二次罚函数法的数值困难

对于等式约束问题:

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

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(1), 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

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

我们将(2)式对比优化问题(1)满足的KKT条件(对最优解(x^*,λ^*)的梯度条件)

$$\nabla f\left(x^*\right)+\sum_{i\in\mathcal{E}}\lambda_i^*\nabla c_i\left(x^*\right)=0, \tag{3}$$

为保证(2)和(3)式在最优解处的一致性, 对充分大的k, 应满足:

$$\lambda_i^*\approx\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right),\quad \forall i\in\mathcal{E}, \tag{4}$$

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$1/\sigma_k$.
- (4)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.

算法 1　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.

1: 检查初始元素.
2: **for** $k = 0, 1, 2, \cdots$ do **do**
3: 以x^k为初始点, 求解$\min_x \quad L_{\sigma_k}\left(x, \lambda^k\right)$, 得到满足精度条件$\left\|\nabla_x L_{\sigma_k}\left(x, \lambda^k\right)\right\| \leqslant \eta_k$的解$x^{k+1}$.
4: **if** $\left\|c\left(x^{k+1}\right)\right\| \leqslant \varepsilon$ **then**
5: 返回近似解(x^{k+1}, λ_k), 终止迭代.
6: **end if**
7: 更新乘子: $\lambda^{k+1} = \lambda^k + \sigma_k c\left(x^{k+1}\right)$.
8: 更新罚因子: $\sigma_{k+1} = \rho\sigma_k$.
9: **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

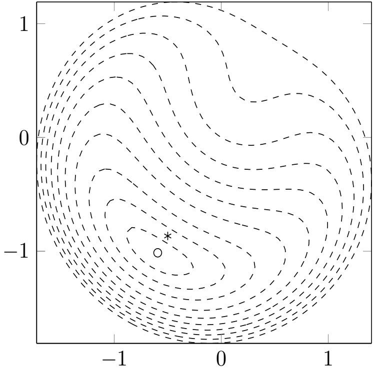

Figure: (a) 二次罚函数

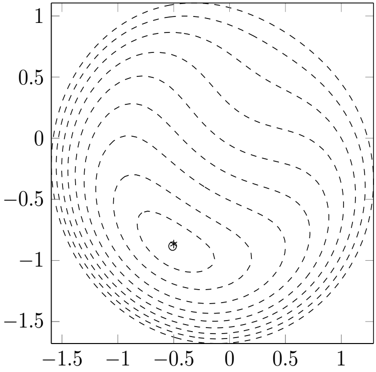

Figure: (b) 增广拉格朗日函数

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.

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(1)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(1)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

$$
\begin{aligned}
\nabla_x L\left(x^*, \lambda^*\right) = \nabla f\left(x^*\right) + \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i\left(x^*\right) = 0, \\
u^{\mathrm{T}} \nabla_{xx}^2 L\left(x^*, \lambda^*\right) u > 0, \\
\nabla c\left(x^*\right)^{\mathrm{T}} u = 0, \quad \forall u.
\end{aligned} \tag{5}
$$

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

$$
\begin{aligned}
\nabla_x L_\sigma\left(x^*, \lambda^*\right) &= \nabla_x L\left(x^*, \lambda^*\right) = 0, \\
\nabla_{xx}^2 L_\sigma\left(x^*, \lambda^*\right) &= \nabla_{xx}^2 L\left(x^*, \lambda^*\right) + \sigma \nabla c\left(x^*\right) \nabla c\left(x^*\right)^{\mathrm{T}}.
\end{aligned} \tag{6}
$$

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

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

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(5)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(1)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(1)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:

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

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

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

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

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(1) 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

一般约束问题的增广拉格朗日函数法

一般的约束优化问题可以写成

$$
\begin{aligned}
\min \quad & f(x), \\
\text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E}, \\
& c_i(x) \leqslant 0, i \in \mathcal{I}.
\end{aligned} \tag{7}
$$

对于问题(7), 我们一般引入松弛变量, 得到如下等价形式:

$$
\begin{aligned}
\min_{x,s} \quad & f(x), \\
\text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E}, \\
& c_i(x) + s_i = 0, i \in \mathcal{I}, \\
& s_i \geqslant 0, i \in \mathcal{I}.
\end{aligned} \tag{8}
$$

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(8)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:

$$\min_{x,s} \quad L_{\sigma_k}\left(x, s, \lambda^k, \mu^k\right), \qquad \text{s.t.} \quad s \geqslant 0, \tag{9}$$

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

$$s_i = \max\left\{-\frac{\mu_i}{\sigma_k} - c_i(x), 0\right\}, \quad i \in \mathcal{I}. \tag{10}$$

- 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(9)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(8), 其最优解x^*, s^* 和乘子λ^*, μ^* 需满足KKT 条件:

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

问题(9)的最优解x^{k+1}, s^{k+1} 满足

$$0=\nabla f\left(x^{k+1}\right)+\sum_{i\in\mathcal{E}}\left(\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right)\right)\nabla c_i\left(x^{k+1}\right)+$$

$$\sum_{i\in\mathcal{I}}\left(\mu_i^k+\sigma_k\left(c_i\left(x^{k+1}\right)+s_i^{k+1}\right)\right)\nabla c_i\left(x^{k+1}\right),$$

$$s_i^{k+1}=\max\left\{-\frac{\mu_i^k}{\sigma_k}-c_i\left(x^{k+1}\right),0\right\},\quad i\in\mathcal{I}.$$

对比问题(8)和问题(9)的KKT 条件, 易知乘子的更新格式为

$$\begin{aligned}\lambda_i^{k+1}&=\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right),\quad i\in\mathcal{E},\\ \mu_i^{k+1}&=\max\left\{\mu_i^k+\sigma_k c_i\left(x^{k+1}\right),0\right\},\quad i\in\mathcal{I}.\end{aligned}\tag{11}$$

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 (10) 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

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

　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**

- 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)

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leqslant 0, i=1,2,\cdots,m. \end{aligned} \tag{12}$$

根据定义，写出问题(12)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(11)式，问题(12)的增广拉格朗日函数法为

$$\begin{cases} x^{k+1} \approx \underset{x\in\mathbb{R}^n}{\arg\min} L_{\sigma_k}\left(x,\lambda^k\right), \\ \lambda^{k+1} = \max\left\{0, \lambda^k + \sigma_k c\left(x^{k+1}\right)\right\}. \end{cases} \tag{13}$$

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:

$$\phi_k\left(x^{k+1}\right) - \inf \phi_k \leqslant \frac{\varepsilon_k^2}{2\sigma_k}, \quad \varepsilon_k \geqslant 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty. \tag{14}$$

由于 $\inf \phi_k$ 是未知的,直接验证(14) 式是数值上不可行的. 但是, 如果ϕ_k是α -强凸函数, 则有

$$\phi_k(x) - \inf \phi_k \leqslant \frac{1}{2\alpha} \operatorname{dist}^2\left(0, \partial\phi_k(x)\right) \tag{15}$$

根据(15)式, 可以进一步构造如下数值可验证的不精确条件:

$$\operatorname{dist}\left(0, \partial\phi_k\left(x^{k+1}\right)\right) \leqslant \sqrt{\frac{\alpha}{\sigma_k}}\varepsilon_k, \quad \varepsilon_k \geqslant 0, \quad \sum_{k=1}^{\infty} \varepsilon_k < +\infty. \tag{16}$$

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(12)通过(13)式生成的序列, x^{k+1} 满足不精确条件(14). 如果问题(12)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(12)的最优解.

提纲

1. 等式约束问题的增广拉格朗日函数法
2. 一般约束问题的增广拉格朗日函数法
3. 凸优化问题的增广拉格朗日函数法
4. 应用:基追踪问题
   - 原始问题的增广拉格朗日函数法
   - 与Bregman算法的等价性
   - 对偶问题的增广拉格朗日函数法
5. 应用: 一般线性规划问题
6. 应用:半定规划问题

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

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad Ax = b. \tag{17}$$

考虑其对偶问题:

$$\min_{y\in\mathbb{R}^m} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad \|A^{\mathrm{T}}y\|_\infty \leqslant 1. \tag{18}$$

通过引入变量s, 对偶问题可以等价地写成

$$\min_{y\in\mathbb{R}^m,\ s\in\mathbb{R}^n} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad A^{\mathrm{T}}y - s = 0,\ \|s\|_\infty \leqslant 1. \tag{19}$$

原始问题的增广拉格朗日函数法

根据问题(17)的形式，引入罚因子σ和乘子λ，其增广拉格朗日函数为

$$L_\sigma(x,\lambda) = \|x\|_1 + \lambda^\mathrm{T}(Ax-b) + \frac{\sigma}{2}\|Ax-b\|_2^2. \tag{20}$$

固定σ，第k步迭代更新格式为

$$\begin{cases} x^{k+1} = \underset{x\in\mathbb{R}^n}{\arg\min}\left\{\|x\|_1 + \frac{\sigma}{2}\left\|Ax-b+\frac{\lambda^k}{\sigma}\right\|_2^2\right\}, \\ \lambda^{k+1} = \lambda^k + \sigma\left(Ax^{k+1}-b\right). \end{cases} \tag{21}$$

设迭代初始点$x^0=\lambda^0=0$，考虑格式(21)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

$$-A^\mathrm{T}\lambda^{k+1} \in \partial\left\|x^{k+1}\right\|_1. \tag{22}$$

例　简单基追踪问题的增广拉格朗日函数解法
考虑标准基追踪问题，其中A是512×1024规模的随机矩阵(每个元素从标准正态分布中抽样)，b定义为

$$b = Au,$$

其中$u \in \mathbb{R}^{1024}$是服从正态分布随机稀疏向量，设其稀疏度$r = 0.1$或0.2，即分别约具有102/205个服从正态分布的非零分量.
进一步地，我们固定罚因子σ，采用近似点梯度法作为求解器(第八章将会详细介绍)，不精确地求解关于x的子问题以得到x^{k+1}.
我们演示的算法中设置了求解精度$\eta_k = 10^{-k}$，并使用**BB步长**作为线搜索的初始步长. 下图展示了算法产生的迭代点与最优点的欧式距离的变化，以及它们约束违反度的变化趋势. 由图可知:

性质

对于本例中的标准基追踪问题，固定的罚因子σ也可以使增广拉格朗日函数法收敛.

BP问题的实例与解

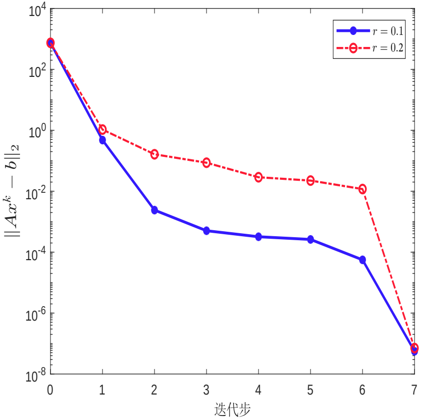

Figure: (a) 约束违反度

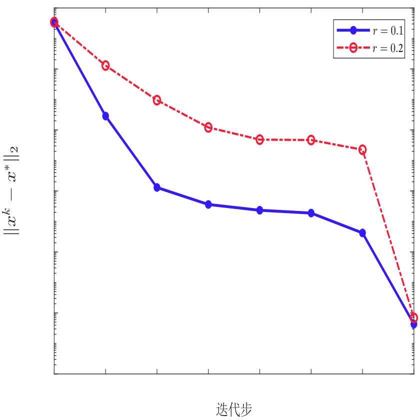

Figure: (b) 与最优点的距离

上述由一个简单基追踪问题导出的结论并非孤例. 我们证明, 对固定的二次罚项系数$\sigma = 1$, 迭代格式**(21)**具有有限终止性. 这是一个很强的结论.

引理1

设迭代序列$\{x^k\}$, $\{\lambda^k\}$ 是算法(20)从初始点$x^0 = \lambda^0 = 0$ 产生的序列, 则若

(a) $\|Ax^k - b\|_2$ 且单调下降的, 则

$$\|Ax^{k+1} - b\|_2 \leqslant \|Ax^k - b\|_2 . \tag{23}$$

(b) 若存在$\tilde{x}$ 满足$A\tilde{x} = b$, 则

$$\frac{\sigma}{2}\|Ax^k - b\|_2^2 \leqslant \frac{1}{k}\|\tilde{x}\|_1 . \tag{24}$$

我们下面将证明这一引理.

(a) 由迭代格式(21)的第一步,

$$\left\|x^{k+1}\right\|_1+\left(\lambda^k\right)^{\mathrm{T}}\left(Ax^{k+1}-b\right)+\frac{\sigma}{2}\left\|Ax^{k+1}-b\right\|_2^2$$
$$\leqslant\left\|x^k\right\|_1+\left(\lambda^k\right)^{\mathrm{T}}\left(Ax^k-b\right)+\frac{\sigma}{2}\left\|Ax^k-b\right\|_2^2.$$

由于$\|x\|_1$ 的凸性和(22)式, 我们有

$$\left\|x^{k+1}\right\|_1\geqslant\left\|x^k\right\|_1+\left\langle -A^{\mathrm{T}}\lambda^k,x^{k+1}-x^k\right\rangle,$$

结合上面两式,

$$\left\|Ax^{k+1}-b\right\|_2\leqslant\left\|Ax^k-b\right\|_2.$$

至此, 引理中(a)所述的结论证毕.

(b) 由迭代格式(21)的第二步,

$$A^{\mathrm{T}}\left(\lambda^{k+1}-\lambda^{k}\right)=\sigma A^{\mathrm{T}}\left(Ax^{k+1}-b\right), \tag{25}$$

于是

$$\begin{aligned}
&\frac{\sigma}{2}\left\|Ax^{k+1}-b\right\|_2^2-\frac{\sigma}{2}\|Ax-b\|_2^2\\
\leqslant&\left\langle A^{\mathrm{T}}\left(\lambda^{k+1}-\lambda^{k}\right), x^{k+1}-x\right\rangle\\
&\text{(}\tfrac{\sigma}{2}\|Ax-b\|_2^2\ \text{点}x^{k+1}\text{处的凸性以及(25)式)}\\
=&\left\langle A^{\mathrm{T}}\lambda^{k+1}, x^{k+1}-x\right\rangle-\left\langle A^{\mathrm{T}}\lambda^{k}, x^{k}-x\right\rangle-\left\langle A^{\mathrm{T}}\lambda^{k}, x^{k+1}-x^{k}\right\rangle\\
\leqslant&\left\langle A^{\mathrm{T}}\lambda^{k+1}, x^{k+1}-x\right\rangle-\left\langle A^{\mathrm{T}}\lambda^{k}, x^{k}-x\right\rangle+\left\|x^{k+1}\right\|_1-\left\|x^{k}\right\|_1.\\
&\text{(}\|x\|_1\ \text{点}x^{k}\text{处的凸性以及(22)式)}
\end{aligned}$$

进一步,

$$
\begin{aligned}
& k\left(\frac{\sigma}{2}\left\|A x^k-b\right\|_2^2-\frac{\sigma}{2}\|A x-b\|_2^2\right) \\
\leqslant & \sum_{j=1}^k\left(\frac{\sigma}{2}\left\|A x^j-b\right\|_2^2-\frac{\sigma}{2}\|A x-b\|_2^2\right) \quad \left(\left\|A x^k-b\right\|_2 \text{ 的单调性}\right) \\
\leqslant & \left\langle A^{\mathrm{T}} \lambda^k, x^k-x\right\rangle+\left\|x^k\right\|_1-\left\langle A^{\mathrm{T}} \lambda^0, x^0-x\right\rangle-\left\|x^0\right\|_1 \quad \left(\|x\|_1 \text{ 的凸性}\right) \\
\leqslant & \|x\|_1,
\end{aligned}
$$

取$x=\tilde{x}$, 我们即有

$$
\frac{\sigma}{2}\left\|A x^k-b\right\|_2^2 \leqslant \frac{1}{k}\|\tilde{x}\|_1 .
$$

至此, 引理中(b)所述的结论证毕.

收敛性证明的引理2

我们再论述一个重要的引理，它论述了迭代解的存在性.

引理2

假设问题(17)的可行域非空，x^k 是由迭代格式(21)得到的满足$Ax^k = b$ 的迭代点，则x^k 是BP 问题(17)的一个解.

Proof 对任意x，由$\|x\|_1$ 的凸性和(22)式，有

$$
\begin{aligned}
\left\|x^k\right\|_1 &\leqslant \|x\|_1 - \left\langle x - x^k, -A^{\mathrm{T}}\lambda^k \right\rangle \\
&= \|x\|_1 + \left\langle Ax - Ax^k, \lambda^k \right\rangle \\
&= \|x\|_1 + \left\langle Ax - b, \lambda^k \right\rangle,
\end{aligned}
$$

因此，对任意的满足$Ax = b$ 的x，都有$\left\|x^k\right\|_1 \leqslant \|x\|_1$. 故$x^k$ 是问题(17) 的最优解.

在论述了**引理1, 引理2**后, 我们可以证明关于简单基追踪问题**(17)**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(17)的可行域非空, 迭代序列$\{x^k\}$, $\{\lambda^k\}$ 是由迭代格式(21) 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(17)的解.

定理的证明

对指标集$\{1,2,\cdots,n\}$的任一划分$\left(I_+^j, I_-^j, E^j\right)$，令

$$U^j \stackrel{\text{def}}{=} U\left(I_+^j, I_-^j, E^j\right) = \left\{x \mid x_i \geqslant 0, i \in I_+^j; x_i \leqslant 0, i \in I_-^j; x_i = 0, i \in E^j\right\},$$

$$H^j \stackrel{\text{def}}{=} \min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|_2^2 \mid x \in U^j\right\}.$$

对于迭代点λ^k，定义指标集$\{1,2,\cdots,n\}$的划分$\left(I_+^{j_k}, I_-^{j_k}, E^{j_k}\right)$为

$$I_+^{j_k} = \left\{i : \left(A^{\mathrm{T}}\lambda^k\right)_i = -1\right\},$$

$$I_-^{j_k} = \left\{i : \left(A^{\mathrm{T}}\lambda^k\right)_i = 1\right\},$$

$$E^{j_k} = \left\{i : \left(A^{\mathrm{T}}\lambda^k\right)_i \in (-1,1)\right\}.$$

并由U^j的定义和$-A^{\mathrm{T}}\lambda^k \in \partial\left\|x^k\right\|_1$知，$x^k \in U^{j_k}$.
因为问题(17)的可行域非空，故存在$\tilde{x}$满足$\|A\tilde{x}-b\| = 0$.
由引理**1** 的**(b)**，对任意满足$H^j > 0$的j，存在一个充分大的K_j使得

$$x^k \notin U^j, \quad \forall k \geqslant K_j.$$

因此, 我们取$K=\max_j\left\{K_j \mid H^j>0\right\}$, 即有

$$H^{j_k}=0, \quad \forall k \geqslant K.$$

结合$\|x\|_1$ 的凸性和(22) 式, 对$k \geqslant K$ 即有

$$\left\|x^k\right\|_1+\left(\lambda^k\right)^{\mathrm{T}} A x^k \leqslant\|x\|_1+\left(\lambda^k\right)^{\mathrm{T}} A x.$$

由于$H^{j_k}=0$, 取$\tilde{x} \in U^{j_k}$ 且$\|A \tilde{x}-b\|=0$, 根据x^k 的最优性, 得到

$$\frac{\sigma}{2}\left\|A x^k-b\right\|^2 \leqslant\|\tilde{x}\|_1-\left\|x^k\right\|_1+\left(\lambda^k\right)^{\mathrm{T}} A\left(\tilde{x}-x^k\right)+\frac{\sigma}{2}\|A \tilde{x}-b\|^2 \leqslant 0.$$

因此由上式联合引理**2**可知, $x^k(\forall k \geqslant K)$都是问题(17) 的最优解.
至此, 简单基追踪问题的ALM收敛性定理证毕.

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(17)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(21)

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

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(21)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(19):

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

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):

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

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

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

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

$$\begin{cases} y^{k+1} = \arg\min_y \left\{ b^\mathrm{T} y + \frac{\sigma}{2} \left\| \psi\left(A^\mathrm{T} y + \frac{\lambda}{\sigma} \right) \right\|_2^2 \right\}, \\ \lambda^{k+1} = \sigma_k \psi\left(A^\mathrm{T} y^{k+1} + \frac{\lambda^k}{\sigma_k} \right), \\ \sigma_{k+1} = \min\left\{ \rho\sigma_k, \bar{\sigma} \right\}. \end{cases} \tag{26}$$

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.

根据凸问题的增广拉格朗日函数法的收敛性定理 (Page 30), 成立对偶问题的收敛性定理.

定理

BP对偶问题增广拉格朗日函数法

假设$\{y^k\}, \{\lambda^k\}$ 是由迭代格式(26)产生的序列, 并且y^{k+1} 的求解精度满足(14)式的不精确条件, 而矩阵 A 是行满秩的. 那么,
(1) 序列$\{y^k\}$ 是有界的, 且其任一聚点均为问题(19)的最优解; *(2)* 序列$\{\lambda^k\}$ 有界且收敛, 其极限为原始问题(17)的某个最优解.

我们对上述2个定理的关系做简单的介绍. 记$\phi_k(y) = L_{\sigma_k}(y, \lambda^k)$, 则根据凸问题的增广拉格朗日函数法, 若要保证算法收敛, 需

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

其中ε_k是人为设定的参数. 当然, ϕ_k凸但非强凸, 因此可按课本的"添项法", 利用不精确条件将其整合为强凸函数, 再进行证明.

提纲

一般线性规划问题的增广拉格朗日函数法

我们介绍基追踪问题及其对偶问题的增广拉格朗日函数法不仅仅是为了向读者展示该方法在一类问题中的应用，更有实际意义的是，它展示了一类处理一般线性规划问题的基本思路.

一般的线性规划问题可描述为

$$\min_{x} \quad c^{\mathrm{T}}x, \qquad \textbf{s.t.} \quad Ax = b, \qquad x \geqslant 0.$$

其对偶问题为

$$\max_{y} \quad b^{\mathrm{T}}y, \qquad \textbf{s.t.} \quad A^{\mathrm{T}}y + s = c, \qquad s \geqslant 0.$$

以上问题中$A \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n, s \in \mathbb{R}^n, y \in \mathbb{R}^m$.
首先还是先写出对偶问题的增广拉格朗日函数

$$L_{\sigma_k}(y, s, \lambda) = -b^{\mathrm{T}}y + \left(\lambda^k\right)^{\mathrm{T}}\left(A^{\mathrm{T}}y + s - c\right) + \frac{\sigma_k}{2}\left\|A^{\mathrm{T}}y + s - c\right\|_2^2, \quad s \geqslant 0.$$

因此根据对偶问题的增广拉格朗日函数, 设计迭代算法为

$$\begin{cases} \left(y^{k+1}, s^{k+1}\right) = \arg \min\limits_{y \in \mathbb{R}^n, s \geqslant 0} L_{\sigma_k}\left(y, s, \lambda^k\right) \\ \qquad\qquad\qquad = \arg \min\limits_{y \in \mathbb{R}^n, s \geqslant 0} \left\{ -b^{\mathrm{T}} y + \dfrac{\sigma_k}{2} \left\| A^{\mathrm{T}} y + s - c + \dfrac{\lambda^k}{\sigma_k} \right\|_2^2 \right\}, \\ \lambda^{k+1} = \lambda^k + \sigma_k \left(A^{\mathrm{T}} y^{k+1} + s^{k+1} - c\right), \\ \sigma_{k+1} = \min\left\{\rho \sigma_k, \bar{\sigma}\right\}. \end{cases} \tag{27}$$

并且还是用最优性条件将s用y表示(消元法), 得到上述关于s的二次型函数中使得s极小化的解为

$$s = \operatorname*{argmin}_{s \geqslant 0} \left\| A^{\mathrm{T}} y + s - c + \frac{\lambda}{\sigma} \right\|_2^2 = \mathcal{P}_{\mathbb{R}^n_+}\left(c - A^{\mathrm{T}} y - \frac{\lambda}{\sigma}\right).$$

将上述s解的表达式代入上述迭代算法, 可得约减式

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathbb{R}^n, s\geqslant 0} \left\{ -b^{\mathrm{T}} y + \frac{\sigma_k}{2} \left\| \psi\left(y, \lambda^k, \sigma_k\right) \right\|_2^2 \right\}, \\ \lambda^{k+1} = \sigma_k \psi\left(y^{k+1}, \lambda^k, \sigma_k\right), \\ \sigma_{k+1} = \min\left\{\rho\sigma_k, \bar{\sigma}\right\}. \end{cases}$$

其中$\psi\left(y, \lambda^k, \sigma_k\right) = \mathcal{P}_{\mathbb{R}^n_+}\left(A^{\mathrm{T}} y + \frac{\lambda^k}{\sigma_k} - c\right)$.

这样, 我们仍利用梯度法或半光滑牛顿法即可求解上述的迭代算法, 以最终解决一般线性规划问题.

提纲

在线性规划问题后，我们考虑半定规划问题：

$$
\begin{aligned}
\min_{X\in\mathcal{S}^n} \quad & \langle C, X\rangle, \\
\text{s.t.} \quad & \langle A_i, X\rangle = b_i, i=1,2,\cdots,m, \\
& X \succeq 0.
\end{aligned} \tag{28}
$$

引入乘子$\lambda\in\mathbb{R}^m$，罚因子σ，记$\mathcal{A}(X)=\left(\langle A_1,X\rangle,\langle A_2,X\rangle,\cdots,\langle A_m,X\rangle\right)^{\mathrm{T}}$，则增广拉格朗日函数为

$$
L_\sigma(X,\lambda)=\langle C,X\rangle-\lambda^{\mathrm{T}}(\mathcal{A}(X)-b)+\frac{\sigma}{2}\|\mathcal{A}(X)-b\|_2^2, \quad X\succeq 0.
$$

根据增广拉格朗日函数，写出算法的迭代格式为

$$
\left\{
\begin{aligned}
X^{k+1} &\approx \underset{X\in\mathcal{S}_+^n}{\operatorname{argmin}} L_{\sigma_k}\left(X,\lambda^k\right), \\
\lambda^{k+1} &= \lambda^k-\sigma_k\left(\mathcal{A}\left(X^{k+1}\right)-b\right), \\
\sigma_{k+1} &= \min\left\{\rho\sigma_k,\bar{\sigma}\right\}.
\end{aligned}
\right.
$$

当迭代收敛时，X^k 和λ^k 分别收敛到原问题和对偶问题的解。

考虑问题(28)的对偶问题:

$$\begin{aligned} \min_{y\in\mathbb{R}^m} \quad & -b^{\mathrm{T}}y, \\ \text{s.t.} \quad & \sum_{i=1}^{m} y_i A_i \preceq C. \end{aligned} \tag{29}$$

引入松他变量$S \succeq 0$, 乘子$\Lambda \in \mathcal{S}^n$以及罚因子$\sigma$, 则增广拉格朗日函数为

$$L_\sigma(y,S,\Lambda) = -b^{\mathrm{T}}y + \left\langle \Lambda, \sum_{i=1}^{m} y_i A_i + S - C \right\rangle + \frac{\sigma}{2}\left\| \sum_{i=1}^{m} y_i A_i + S - C \right\|_F^2 .$$

根据上述函数, 增广拉格朗日函数法的迭代格式为

$$\left(y^{k+1}, S^{k+1}\right) \approx \underset{y\in\mathbb{R}^m}{\arg\min} L_{\sigma_k}\left(y,S,\Lambda^k\right),$$

$$\Lambda^{k+1} = \Lambda^k + \sum_{i=1}^{m} y_i^{k+1} A_i + S^{k+1} - C,$$

$$\sigma_{k+1} = \min\left\{\rho\sigma_k, \bar{\sigma}\right\}.$$

类似地，我们利用消元法，由最优性条件消去S，得

$$L_\sigma(y,\Lambda) = -b^\mathrm{T}y + \frac{\sigma}{2}\left(\left\|\mathcal{P}_{\mathcal{S}_+^n}\left(\sum_{i=1}^m y_iA_i - C + \frac{\Lambda}{\sigma}\right)\right\|_F^2 - \frac{\|\Lambda\|_F^2}{\sigma^2}\right).$$

其中$\mathcal{P}_{\mathcal{S}_+^n}$为到半定锥集$\mathcal{S}_+^n$的投影算子.

因此迭代算法的格式变为

$$\begin{aligned}
y^{k+1} &\approx \mathop{\arg\min}_{y\in\mathbb{R}^m} L_{\sigma_k}\left(y,\Lambda^k\right),\\
\Lambda^{k+1} &= \sigma\mathcal{P}_{\mathcal{S}_+^n}\left(\sum_{i=1}^m y_i^{k+1}A_i - C + \frac{\Lambda^k}{\sigma_k}\right),\\
\sigma_{k+1} &= \min\left\{\rho\sigma_k, \bar{\sigma}\right\}.
\end{aligned}$$

基于原问题和对偶问题的算法比较

我们对比半定规划及其对偶问题的增广拉格朗日函数法.

- 半定规划问题的迭代格式中X^{k+1}是在半正定锥$\mathcal{S}_+^n$中求解, 还是约束优化;
- 对偶问题中y^{k+1}在$\mathbb{R}^m$中求解, 对应于可微的无约束优化问题.

注: 容易验证$L_{\sigma_k}(y,\Lambda^k)$关于$y$连续可微, 因此可用梯度法求$y^{k+1}$.
因此, 若m较小, 我们推荐考虑对偶问题的增广拉格朗日函数法, 以求解半定规划问题.

更进一步, 我们指出, 实际上$L_{\sigma_k}(y,\Lambda^k)$关于$y$还是强半光滑的, 因此我们考虑对偶问题时还可以用半光滑牛顿法更快速地求解半定规划问题.