迅速。对偶法的核心思想是在牛顿法解方程中自适应调整对偶间隔 τ 。假设进行第 k 次迭代时，对偶间隔为

$\eta_k = -\mu(h(x_k) - t)$ ，则设置下一次迭代的对偶间隔 $\tau_{k+1} = \sigma \eta_k$ ，这里 $\sigma \in [0, 1]$ 是用于控制对偶间隔下降速度的因子。原始对偶法与障碍函数法的区别是不再直接初始化 τ ，而是在迭代的每一步中根据当前的对偶间隔更新 τ 。

为了保证求解的正确性，牛顿法解方程停止的条件必须外加一条 $\tau_k \leq \epsilon$

步长选择：

步长选择必须满足约束条件，即满足

$$\begin{cases} \mu_k + \alpha \Delta \mu_k \geq 0 \\ h(x_k + \alpha \Delta x_k) \leq t \end{cases}$$

障碍函数法和原始对偶法极为相似，但是其区别也很明显。障碍函数法对每个 τ 下的方程组需要进行牛顿迭代求解，而外层还需要对 τ 进行不断缩小的迭代尝试。而对偶法仅在求方程组时进行迭代，只有一层迭代。

编辑于 2023-03-16 11:26 · IP 属地未知