增广拉格朗日函数法

文再文

北京大学北京国际数学研究中心

教材《最优化:建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢:本教案由陈乐恒、丁思哲、邓展望协助准备

提纲

等式约束的二次罚函数法

首先考虑简单情形：仅包含等式约束的约束优化问题

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned} \tag{1}$$

其中$x \in \mathbb{R}^n$，$\mathcal{E}$为等式约束的指标集，$c_i(x)$为连续函数

定义该问题的二次罚函数为:

$$P_E(x, \sigma) = f(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(x) \tag{2}$$

其中等式右端第二项称为**罚函数**，$\sigma > 0$称为**罚因子**

- 由于这种罚函数对不满足约束的点进行惩罚，在迭代过程中点列一般处于可行域之外，因此它也被称为**外点罚函数**.

例1

为了直观理解罚函数的作用，我们给出一个例子:

考虑优化问题

$$\begin{aligned} \min \quad & x+\sqrt{3}y \\ \text{s.t.} \quad & x^2+y^2=1 \end{aligned}$$

容易求得最优解为$\left(-\frac{1}{2},-\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}$，考虑二次罚函数

$$P_E(x,y,\sigma)=x+\sqrt{3}y+\frac{\sigma}{2}\left(x^2+y^2-1\right)^2$$

并在下图中绘制出$\sigma=1$和$\sigma=10$对应的罚函数的等高线.

例1

取不同的值时二次罚函数$P_E(x,y,\sigma)$的等高线

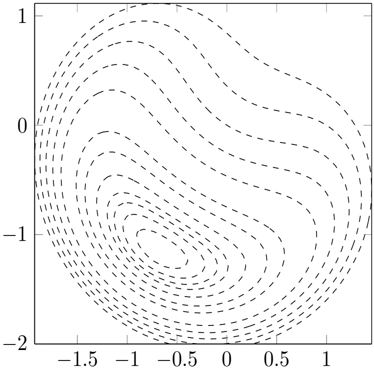

Figure: (a) $\sigma = 1$

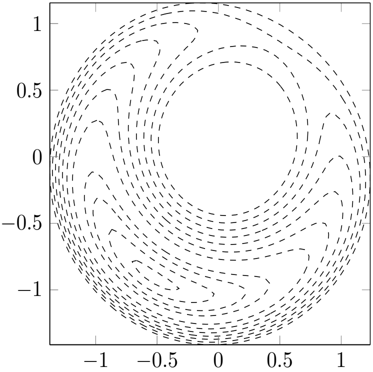

Figure: (b) $\sigma = 10$

例2

下面这个例子表明，当σ选取过小时罚函数可能无下界.

考虑优化问题

$$\begin{aligned} \min \quad & -x^2+2y^2 \\ \text{s.t.} \quad & x=1 \end{aligned}$$

容易求得最优解为$(1,0)^T$，然而考虑罚函数

$$P_E(x,y,\sigma)=-x^2+2y^2+\frac{\sigma}{2}(x-1)^2$$

对任意的$\sigma\leq 2$，该罚函数无下界

二次罚函数法算法

Algorithm 1 二次罚函数法

给定$\sigma_1 > 0, x_0, k \leftarrow 1$. 罚因子增长系数$\rho > 1$.
while 未达到收敛准则 **do**
  以x^k 为初始点，求解$x^{k+1} = \arg\min_x P_E(x, \sigma_k)$
  选取$\sigma^{k+1} = \rho\sigma_k$.
  $k \leftarrow k + 1$
end while

注意事项：

- 选取合适的参数ρ: σ_k增长过快会使子问题求解困难，σ_k增长过慢则会增加迭代次数. 另外，也可以自适应地调整ρ
- 检测到迭代点发散就应该立即终止迭代并增大罚因子
- 为保证收敛，子问题求解误差需要趋于零

分析KKT条件

从KKT条件角度分析：

- 原问题的KKT条件：

$$\nabla f(x^*) - \sum_{i\in\mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0$$

$$c_i(x^*) = 0, \quad \forall i \in \mathcal{E}$$

- 添加罚函数项问题的KKT条件：

$$\nabla f(x) + \sum_{i\in\mathcal{E}} \sigma c_i(x) \nabla c_i(x) = 0$$

假设两个问题收敛到同一点，对比KKT条件(梯度式)，应有下式成立：

$$\sigma c_i\,(x) \approx -\lambda_i^*, \quad \forall i \in \mathcal{E}$$

最优点处乘子λ^*固定，为使约束$c_i(x) = 0$成立，需要$\sigma \to \infty$

分析数值困难

- 考虑罚函数$P_E(x,\sigma)$的海瑟矩阵:

$$\nabla^2_{xx}P_E(x,\sigma)=\nabla^2 f(x)+\sum_{i\in\mathcal{E}}\sigma c_i(x)\nabla^2 c_i(x)+\sigma\nabla c(x)\nabla c(x)^{\mathrm{T}}$$

- 等号右边的前两项可以使用拉格朗日函数$L(x,\lambda^*)$ 来近似，即：

$$\nabla^2_{xx}P_E(x,\sigma)\approx\nabla^2_{xx}L\left(x,\lambda^*\right)+\sigma\nabla c(x)\nabla c(x)^{\mathrm{T}}$$

- 右边为一个定值矩阵和一个最大特征值趋于正无穷的矩阵，这导致$\nabla^2_{xx}P_E(x,\sigma)$ 条件数越来越大，求解子问题的难度也会相应地增加.
- 此时使用梯度类算法求解将会变得非常困难. 若使用牛顿法，则求解牛顿方程本身就是一个非常困难的问题. 因此在实际应用中，我们不可能令罚因子趋于正无穷.

考虑不等式约束问题：

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c_i(x) \leqslant 0, i \in \mathcal{I} \end{array}$$

定义该问题的二次罚函数为:

$$P_I(x,\sigma) = f(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{I}} \tilde{c}_i^2(x)$$

其中 $\tilde{c}_i(x)$ 定义为：

$$\tilde{c}_i(x) = \max\{c_i(x), 0\}$$

注：$h(t) = (\min\{t, 0\})^2$ 关于t可导，故$P_I(x,\sigma)$梯度存在，所以可以使用梯度类算法求解，或者考虑能处理不可微情形的二阶算法，如半光滑牛顿法等等。

现在考虑一般约束问题：

$$
\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E} \\
& c_i(x) \leqslant 0, i \in \mathcal{I}
\end{aligned}
$$

定义该问题的二次罚函数为:

$$
P(x, \sigma) = f(x) + \frac{1}{2}\sigma \left[\sum_{i \in \mathcal{E}} c_i^2(x) + \sum_{i \in \mathcal{I}} \tilde{c}_i^2(x) \right]
$$

其中等式右端第二项称为惩罚项，$\tilde{c}_i(x)$的定义如(3)式，常数$\sigma > 0$称为罚因子.

提纲

二次罚函数法的数值困难

对于等式约束问题:

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned} \tag{3}$$

二次罚函数法需要求解最小化罚函数的子问题:

$$\min_{x} \quad P_E(x,\sigma) = f(x) + \frac{1}{2}\sigma \sum_{i\in\mathcal{E}} c_i^2(x).$$

由$c_i\left(x^{k+1}\right) \approx -\frac{\lambda_i^*}{\sigma_k}$, 为了满足可行性条件, 必须使$\sigma_k$趋于$\infty$, 这造成了子问题求解的数值困难.

我们接下来介绍的增广拉格朗日函数法可以利用有限的罚因子逼近最优解, 从而避免了上述必须使罚因子迅速膨胀的数值困难.

等式约束问题的增广拉格朗日函数法

增广拉格朗日函数法的每步都需要构造增广拉格朗日函数. 根据不同的约束, 增广拉格朗日函数的形式也不同, 因此我们分别论述.

定义

等式约束问题的增广拉格朗日函数

对于等式约束问题(3), 定义**增广拉格朗日函数**为:

$$L_\sigma(x, \lambda) = f(x) + \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2}\sigma \sum_{i \in \mathcal{E}} c_i^2(x).$$

这即是在拉格朗日函数的基础上添加等式约束的二次罚函数.

由定义可得, 在第k步迭代, 给定罚因子σ_k和乘子λ^k, $L_{\sigma_k}\left(x, \lambda^k\right)$的最小值点$x^{k+1}$应满足梯度条件

$$\nabla_x L_{\sigma_k}\left(x^{k+1}, \lambda^k\right) = \nabla f\left(x^{k+1}\right) + \sum_{i \in \mathcal{E}} \left(\lambda_i^k + \sigma_k c_i\left(x^{k+1}\right)\right) \nabla c_i\left(x^{k+1}\right) = 0. \tag{4}$$

等式约束问题的增广拉格朗日函数法

我们将(4)式对比优化问题(3)满足的KKT条件(对最优解(x^*,λ^*)的梯度条件)

$$\nabla f\left(x^*\right)+\sum_{i\in\mathcal{E}}\lambda_i^*\nabla c_i\left(x^*\right)=0, \tag{5}$$

为保证(4)和(5)式在最优解处的一致性，对充分大的k，应满足:

$$\lambda_i^*\approx\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right), \quad \forall i\in\mathcal{E}, \tag{6}$$

即等价于

$$c_i\left(x^{k+1}\right)\approx\frac{1}{\sigma_k}\left(\lambda_i^*-\lambda_i^k\right).$$

由此得出我们希望设计的增广拉格朗日算法具有如下的特性.

性质

- 增广拉格朗日函数法通过合理更新乘子, 即通过控制$\lambda_i^* - \lambda_i^k$降低约束违反度.
  因为根据约束违反度满足的公式, 当λ_i^k足够接近λ_i^*时, $c_i(x^{k+1})$将远小于$1/\sigma_k$.
- (6)式的一个截断近似可以写为:

$$\lambda_i^{k+1} = \lambda_i^k + \sigma_k c_i\left(x^{k+1}\right), \quad \forall i \in \mathcal{E},$$

  这可以作为算法中乘子的更新方式.

根据如上讨论, 并对$c(x)$, $\nabla c(x)$沿用罚函数法的定义, 我们将在下文写出等式约束问题增广拉格朗日函数法的具体算法.

等式约束问题的增广拉格朗日函数法

算法 2 增广拉格朗日函数法

Require: 初始坐标$x^0 \in \mathbb{R}^n$, 乘子λ^0, 罚因子$\sigma_0 > 0$, 约束违反度常数$\varepsilon > 0$, 精度$\eta_k > 0$, 迭代步$k = 0$.

Ensure: x^{k+1}, λ^k.

检查初始元素.
for $k = 0, 1, 2, \cdots$ do **do**
    以x^k为初始点, 求解$\min_x \quad L_{\sigma_k}\left(x, \lambda^k\right)$, 得到满足需求的精度条件$\left\|\nabla_x L_{\sigma_k}\left(x, \lambda^k\right)\right\| \leqslant \eta_k$的解$x^{k+1}$.
    if $\left\|c\left(x^{k+1}\right)\right\| \leqslant \varepsilon$ **then**
        返回近似解(x^{k+1}, λ_k), 终止迭代.
    end if
    更新乘子: $\lambda^{k+1} = \lambda^k + \sigma_k c\left(x^{k+1}\right)$.
    更新罚因子: $\sigma_{k+1} = \rho\sigma_k$.
end for

使用增广拉格朗日函数法的实例

我们考虑优化问题

$$\begin{aligned} \min \quad & x + \sqrt{3}y, \\ \text{s.t.} \quad & x^2 + y^2 = 1. \end{aligned}$$

容易求得最优解为$x^* = \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)^{\mathrm{T}}$, 相应的拉格朗日乘子$\lambda^* = 1$.

根据增广拉格朗日函数的形式, 写出本问题的增广拉格朗日函数:

$$L_\sigma(x, y, \lambda) = x + \sqrt{3}y + \lambda\left(x^2 + y^2 - 1\right) + \frac{\sigma}{2}\left(x^2 + y^2 - 1\right)^2,$$

并在下图中绘制$L_2(x, y, 0.9)$的等高线.

二次罚函数法与增广拉格朗日函数法求解的等高线

图中标“$*$”的点为原问题的最优解x^*
标“$\circ$”的点为罚函数或增广拉格朗日函数的最优解

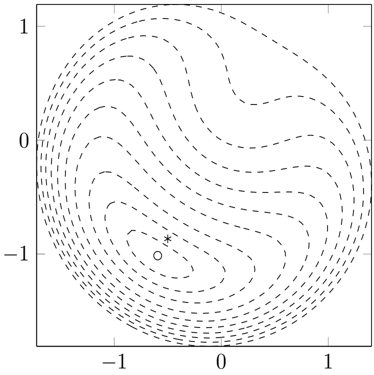

Figure: (a) 二次罚函数

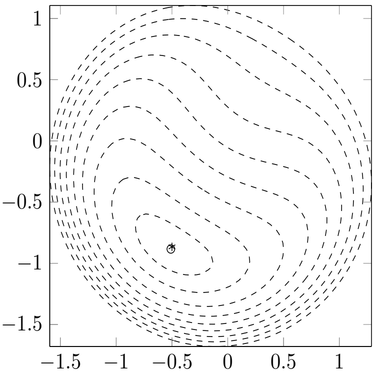

Figure: (b) 增广拉格朗日函数

比较与结论

我们比较二次罚函数和增广拉格朗日函数在最优解探寻方面的有效性.

- 二次罚函数法求出的最优解为$(-0.5957, -1.0319)$, 与最优解的欧氏距离约0.1915, 约束违反度为0.4197.
- 增广拉格朗日罚函数法求出的最优解为$(-0.5100, -0.8833)$, 与最优解的欧氏距离约0.02, 约束违反度为0.0403.

由此可见, 成立如下的经验性结论.

性质

增广拉格朗日函数法可具有比二次罚函数法更精确的寻优能力, 且约束违反度一般更低.

ρ与σ_k的取值指导

在每次迭代确定σ_k时，应考虑如下的问题.

- σ_k不应增长过快:
  (1)随着罚因子σ_k的增大，可见$L_{\sigma_k}(x,\lambda^k)$关于$x$的海瑟矩阵的条件数也将增大，这将导致数值困难;
  (2)σ_k与σ_{k+1}接近时，x^k可以作为求解x^{k+1}的初始点，以加快收敛.
- σ_k不应增长过慢：算法整体的收敛速度将变慢(惩罚不足).

因此在实际中，我们应该控制σ_k的增长维持在一个合理的速度区间内. 一个简单的方法是维持$\rho \in [2,10]$，不过近年来也有学者设计了更合理的自适应方法.

收敛性分析

我们阐述由增广拉格朗日函数法导出的极小值点和原问题的极小值点有什么关系. 实际上, 增广拉格朗日函数在一定条件下将成为精确罚函数.

定理

严格局部极小解定理

设x^*, λ^*分别为问题(3)的局部极小解和相应的乘子, 且点x^* 处$LICQ$ 和二阶充分条件成立. 则:
存在有限的常数$\bar{\sigma}$, 对任意的$\sigma \geqslant \bar{\sigma}$, x^*都是$L_\sigma(x, \lambda^*)$ 的严格局部极小解.

反之, 若x^* 为$L_\sigma(x, \lambda^*)$的局部极小解且满足$c_i(x^*) = 0, i \in \mathcal{E}$, 则$x^*$为问题(3)的局部极小解.

因为x^* 为问题(1) 的局部极小解且二阶充分条件成立, 所以

$$
\begin{aligned}
\nabla_x L\left(x^*, \lambda^*\right) = \nabla f\left(x^*\right) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i\left(x^*\right) = 0, \\
u^{\mathrm{T}} \nabla_{xx}^2 L\left(x^*, \lambda^*\right) u > 0, \\
\nabla c\left(x^*\right)^{\mathrm{T}} u = 0, \quad \forall u.
\end{aligned} \tag{7}
$$

对比$L_\sigma\left(x^*, \lambda^*\right)$和$L\left(x^*, \lambda^*\right)$的表达式, 由$c_i\left(x^*\right) = 0, i \in \mathcal{E}$, 得

$$
\begin{aligned}
\nabla_x L_\sigma\left(x^*, \lambda^*\right) &= \nabla_x L\left(x^*, \lambda^*\right) = 0, \\
\nabla_{xx}^2 L_\sigma\left(x^*, \lambda^*\right) &= \nabla_{xx}^2 L\left(x^*, \lambda^*\right) + \sigma \nabla c\left(x^*\right) \nabla c\left(x^*\right)^{\mathrm{T}}.
\end{aligned} \tag{8}
$$

为了证明x^*是$L_\sigma\left(x^*, \lambda^*\right)$的严格局部极小解, 只需证对于充分大的$\sigma$成立

$$
\nabla_{xx}^2 L_\sigma\left(x^*, \lambda^*\right) \succ 0.
$$

定理证明

假设该结论不成立，则对任意大的σ，存在u_k 满足$\|u_k\| = 1$，且满足：

$$u_k^{\mathrm{T}} \nabla_{xx}^2 L_\sigma (x^*, \lambda^*) u_k = u_k^{\mathrm{T}} \nabla_{xx}^2 L (x^*, \lambda^*) u_k + \sigma \left\| \nabla c (x^*)^{\mathrm{T}} u_k \right\|^2 \leqslant 0,$$

则

$$\left\| \nabla c (x^*)^{\mathrm{T}} u_k \right\|^2 \leqslant -\frac{1}{\sigma} u_k^{\mathrm{T}} \nabla_{xx}^2 L (x^*, \lambda^*) u_k \to 0, \quad \sigma \to \infty.$$

因为$\{u_k\}$ 为有界序列，必存在聚点，设为 u. 那么

$$\nabla c (x^*)^{\mathrm{T}} u = 0, \quad u^{\mathrm{T}} \nabla_{xx}^2 L (x^*, \lambda^*) u \leqslant 0.$$

这与(7)式矛盾，故结论成立.

反之，若x^* 满足$c_i(x^*) = 0$ 且为$L_\sigma(x, \lambda^*)$的局部极小解，那么对于任意与x^* 充分接近的可行点x，我们有

$$f (x^*) = L_\sigma (x^*, \lambda^*) \leqslant L_\sigma (x, \lambda^*) = f(x),$$

因此，x^* 为原问题(3)的一个局部极小解，证毕.

对于增广拉格朗日方法，通过进一步假设乘子点列的有界性和收敛点处的约束品性，算法迭代生成的序列$\{x^k\}$会有子列收敛至问题(3)的一阶稳定点.

定理

增广拉格朗日函数法的收敛性

假设乘子列$\{\lambda^k\}$是有界的，罚因子$\sigma_k \to +\infty,\ k \to \infty$，上述增广拉格朗日方法中精度$\mu_k \to 0$，迭代点列$\{x^k\}$的一个子序列$\{x^{k_j+1}\}$收敛到$x^*$，并且在点$x^*$处$LICQ$成立.

那么存在λ^*，满足:

$$\lambda^{k_j+1} \to \lambda^*, \quad j \to \infty,$$

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0, \quad c(x^*) = 0.$$

定理证明

对于增广拉格朗日函数$L_{\sigma_k}\left(x,\lambda^k\right)$,

$$\begin{aligned}\nabla_x L_{\sigma_k}\left(x^{k+1},\lambda^k\right) &= \nabla f\left(x^{k+1}\right)+\nabla c\left(x^{k+1}\right)\left(\lambda^k+\sigma_k c\left(x^{k+1}\right)\right)\\ &= \nabla f\left(x^{k+1}\right)+\nabla c\left(x^{k+1}\right)\lambda^{k+1}=\nabla_x L\left(x^{k+1},\lambda^{k+1}\right).\end{aligned}$$

由于点x^*处$LICQ$成立，故$\operatorname{rank}\left(\nabla c\left(x^{k_j+1}\right)\right)=|\mathcal{E}|$成立（当$x^{k_j+1}$充分接近$x^*$时），从而下式成立:

$$\lambda^{k_j+1}=\left(\nabla c(x^{k_j+1})^{\mathrm{T}}\nabla c(x^{k_j+1})\right)^{-1}\nabla c(x^{k_j+1})^{\mathrm{T}}\left(\nabla_x L_{\sigma_k}(x^{k_j+1},\lambda^{k_j})-\nabla f(x^{k_j+1})\right).$$

因为$\left\|\nabla_x L_{\sigma_k}\left(x^{k_j+1},\lambda^{k_j}\right)\right\|\leqslant\eta_{k_j}\to 0$, 我们有

$$\lambda^{k_j+1}\to\lambda^*\overset{\text{def}}{=}-\left(\nabla c\left(x^*\right)^{\mathrm{T}}\nabla c\left(x^*\right)\right)^{-1}\nabla c\left(x^*\right)^{\mathrm{T}}\nabla f\left(x^*\right),$$

$$\nabla_x L\left(x^*,\lambda^*\right)=0.$$

而乘子列$\{\lambda^k\}$是有界的，且$\lambda^{k_j}+\sigma_{k_j}c\left(x^{k_j+1}\right)\to\lambda^*$，故$\{\sigma_{k_j}c\left(x^{k_j+1}\right)\}$有界. 又$\sigma_k\to+\infty$, 则$c\left(x^*\right)=0$.

上述收敛性定理的条件还可以进一步放宽，但证明将会更复杂，我们就不多述了.

定理

增广拉格朗日函数法的收敛性(基于更弱的假设)

假设x^*, λ^*分别是问题(3) 的严格局部极小解和相应的乘子，则存在充分大的常数$\bar{\sigma} > 0$ 和充分小的常数$\delta > 0$，如果对某个k，有

$$\frac{1}{\sigma_k}\left\|\lambda^k - \lambda^*\right\| < \delta, \quad \sigma_k \geqslant \bar{\sigma},$$

则

$$\lambda^k \to \lambda^*, \quad x^k \to x^*.$$

同时，如果

(1) $\limsup \sigma_k < +\infty$ 且$\lambda^k \neq \lambda^*, \forall k$，则$\left\{\lambda^k\right\}$ 收敛的速度是Q-线性;
(2) $\limsup \sigma_k = +\infty$ 且$\lambda^k \neq \lambda^*, \forall k$，则$\left\{\lambda^k\right\}$ 收敛的速度是Q-超线性.

提纲

一般的约束优化问题可以写成

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \leqslant 0, i \in \mathcal{I}. \end{aligned} \tag{9}$$

对于问题(9), 我们一般引入松弛变量, 得到如下等价形式:

$$\begin{aligned} \min_{x,s} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) + s_i = 0, i \in \mathcal{I}, \\ & s_i \geqslant 0, i \in \mathcal{I}. \end{aligned} \tag{10}$$

这样的做法我们已经用过多次了, 读者应熟练掌握.

构造增广拉格朗日函数

保留非负约束，可以构造拉格朗日函数

$$L(x,s,\lambda,\mu)=f(x)+\sum_{i\in\mathcal{E}}\lambda_i c_i(x)+\sum_{i\in\mathcal{I}}\mu_i\left(c_i(x)+s_i\right),s_i\geqslant 0,i\in\mathcal{I}.$$

记问题(10)中等式约束的二次罚函数为$p(x,s)$，即

$$p(x,s)=\sum_{i\in\mathcal{E}}c_i^2(x)+\sum_{i\in\mathcal{I}}\left(c_i(x)+s_i\right)^2,$$

那么可以同样构造增广拉格朗日函数如下：

$$\begin{aligned}L_\sigma(x,s,\lambda,\mu)=&f(x)+\sum_{i\in\mathcal{E}}\lambda_i c_i(x)+\sum_{i\in\mathcal{I}}\mu_i\left(c_i(x)+s_i\right)+\\&\frac{\sigma}{2}p(x,s),\quad s_i\geqslant 0,i\in\mathcal{I}.\end{aligned}$$

消元法求解子问题

在第 k 步迭代中，给定乘子λ^k, μ^k 和罚因子σ_k，需要求解如下问题：

$$\min_{x,s} \quad L_{\sigma_k}\left(x, s, \lambda^k, \mu^k\right), \quad \text{s.t.} \quad s \geqslant 0, \tag{11}$$

以得到 x^{k+1}, s^{k+1}.

我们现在介绍一种基于消元的方法，即考虑消去 s，求解只关于 x 的优化问题.

- 首先固定 x，关于 s 的子问题化为

$$\min_{s\geqslant 0} \quad \sum_{i\in\mathcal{I}} \mu_i \left(c_i(x) + s_i\right) + \frac{\sigma_k}{2} \sum_{i\in\mathcal{I}} \left(c_i(x) + s_i\right)^2 .$$

容易直接解得使子问题最优且满足非负约束的s_i为

$$s_i = \max\left\{-\frac{\mu_i}{\sigma_k} - c_i(x), 0\right\}, \quad i \in \mathcal{I}. \tag{12}$$

消元法求解子问题

- 将 s_i 的表达式代入 L_{σ_k} 我们有

$$L_{\sigma_k}\left(x,\lambda^k,\mu^k\right)=f(x)+\sum_{i\in\mathcal{E}}\lambda_i c_i(x)+\frac{\sigma_k}{2}\sum_{i\in\mathcal{E}}c_i^2(x)+\frac{\sigma_k}{2}\sum_{i\in\mathcal{I}}\left(\max\left\{\frac{\mu_i}{\sigma_k}+c_i(x),0\right\}^2-\frac{\mu_i^2}{\sigma_k^2}\right).$$

其为关于x 的连续可微函数(假设$f(x),c_i(x),i\in\mathcal{I}\cup\mathcal{E}$ 连续可微). 因此, 问题(11)等价于

$$\min_{x\in\mathbb{R}^n}\quad L_{\sigma_k}\left(x,\lambda^k,\mu^k\right).$$

并可以利用梯度法进行求解.

注意: 这里, 我们消去了变量s, 因此可以只考虑关于x的优化问题.

更新乘子

对于问题(10), 其最优解x^*, s^* 和乘子λ^*, μ^* 需满足KKT 条件:

$$
\begin{aligned}
&0=\nabla f\left(x^*\right)+\sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i\left(x^*\right)+\sum_{i \in \mathcal{I}} \mu_i^* \nabla c_i\left(x^*\right), \\
&\mu_i^* \geqslant 0,\ s_i^* \geqslant 0, i \in \mathcal{I}
\end{aligned}
$$

问题(11)的最优解x^{k+1}, s^{k+1} 满足

$$
\begin{aligned}
0= & \nabla f\left(x^{k+1}\right)+\sum_{i \in \mathcal{E}}\left(\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right)\right) \nabla c_i\left(x^{k+1}\right)+ \\
& \sum_{i \in \mathcal{I}}\left(\mu_i^k+\sigma_k\left(c_i\left(x^{k+1}\right)+s_i^{k+1}\right)\right) \nabla c_i\left(x^{k+1}\right), \\
s_i^{k+1}= & \max \left\{-\frac{\mu_i^k}{\sigma_k}-c_i\left(x^{k+1}\right), 0\right\}, \quad i \in \mathcal{I} .
\end{aligned}
$$

对比问题(10)和问题(11)的KKT 条件, 易知乘子的更新格式为

$$
\begin{aligned}
\lambda_i^{k+1} & =\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right), \quad i \in \mathcal{E}, \\
\mu_i^{k+1} & =\max \left\{\mu_i^k+\sigma_k c_i\left(x^{k+1}\right), 0\right\}, \quad i \in \mathcal{I} .
\end{aligned} \tag{13}
$$

约束违反度与参数更新

对于等式约束，我们定义约束违反度为

$$v_k\left(x^{k+1}\right)=\sqrt{\sum_{i\in\mathcal{E}} c_i^2\left(x^{k+1}\right)+\sum_{i\in\mathcal{I}}\left(c_i\left(x^{k+1}\right)+s_i^{k+1}\right)^2}.$$

根据 (12) 式消去s，得

$$v_k\left(x^{k+1}\right)=\sqrt{\sum_{i\in\mathcal{E}} c_i^2\left(x^{k+1}\right)+\sum_{i\in\mathcal{I}}\max\left\{c_i\left(x^{k+1}\right),-\frac{\mu_i^k}{\sigma_k}\right\}^2}.$$

在算法中，需要根据约束违反度的大小判断参数的更新方式：

- 若$v_k\left(x^{k+1}\right)$满足精度条件，则进行乘子的更新，并提高子问题求解精度，罚因子不变；
- 若不满足，则不进行乘子的更新，并适当增大罚因子以便得到约束违反度更小的解.

算法2　一般约束增广拉格朗日函数法

- 选取初始点x^0, 乘子λ^0, μ^0, 罚因子$\sigma_0 > 0$, 约束违反度常数$\varepsilon > 0$, 精度常数$\eta > 0$, 以及常数$0 < \alpha \leqslant \beta \leqslant 1$ 和$\rho > 1$. 令$\eta_0 = \frac{1}{\sigma_0}, \varepsilon_0 = \frac{1}{\sigma_0^\alpha}$ 以及$k = 0$.
- **for** $k = 0, 1, 2, \ldots$ **do**
-     以x^k 为初始点, 求解

$$\min_x \quad L_{\sigma_k}\left(x, \lambda^k, \mu^k\right),$$

得到满足精度条件

$$\left\|\nabla L_{\sigma_k}\left(x^{k+1}, \lambda^k, \mu^k\right)\right\|_2 \leqslant \eta_k$$

的解x^{k+1}.

-     **if** $v_k\left(x^{k+1}\right) \leqslant \varepsilon_k$ **then**
-         **if** $v_k\left(x^{k+1}\right) \leqslant \varepsilon$ 且$\left\|\nabla_x L_{\sigma_k}\left(x^{k+1}, \lambda^k, \mu^k\right)\right\|_2 \leqslant \eta$ **then**
-             得到逼近解$x^{k+1}, \lambda^k, \mu^k$, 终止迭代.
-         **end if**

一般约束增广拉格朗日函数法算法(续)

- 更新乘子:

$$\begin{aligned} \lambda_i^{k+1} &= \lambda_i^k + \sigma_k c_i\left(x^{k+1}\right), \quad i \in \mathcal{E}, \\ \mu_i^{k+1} &= \max\left\{\mu_i^k + \sigma_k c_i\left(x^{k+1}\right), 0\right\}, \quad i \in \mathcal{I}. \end{aligned}$$

- 罚因子不变: $\sigma_{k+1} = \sigma_k$.
- 减小子问题求解误差和约束违反度:

$$\eta_{k+1} = \frac{\eta_k}{\sigma_{k+1}}, \quad \varepsilon_{k+1} = \frac{\varepsilon_k}{\sigma_{k+1}^{\beta}}.$$

- **else** (注:约束违反度不满足精度条件)
- 乘子不变: $\lambda^{k+1} = \lambda^k$.
- 更新罚因子: $\sigma_{k+1} = \rho\sigma_k$.
- 调整子问题求解误差和约束违反度:

$$\eta_{k+1} = \frac{1}{\sigma_{k+1}}, \quad \varepsilon_{k+1} = \frac{1}{\sigma_{k+1}^{\alpha}}.$$

- **end if**
- **end for**

提纲

考虑凸优化问题(不等式形式)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leqslant 0, i = 1, 2, \cdots, m. \end{aligned} \tag{14}$$

根据定义, 写出问题(14)的增广拉格朗日函数:

$$L_\sigma(x, \lambda) = f(x) + \frac{\sigma}{2} \sum_{i=1}^{m} \left(\max \left\{ \frac{\lambda_i}{\sigma} + c_i(x), 0 \right\}^2 - \frac{\lambda_i^2}{\sigma^2} \right).$$

给定一列单调递增的乘子$\sigma_k \uparrow \sigma_\infty$, 以及初始乘子$\lambda^0$, 结合(13)式, 问题(14)的增广拉格朗日函数法为

$$\left\{ \begin{aligned} x^{k+1} & \approx \underset{x \in \mathbb{R}^n}{\arg\min} L_{\sigma_k}\left(x, \lambda^k\right), \\ \lambda^{k+1} & = \max\left\{0, \lambda^k + \sigma_k c\left(x^{k+1}\right)\right\}. \end{aligned} \right. \tag{15}$$

不精确条件

定义$\phi_k(x) = L_{\sigma_k}\left(x, \lambda^k\right)$. 由于 $\phi_k(x)$ 的最小值点的显式表达式通常是未知的, 我们往往调用迭代算法求其一个近似解. 为保证收敛性, 我们要求该近似解至少满足不精确条件. 例如:

$$\phi_k\left(x^{k+1}\right) - \inf \phi_k \leqslant \frac{\varepsilon_k^2}{2\sigma_k}, \quad \varepsilon_k \geqslant 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty. \tag{16}$$

由于 $\inf \phi_k$ 是未知的,直接验证(16) 式是数值上不可行的. 但是, 如果ϕ_k是α -强凸函数, 则有

$$\phi_k(x) - \inf \phi_k \leqslant \frac{1}{2\alpha} \operatorname{dist}^2\left(0, \partial \phi_k(x)\right) \tag{17}$$

根据(17)式, 可以进一步构造如下数值可验证的不精确条件:

$$\operatorname{dist}\left(0, \partial \phi_k\left(x^{k+1}\right)\right) \leqslant \sqrt{\frac{\alpha}{\sigma_k}} \varepsilon_k, \quad \varepsilon_k \geqslant 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty. \tag{18}$$

下面我们给出不精确条件下增广拉格朗日函数法的收敛性定理.

定理

凸问题的增广拉格朗日函数法的收敛性

假设$\{x^k\}$, $\{\lambda^k\}$ 为问题(14)通过(15)式生成的序列, x^{k+1} 满足不精确条件(16). 如果问题(14)的*Slater* 约束品性成立, 那么序列$\{\lambda^k\}$ 是有界序列且收敛到λ^∞ *(* λ^∞ 为对偶问题的一个最优解*)*.

如果存在一个γ, 使得下水平集$\{x \in \mathcal{X} \mid f(x) \leqslant \gamma\}$ 是非空有界的, 那么序列$\{x^k\}$ 也是有界的, 并且其所有的聚点都是问题(14)的最优解.

提纲

基追踪问题(BP)

考虑一类简单的基追踪问题. 设 $A \in \mathbb{R}^{m\times n}(m \leqslant n)$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, 基追踪问题被描述为

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad Ax = b. \tag{19}$$

考虑其对偶问题:

$$\min_{y\in\mathbb{R}^m} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad \left\|A^{\mathrm{T}}y\right\|_\infty \leqslant 1. \tag{20}$$

通过引入变量s, 对偶问题可以等价地写成

$$\min_{y\in\mathbb{R}^m,\ s\in\mathbb{R}^n} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad A^{\mathrm{T}}y - s = 0,\ \|s\|_\infty \leqslant 1. \tag{21}$$

原始问题的增广拉格朗日函数法

根据问题(19)的形式，引入罚因子σ和乘子λ，其增广拉格朗日函数为

$$L_\sigma(x,\lambda) = \|x\|_1 + \lambda^{\mathrm{T}}(Ax-b) + \frac{\sigma}{2}\|Ax-b\|_2^2. \tag{22}$$

固定σ，第k步迭代更新格式为

$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\left\{\|x\|_1 + \frac{\sigma}{2}\left\|Ax-b+\frac{\lambda^k}{\sigma}\right\|_2^2\right\}, \\ \lambda^{k+1} = \lambda^k + \sigma\left(Ax^{k+1}-b\right). \end{cases} \tag{23}$$

设迭代初始点$x^0=\lambda^0=0$，考虑格式(23)中的第一步，并假设x^{k+1}为$L_\sigma(x,\lambda^k)$的一个全局极小解，则对$L_\sigma(x,\lambda^k)$利用极小性条件得

$$0 \in \partial\left\|x^{k+1}\right\|_1 + \sigma A^{\mathrm{T}}\left(Ax^{k+1}-b+\frac{\lambda^k}{\sigma}\right).$$

因此成立

$$-A^{\mathrm{T}}\lambda^{k+1} \in \partial\left\|x^{k+1}\right\|_1. \tag{24}$$

BP问题的实例与解

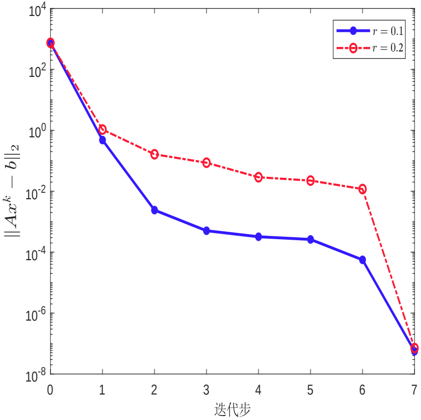

Figure: (a) 约束违反度

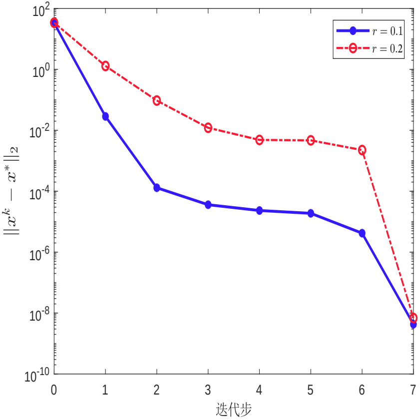

Figure: (b) 与最优点的距离

基追踪问题的收敛性定理

定理

简单基追踪问题的收敛性定理

假设问题(19)的可行域非空, 迭代序列$\{x^k\}$, $\{\lambda^k\}$ 是由迭代格式(23) 从初始点 $x^0 = \lambda^0 = 0$ 产生的, 则存在正整数 K 使得任意的$x^k, k \geqslant K$ 是问题(19)的解.

基追踪问题的Bregman算法

我们介绍了对基追踪问题的增广拉格朗日函数法. 另外, 求解基追踪问题的一个通用方法是Bregman迭代算法.

对于凸函数$h(x) = \|x\|_1$, 定义其 Bregman距离:

$$D_h^g(x, y) = h(x) - h(y) - \langle g, x - y \rangle,$$

其中$g \in \partial h(y)$ 为函数 h 在点 y 处的一个次梯度.

定义

基追踪问题的Bregman迭代算法

根据*Bregman*距离的定义, 为基追踪问题(19)设计的*Bregman* 迭代算法如下所示. 其中, $g^{k+1} \in \partial h(x^{k+1})$.

$$\begin{cases} x^{k+1} = \underset{x \in \mathbb{R}^n}{\arg\min} \left\{ D_h^{g^k}\left(x, x^k\right) + \frac{1}{2}\|Ax - b\|_2^2 \right\}, \\ g^{k+1} = g^k - A^{\mathrm{T}}\left(Ax^{k+1} - b\right). \end{cases}$$

基追踪问题的2种算法之比较

对比基于增广拉格朗日函数设计的算法(23)

$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\left\{\|x\|_1 + \frac{\sigma}{2}\left\|Ax - b + \frac{\lambda^k}{\sigma}\right\|_2^2\right\}, \\ \lambda^{k+1} = \lambda^k + \sigma\left(Ax^{k+1} - b\right). \end{cases}$$

这两种算法具有如下的等价性质.

性质

如果基于增广拉格朗日函数设计的算法(23)的初始点设置为$(x^0, -A^{\mathrm{T}}\lambda^0)$, 则有

$$g^k = -A^{\mathrm{T}}\lambda^k.$$

即在合理选取初始点时, 两个算法得到的迭代点列完全一致(此时增广拉格朗日函数法中$\sigma = 1$固定), 因此算法是等价的.

两种算法的内在关系如何, 与它们求解的效率高低比较直接相关. 这进一步说明, 在合理选取初始点的情况下, 以上2种方法的效率一致.

现在我们考虑另一重要的问题. 设对偶问题(21):

$$\min_{y\in\mathbb{R}^m, s\in\mathbb{R}^n} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad A^{\mathrm{T}}y - s = 0, \quad \|s\|_\infty \leqslant 1.$$

引入拉格朗日乘子λ和罚因子σ, 作增广拉格朗日函数

$$L_\sigma(y,s,\lambda) = b^{\mathrm{T}}y + \lambda^{\mathrm{T}}\left(A^{\mathrm{T}}y - s\right) + \frac{\sigma}{2}\left\|A^{\mathrm{T}}y - s\right\|_2^2, \quad \|s\|_\infty \leqslant 1.$$

增广拉格朗日函数法的迭代格式为($\rho > 1$ 和 $\bar{\sigma} < +\infty$ 为算法参数):

$$\begin{cases} \left(y^{k+1}, s^{k+1}\right) = \mathop{\arg\min}\limits_{y, \|s\|_\infty \leqslant 1} \left\{ b^{\mathrm{T}}y + \dfrac{\sigma_k}{2}\left\|A^{\mathrm{T}}y - s + \dfrac{\lambda}{\sigma_k}\right\|_2^2 \right\}, \\ \lambda^{k+1} = \lambda^k + \sigma_k\left(A^{\mathrm{T}}y^{k+1} - s^{k+1}\right), \\ \sigma_{k+1} = \min\{\rho\sigma_k, \bar{\sigma}\}. \end{cases}$$

其中$\left(y^{k+1}, s^{k+1}\right)$的显式表达式未知, 需要迭代求解.

消元法求解子问题

除了利用投影梯度法求解关于 (y, s) 的联合最小化问题外，还可以利用最优性条件将 s 用 y 来表示，转而求解只关于 y 的最小化问题.

关于s的极小化问题为

$$\min_{s} \quad \frac{\sigma}{2}\left\|A^{\mathrm{T}}y - s + \frac{\lambda}{\sigma}\right\|_2^2, \qquad \textbf{s.t.} \quad \|s\|_\infty \leqslant 1.$$

这是一个关于s的二次型函数，因此问题的解为

$$s = \mathcal{P}_{\|s\|_\infty \leqslant 1}(A^{\mathrm{T}}y + \frac{\lambda}{\sigma}),$$

其中$\mathcal{P}_{\|s\|_\infty \leqslant 1}(z)$为集合$\{s | \, \|s\|_\infty \leqslant 1\}$的**投影算子**，即

$$\mathcal{P}_{\|s\|_\infty \leqslant 1}(z) = \max\left\{\min\left\{z, \ 1\right\}, \ -1\right\}.$$

消元法求解子问题

将上述s的表达式代入的增广拉格朗日函数法的迭代格式，得

$$
\begin{cases}
y^{k+1} = \underset{y}{\arg\min} \left\{ b^\mathrm{T} y + \frac{\sigma}{2} \left\| \psi\left(A^\mathrm{T} y + \frac{\lambda}{\sigma}\right) \right\|_2^2 \right\}, \\
\lambda^{k+1} = \sigma_k \psi\left(A^\mathrm{T} y^{k+1} + \frac{\lambda^k}{\sigma_k}\right), \\
\sigma_{k+1} = \min\left\{\rho\sigma_k, \bar{\sigma}\right\}.
\end{cases}
\tag{25}
$$

其中$\psi(x) = \mathrm{sign}(x)\max\{|x| - 1, 0\}$.

我们不能得到关于 y^{k+1} 的显式表达式. 但是由于 $L_{\sigma_k}\left(y, \lambda^k\right)$ 关于 y 连续可微, 故可以利用梯度法求解.