MATH 3011 运筹学 **秋季 2023**

Lecture 17: 约束优化 增广拉格朗日函数法

Lecturer: 陈士祥 *Scribes:* 陈士祥

1 问题形式

考虑一般的约束优化问题，可以写成

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i(x) = 0, i \in \mathcal{E}, \\ & c_i(x) \leqslant 0, i \in \mathcal{I}. \end{aligned} \tag{17.1}$$

目标函数 $f(x)$ 和 $c_i(x), i \in \mathcal{E}, \mathcal{I}$ 都是可微的。

2 等式约束问题的增广拉格朗日函数法

2.1 二次罚函数法的数值困难

对于等式约束问题:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned} \tag{17.2}$$

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 σ_k 趋于 ∞, 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, 从而避免了上述困难.

2.2 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.

我们首先考虑 等式约束问题的增广拉格朗日函数。

Definition 17.1 *对于等式约束问题(17.2), 定义***增广拉格朗日函数***为:*

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

即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

$$\nabla_x L_{\sigma_k}\left(x^{k+1},\lambda^k\right)=\nabla f\left(x^{k+1}\right)+\sum_{i\in\mathcal{E}}\left(\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right)\right)\nabla c_i\left(x^{k+1}\right)=0. \tag{17.3}$$

我们将(17.3)式对比优化问题(17.2)满足的 KKT 条件 (对最优解 (x^*,λ^*) 的梯度条件)

$$\nabla f\left(x^*\right)+\sum_{i\in\mathcal{E}}\lambda_i^*\nabla c_i\left(x^*\right)=0, \tag{17.4}$$

为保证(17.3)和(17.4)式在最优解处的一致性, 对充分大的 k, 应满足:

$$\lambda_i^*\approx\lambda_i^k+\sigma_k c_i\left(x^{k+1}\right),\quad \forall i\in\mathcal{E}, \tag{17.5}$$

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 $1/\sigma_k$. 如此，避免了 σ_k 趋向无穷大，从而避免了数值困难。
- (17.5)启发我们这样更新 λ^k:

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

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.

在这之前我们先看一个数值例子，与二次罚函数对比，说明增广拉格朗日函数的优势。

Example 17.1 我们考虑优化问题

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

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.

下图中标“$*$”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

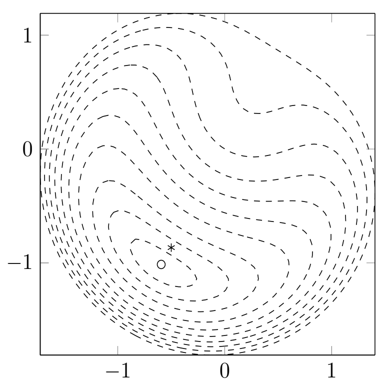

图 17.1: (a) 二次罚函数等高线

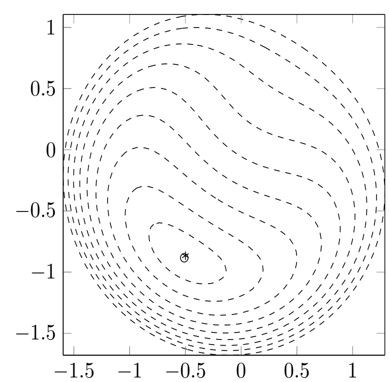

图 17.2: (b) 增广拉格朗日函数等高线

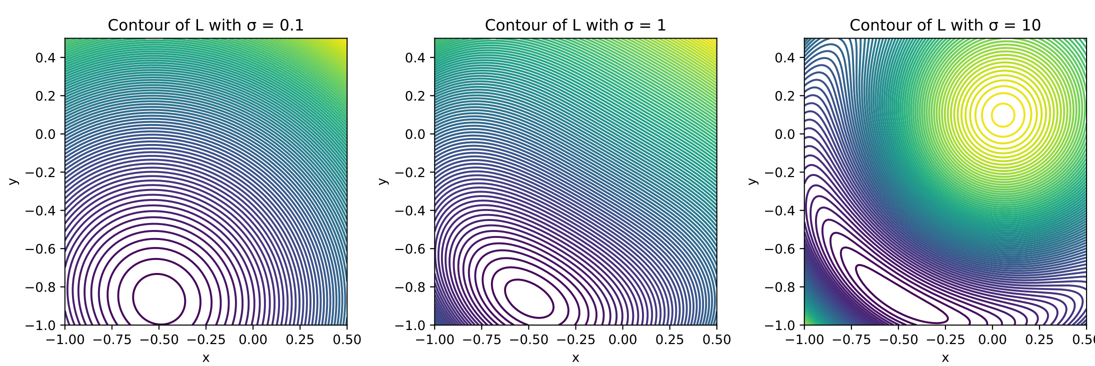

图 17.3: 增广拉格朗日函数 $L_\sigma(x,y,0.95)$ 等高线。紫色表示小的函数值,绿色表示大的函数值。三种情况的局部最优: $\sigma=0.1:[-0.54924849-0.95132628]$; Critical point for $\sigma=1:[-0.52528898-0.9098272]$; Critical point for $\sigma=10:[-0.50453049-0.87387245]$.

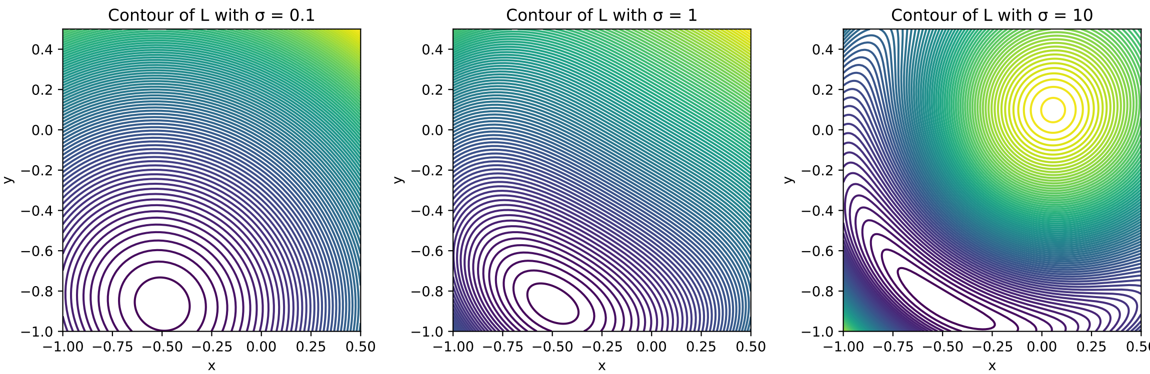

图 17.4: 增广拉格朗日函数 $L_\sigma(x,y,1)$ 等高线。紫色表示小的函数值，绿色表示大的函数值。此时，$x=-1/2, y=-\sqrt{3}/2$ 是局部最小值点。

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.

- 对偶乘子 λ 更接近 KKT 对处的 λ^* 时，增广拉格朗日函数的局部极小则更接近原问题的局部极小。

由此可见, 成立如下的经验性结论.

结论： 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.

基于上述讨论，我们给出如下等式约束的增广拉格朗日函数法。

Algorithm 1 等式约束的增广拉格朗日函数法 (ALM: Augmented Lagrangian Method)

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 $\eta > 0$, 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}.
4: **if** $\left\|c\left(x^{k+1}\right)\right\| \leqslant \varepsilon$ 且 $\eta_k \leq \eta$ **then**
5: 返回近似解 (x^{k+1}, λ_k), 终止迭代.
6: **end if**
7: 更新对偶乘子: $\lambda^{k+1} = \lambda^k + \sigma_k c\left(x^{k+1}\right)$.
8: 更新罚因子: $\sigma_{k+1} = \rho\sigma_k$. 减小子问题精度 η_k.
9: **end for**

注 17.1
- *算法中第 3 行意思为，我们可以求解关于 x 的子问题一个非精确解。*
- *算法第 4 行为终止条件。若找到一个可行点且为子问题的解，则终止迭代。*
- *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.*

2.3 收敛性结论

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, 增广拉格朗朗日函数在一定条件下将成为精确罚函数.

Theorem 17.1 *严格局部极小解定理*

设 x^*, λ^* 分别为问题 (17.2)*的局部极小解和相应的乘子, 且点 x^* 处满足 LICQ,*

若 x^* 处的二阶充分条件成立. 则: 存在有限的常数 $\bar{\sigma}$, 对任意的 $\sigma \geqslant \bar{\sigma}$, x^* 都是 $L_\sigma\left(x, \lambda^*\right)$ 的严格局部极小解.

反之, 若 x^* 为 $L_\sigma\left(x, \lambda^*\right)$ 的局部极小解且满足 $c_i\left(x^*\right) = 0, i \in \mathcal{E}$, 则 x^* 为问题 (17.2)*的局部极小解.*

Proof: 因为 x^* 为问题 (1) 的局部极小解且二阶充分条件成立, 所以

$$\begin{aligned} \nabla_x L\left(x^*, \lambda^*\right) &= \nabla f\left(x^*\right)+\sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i\left(x^*\right)=0, \\ u^{\mathrm{T}} \nabla_{x x}^2 L\left(x^*, \lambda^*\right) u &> 0, \forall u \in\left\{u \mid \nabla c\left(x^*\right)^{\mathrm{T}} u=0\right\}. \end{aligned} \tag{17.6}$$

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

$$\begin{aligned} \nabla_x L_\sigma\left(x^*, \lambda^*\right) &= \nabla_x L\left(x^*, \lambda^*\right)=0, \\ \nabla_{x x}^2 L_\sigma\left(x^*, \lambda^*\right) &= \nabla_{x x}^2 L\left(x^*, \lambda^*\right)+\sigma \nabla c\left(x^*\right) \nabla c\left(x^*\right)^{\mathrm{T}}. \end{aligned} \tag{17.7}$$

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 σ 成立

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

假设该结论不成立, 则对任意 k 以及 $\sigma_k>0$, 存在 u_k 满足 $\left\|u_k\right\|=1$, 且满足:

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

则

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

因为 $\left\{u_k\right\}$ 为有界序列, 必存在聚点, 设为 u. 那么

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

这与(17.6)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

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

因此, x^* 为原问题(17.2)的一个局部极小解, 证毕. ■

关于算法的收敛性结果，我们这里只给出基于较强假设下的结论，并不讨论一般的收敛结果。

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正则条件, 算法迭代生成的序列 $\{x^k\}$ 会有子列收敛至问题(17.2)的一阶稳定点.

Theorem 17.2 增广拉格朗日函数法的收敛性

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

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

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

该定理中的关于乘子列 $\left\{\lambda^k\right\}$ 是有界假设比较强，需要对具体问题具体证明。我们不过多赘述。

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.

Theorem 17.3 增广拉格朗日函数法的收敛性 *(基于更弱的假设)*
假设 x^*,λ^* 是问题 (17.2) 的满足严格局部二阶最优条件的 KKT 对，且 $LICQ$ 成立。给定 $\{\sigma_k\}$ 是递增序列，那么，存在充分大的常数 $\bar{\sigma}>0$ 和充分小的常数 $\delta>0$, 如果初始点满足

$$\frac{1}{\sigma_0}\left\|\lambda^0-\lambda^*\right\|<\delta, \quad \sigma_0 \geqslant \bar{\sigma},$$

则

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

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.

3 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

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i(x)=0, i \in \mathcal{E}, \\ & c_i(x) \leqslant 0, i \in \mathcal{I}. \end{aligned} \tag{17.8}$$

对于问题(17.8), 我们一般引入松弛变量 $s \in \mathbb{R}^{|\mathcal{I}|}$, 得到如下等价形式:

$$\begin{aligned} \min_{x,s} \quad & f(x), \\ \text{s.t.} \quad & c_i(x)=0, i \in \mathcal{E}, \\ & c_i(x)+s_i=0, i \in \mathcal{I}, \\ & s_i \geqslant 0, i \in \mathcal{I}. \end{aligned} \tag{17.9}$$

这样的做法我们已经用过多次了, 应熟练掌握.

保留关于 s 的非负约束, 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(17.9)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:

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

下面我们将看到，添加这种非负约束变量，子问题同样可以求解。

3.1 求解非负约束的子问题

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:

$$\min_{x,s} \quad L_{\sigma_k}\left(x,s,\lambda^k,\mu^k\right), \qquad \text{s.t.} \quad s \geqslant 0, \tag{17.10}$$

可以得到 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.

1. 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

$$s_i = \max\left\{-\frac{\mu_i}{\sigma_k} - c_i(x), 0\right\}, \quad i \in \mathcal{I}. \tag{17.11}$$

2. 将 s_i 的表达式代入 L_{σ_k} 我们有

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

   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(17.10)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.

3.2 更新对偶乘子

对于问题(17.9), 其最优解 x^*, s^* 和乘子 λ^*, μ^* 需满足 KKT 条件:

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

问题(17.10)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(17.9)和问题(17.10)的 KKT 条件, 易知乘子的更新格式为

$$\begin{aligned}
\lambda_i^{k+1} &= \lambda_i^k + \sigma_k c_i\left(x^{k+1}\right), \quad i \in \mathcal{E}, \\
\mu_i^{k+1} &= \max\left\{\mu_i^k + \sigma_k c_i\left(x^{k+1}\right), 0\right\}, \quad i \in \mathcal{I}.
\end{aligned} \tag{17.12}$$

3.3 约束违反度

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 (17.11) 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.

综上，一般约束的增广拉格朗日函数算法描述见算法 2.

Algorithm 2 一般约束增广拉格朗日函数法 (ALM: Augmented Lagrangian Method)

Require: 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 $k = 0$.

Ensure: 输出 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

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

  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

    更新乘子:

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

    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

    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

4 应用：基追踪问题

基追踪（Basis Pursuit）的应用发展历史与压缩感知（Compressed Sensing）和信号处理的历史紧密相关。下面是其应用发展的简要概述：

1. 20 世纪末的理论基础：基追踪的理论基础在 20 世纪末由数学和工程领域的研究者建立。最初，它是作为一种稀疏表示和信号恢复的技术而被探索。
2. 压缩感知的兴起（2000 年代初）：2006 年，Emmanuel Candes、Terence Tao 和 David Donoho 提出了压缩感知理论，这标志着基追踪应用的一个重要转折点。压缩感知理论显示，在某些条件下，可以从远少于 Nyquist 采样定理要求的样本中恢复稀疏信号，而基追踪成为实现这一目标的关键技术之一。
3. 信号处理和图像重构：随后，基追踪在信号处理和图像重构领域得到了广泛应用。它被用于从有限的观测数据中恢复图像和信号，特别是在 MRI（磁共振成像）和雷达成像等领域。
4. 机器学习和数据科学：随着机器学习和数据科学的发展，基追踪也被应用于这些领域，特别是在特征选择和稀疏建模方面。它帮助分析和处理高维数据集，提高了模型的解释性和效率。
5. 算法和计算方法的进步：为了应对基追踪中的计算挑战，研究者开发了多种算法，如正交匹配追踪（OMP）、迭代阈值算法等，提高了问题求解的效率和可行性。
6. 其他领域的应用：基追踪还被应用于无线通信、生物信息学、金融数据分析等领域，展示了其在处理复杂和高维数据问题上的广泛潜力。

总的来说，基追踪的发展和应用反映了现代科学和工程中对高效、稀疏数据表示和处理方法的持续需求。随着技术的不断进步，预计基追踪将在更多领域发挥重要作用。

基追踪是一个数学上的优化问题，它的目的是从一组过完备基（overcomplete basis）中选择出最少的基元素，以便这些基元素的线性组合可以精确地或近似地表示给定的信号。基追踪问题可以被表述为一种特殊的优化问题，其核心在于寻找最稀疏的解。

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

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad Ax = b. \tag{17.13}$$

这里, ℓ_1 范数用来使得 x 尽可能稀疏。

4.1 原问题的增广拉格朗日函数法

根据问题(17.13)的形式, 引入罚因子 σ 和乘子 λ, 其增广拉格朗日函数为

$$L_\sigma(x,\lambda) = \|x\|_1 + \lambda^{\mathrm{T}}(Ax-b) + \frac{\sigma}{2}\|Ax-b\|_2^2. \tag{17.14}$$

固定 σ, 第 k 步迭代更新格式为

$$\left\{\begin{array}{l} x^{k+1}=\underset{x\in\mathbb{R}^n}{\arg\min}\left\{\|x\|_1+\frac{\sigma}{2}\left\|Ax-b+\frac{\lambda^k}{\sigma}\right\|_2^2\right\}, \\ \lambda^{k+1}=\lambda^k+\sigma\left(Ax^{k+1}-b\right). \end{array}\right. \tag{17.15}$$

这里，关于 x 的子问题是一个类似 LASSO（least absolute shrinkage and selection operator）的非光滑问题，我们将会在后面的课程介绍求解方法。

Theorem 17.4 简单基追踪问题的收敛性定理

假设问题 (17.13)的可行域非空, 迭代序列 $\left\{x^k\right\},\left\{\lambda^k\right\}$ 是由迭代格式 (17.15) 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(17.13)的解.

4.2 对偶问题的增广拉格朗日函数法

考虑其对偶问题:

$$\min_{y\in\mathbb{R}^m} \quad b^{\mathrm{T}}y, \quad \text{s.t.} \quad \left\|A^{\mathrm{T}}y\right\|_\infty\leqslant 1. \tag{17.16}$$

通过引入变量 s, 对偶问题可以等价地写成

$$\min_{y\in\mathbb{R}^m,\ s\in\mathbb{R}^n} b^{\mathrm{T}}y, \quad \text{s.t.} \quad A^{\mathrm{T}}y-s=0,\ \|s\|_\infty\leqslant 1. \tag{17.17}$$

现在我们考虑另一重要的问题. 设对偶问题(17.17):

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

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):

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

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

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.

先只考虑关于 s 的极小化问题

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

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

$$\begin{cases} y^{k+1} = \arg\min\limits_{y}\left\{b^{\mathrm{T}}y + \frac{\sigma}{2}\left\|\psi\left(A^{\mathrm{T}}y + \frac{\lambda}{\sigma}\right)\right\|_2^2\right\}, \\ \lambda^{k+1} = \sigma_k \psi\left(A^{\mathrm{T}}y^{k+1} + \frac{\lambda^k}{\sigma_k}\right), \\ \sigma_{k+1} = \min\left\{\rho\sigma_k, \bar{\sigma}\right\}. \end{cases} \tag{17.18}$$

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

性质:

- 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.
- 注意 y 的维度是 m, 原问题 x 的变量维度是 n. 若 $m < n$, 则求解对偶问题更有优势。