



综述

约束优化的内点增广 Lagrange 函数方法

献给韩继业教授 90 寿辰

刘新为^{1*}, 戴彧虹²

1. 河北工业大学数学研究院, 天津 300401;

2. 中国科学院数学与系统科学研究院计算数学与科学工程计算研究所, 北京 100190

E-mail: mathlxw@hebut.edu.cn, dyh@lsec.cc.ac.cn

收稿日期: 2024-04-28; 接受日期: 2024-09-22; 网络出版日期: 2024-12-18; * 通信作者

中国科学院战略性先导科技专项 (A 类) (批准号: XDA27010101) 和国家自然科学基金 (批准号: 12471286, 12071108, 12021001 和 11991021) 资助项目

摘要 内点方法是求解不等式约束优化的一类重要和有效的方法, 而经典增广 Lagrange 函数方法是求解等式约束优化和凸优化的有效方法. 最近的研究表明, 结合内点技术的增广 Lagrange 函数具有很好的高阶光滑性质, 并能够保持优化问题的凸性. 因此该函数既能够用于有效求解非凸优化, 也能够用于有效求解凸优化问题. 本文综述近年来我们关于内点增广 Lagrange 函数方法的系列研究进展, 并对下一步的研究方向进行展望.

关键词 约束优化 非凸优化 凸优化 线性规划 增广 Lagrange 函数 内点方法 交替极小化方法

MSC (2020) 主题分类 90C25, 90C26, 90C30, 90C51

1 引言

考虑一般约束优化

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } h(x) = 0, \\ g(x) \geq 0, \end{aligned} \quad (1.1)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$, $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是 $\mathbb{R}^n$ 上的二次连续可微函数, $n > 0$, $p > 0$, $m > 0$ 是正整数. 另外, $p = 0$ 表明 (1.1) 没有等式约束只有不等式约束, $m = 0$ 表明 (1.1) 是等式约束优化问题. 这里的“一般”是指除可微性外对函数的结构和性质 (如线性性和凸性) 没有特别的要求, 因此其理论和方法可应用于非凸优化和凸优化. 需要特别指出的是, 作为国内非线性优化研究的先驱之一, 韩继

英文引用格式: Liu X W, Dai Y-H. Interior-point augmented Lagrangian methods for constrained optimization (in Chinese). Sci Sin Math, 2025, 55: 535–550, doi: 10.1360/SSM-2024-0132

业教授在涉及非线性规划、不可微优化、变分不等式与互补问题、双层规划、半定规划和组合优化等的运筹学领域作出了突出贡献. 在韩继业教授 90 寿辰来临之际, 祝他健康长寿, 继续为中国运筹学事业作出更大贡献.

增广 Lagrange 函数首先分别由 Hestenes^[11] 和 Powell^[22] 为求解等式约束优化独立提出, 然后由 Rockafellar^[24–26] 扩展到了凸优化. 具体地, 通过在 (1.1) 中引进松弛变量 $s \in \mathbb{R}_+^m$, 可得到一个具有非负约束和等式约束的等价的优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0, \\ & g(x) - s = 0, \quad s \geq 0. \end{aligned} \quad (1.2)$$

然后基于问题 (1.2), 利用等式约束优化的增广 Lagrange 函数将 (1.1) 转化为

$$\begin{aligned} \min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & f(x) - \lambda^T h(x) - \mu^T (g(x) - s) + \frac{1}{2} \rho \|h(x)\|^2 + \frac{1}{2} \rho \|g(x) - s\|^2 \\ \text{s.t.} \quad & s \geq 0, \end{aligned} \quad (1.3)$$

其中 $\rho > 0$ 是罚参数, $\lambda \in \mathbb{R}^p$ 和 $\mu \in \mathbb{R}^m$ 分别为对应于等式和不等式约束的 Lagrange 乘子估计. 经过简单的计算, (1.3) 可以等价地化简为无约束优化

$$\min_{x \in \mathbb{R}^n} L_A(x, \lambda, \mu; \rho) \triangleq f(x) - \lambda^T h(x) + \frac{1}{2} \rho \|h(x)\|^2 + \sum_{i=1}^m \phi_i(x, \mu_i; \rho), \quad (1.4)$$

其中

$$\phi_i(x, \mu_i; \rho) = \begin{cases} -\mu_i g_i(x) + \frac{1}{2} \rho g_i(x)^2, & \text{若 } \rho g_i(x) \leq \mu_i, \\ -\frac{1}{2} \frac{\mu_i^2}{\rho}, & \text{否则.} \end{cases}$$

显然, 由于最后一项的存在, (1.4) 中的目标函数 $L_A(x, \lambda, \mu; \rho)$ 只是关于 x 的一阶光滑函数, 但不是二阶光滑的函数. 它是分片二阶光滑函数. 它的梯度是半光滑函数, 并在适当条件下可以是强半光滑函数 (参见文献 [23]).

内点方法在求解线性优化中取得了巨大成功, 受到了国际上优化专家甚至社会领域工作者的广泛关注 (参见文献 [10, 12, 20, 27, 31]), 纽约时报在 1984 年还为此专门发布了一条消息. 人们很快发现, 内点方法的思想可以追溯到早期的约束优化的对数障碍方法^[21]. 自 20 世纪 80 年代中期以来的三十多年里, 人们对线性和非线性优化的内点方法开展了广泛的研究, 并在凸优化领域取得了显著的进展, 极大促进了优化算法与理论的发展^[19]. 然而, 相对于凸优化, 内点方法在非凸优化领域的进展较为有限 (参见文献 [20, 27]). 对应于优化问题 (1.2) 的对数障碍问题是

$$\begin{aligned} \min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} \quad & f(x) - \sum_{i=1}^m \xi \ln s_i \\ \text{s.t.} \quad & h(x) = 0, \quad g(x) - s = 0, \end{aligned} \quad (1.5)$$

其中 $\xi > 0$ 是障碍参数, 且目标函数只在 $s > 0$ 时才有意义. 注意到优化问题 (1.5) 是等式约束优化, 因此可以写出它的增广 Lagrange 函数, 并将它转化为下列无约束优化:

$$\min_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} f(x) - \sum_{i=1}^m \xi \ln s_i - \lambda^T h(x) - \mu^T (g(x) - s) + \frac{1}{2} \rho \|h(x)\|^2 + \frac{1}{2} \rho \|g(x) - s\|^2. \quad (1.6)$$

通过按分量逐个计算它关于 s_i 极小, 可进一步将它等价地简化为

$$\min_{x \in \mathbb{R}^n} L_B(x, \lambda, \mu; \rho, \xi) \triangleq f(x) - \lambda^T h(x) + \frac{1}{2} \rho \|h(x)\|^2 + \sum_{i=1}^m \psi_i(x, \mu_i; \rho, \xi), \quad (1.7)$$

其中

$$\psi_i(x, \mu_i; \rho, \xi) = -\xi \ln s_i(x, \mu_i; \rho, \xi) - \mu_i(g_i(x) - s_i(x, \mu_i; \rho, \xi)) + \frac{1}{2} \rho (g_i(x) - s_i(x, \mu_i; \rho, \xi))^2,$$

这里

$$s_i(x, \mu_i; \rho, \xi) = \frac{1}{2} \left[\sqrt{\left(g_i(x) - \frac{\mu_i}{\rho}\right)^2 + 4\frac{\xi}{\rho}} + \left(g_i(x) - \frac{\mu_i}{\rho}\right) \right].$$

由于当 $\xi > 0$ 时 $s_i(x, \mu_i; \rho, \xi)$ 和 $\psi_i(x, \mu_i; \rho, \xi)$ 均为 x 的二阶光滑函数, 且当 $\xi = 0$ 时 $s_i(x, \mu_i; \rho, \xi) = \max\{0, g_i(x) - \frac{\mu_i}{\rho}\}$ 和 $\psi_i(x, \mu_i; \rho, \xi) = \phi_i(x, \mu_i; \rho)$, 因此 $\psi_i(x, \mu_i; \rho, \xi)$ 实际上是非二阶光滑函数 $\phi_i(x, \mu_i; \rho)$ 的一种基于内点的高阶光滑化函数, 其中 $\xi > 0$ 既是内点方法的对数障碍参数, 也是非二阶光滑函数的光滑化参数.

称函数 $L_B(x, \lambda, \mu; \rho, \xi)$ 为内点增广 Lagrange 函数. 它除了依赖于原始变量 x 和对偶变量 λ 和 μ 以外, 还依赖于罚参数 ρ 和光滑化参数 ξ . 我们关于非线性约束优化内点增广 Lagrange 函数方法的研究正是基于将原问题 (1.1) 转化为求解极大极小问题

$$\min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m} L_B(x, \lambda, \mu; \rho, \xi) \quad (1.8)$$

提出的. 在求解问题 (1.8) 的过程中, 我们需要根据迭代点处的函数和梯度值信息逐步修正罚参数和光滑化参数, 并在光滑化参数靠近零时终止我们的算法.

本文余下内容的结构如下. 第 2 节给出内点增广 Lagrange 函数 $L_B(x, \lambda, \mu; \rho, \xi)$ 的分析性质, 包括它的一阶和二阶可微性, 以及它对应的极大极小问题 (1.8) 与优化问题 (1.1) 和 (1.5) 的解的关系. 第 3 节给出一个求解非凸约束优化的内点松弛方法及其收敛性结果. 第 4 节给出求解非凸约束优化的一个交替极小化方法. 然后在第 5 节给出内点增广 Lagrange 函数在求解半定规划、二阶锥规划和线性规划等凸优化中的应用. 第 6 节给出本文的结论, 并在最后指出下一步可能的研究方向.

本文中使用的符号均为常用的符号, 所有的向量均为列向量. 特别地, $\|\cdot\|$ 表示 l_2 -范数, 下标 i 表示一个向量函数的第 i 个分量函数, 或者是一个向量的第 i 个分量, 下标 k 表示与第 k 次迭代有关. 矩阵或者向量的转置用上标 T 表示. 符号 $\circ$ 表示两个向量的 Hadamard 积. 除第 5.2 小节外, e 表示所有分量为 1 的向量. 为简单起见, 将形式为 $(x^T, y^T, z^T)^T$ 的向量简记为 (x, y, z) , 记 $f_k = f(x_k)$, $s_i = s_i(x, \mu_i; \rho, \xi)$, $s_{ki} = s_i(x_k, \mu_{ki}; \rho, \xi_k)$, $s_k = (s_{ki}, i = 1, \dots, m)$.

2 内点增广 Lagrange 函数的基本性质

本节的结论可在文献 [14] 或者 [16] 中找到, 或者与其中相关的结论类似地证明.

引理 2.1 (可微性) 给定 $\rho > 0$ 和 $\xi > 0$. 如果 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^p$ 和 $g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 均在 $\mathbb{R}^n$ 上二次连续可微, 则

(1) $L_B(x, \lambda, \mu; \rho, \xi)$ 是关于 x 的二次连续可微函数, 且

$$\nabla_x L_B(x, \lambda, \mu; \rho, \xi) = \nabla f(x) - \nabla h(x)(\lambda - \rho h(x)) - \rho \nabla g(x)z, \quad (2.1)$$

$$\begin{aligned}\nabla_{xx}^2 L_B(x, \lambda, \mu; \rho, \xi) = & \left(\nabla^2 f(x) - \sum_{i=1}^p (\lambda_i - \rho h_i(x)) \nabla^2 h_i(x) - \sum_{i=1}^m \rho z_i \nabla^2 g_i(x) \right) \\ & + \sum_{i=1}^m \rho \nabla h_i(x) \nabla h_i(x)^T + \sum_{i=1}^m \rho \frac{z_i}{s_i + z_i} \nabla g_i(x) \nabla g_i(x)^T,\end{aligned}\quad (2.2)$$

其中

$$z_i = \frac{1}{2} \left[\sqrt{\left(g_i(x) - \frac{\mu_i}{\rho} \right)^2 + 4 \frac{\xi}{\rho}} - \left(g_i(x) - \frac{\mu_i}{\rho} \right) \right],$$

满足 $\rho z = \mu - \rho g(x) + \rho s$;

(2) $L_B(x, \lambda, \mu; \rho, \xi)$ 是关于 λ 和 μ 的二次连续可微函数, 且

$$\nabla_{\lambda} L_B(x, \lambda, \mu, \rho, \xi) = -h(x), \quad \nabla_{\lambda\lambda}^2 L_B(x, \lambda, \mu; \rho, \xi) = 0, \quad (2.3)$$

$$\nabla_{\mu} L_B(x, \lambda, \mu; \rho, \xi) = s - g(x), \quad \nabla_{\mu\mu}^2 L_B(x, \lambda, \mu; \rho, \xi) = -\frac{1}{\rho} (S + Z)^{-1} S, \quad (2.4)$$

其中 $S = \text{diag}(s)$ 和 $Z = \text{diag}(z)$ 是对角矩阵.

引理 2.2 (凸性) 给定 $\rho > 0$ 和 $\xi > 0$.

(1) 若 f 和 $-g_i$ ($i = 1, \dots, m$) 均为 $\mathbb{R}^n$ 上的凸函数, h_i ($i = 1, \dots, p$) 为仿射函数, 则 $L_B(x, \lambda, \mu; \rho, \xi)$ 是关于 x 在 $\mathbb{R}^n$ 上的凸函数;

(2) $L_B(x, \lambda, \mu; \rho, \xi)$ 是关于 λ 的线性函数, 也是关于 μ 的严格凹函数.

引理 2.3 (稳定性) 下列结论成立:

(1) 给定 $\xi > 0$ 和 $\rho > 0$, 令 $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ 是极大极小问题 (1.8) 的一个局部解 (也就是一个鞍点), 则

$$\nabla f(x^*) - \nabla h(x^*) \lambda^* - \nabla g(x^*) \mu^* = 0, \quad (2.5)$$

$$h(x^*) = 0, \quad (2.6)$$

$$g(x^*) - s^* = 0, \quad (2.7)$$

其中 $s^* = (s_i^*, i = 1, \dots, m)$, 且

$$s_i^* = \frac{1}{2} \left[\sqrt{\left(g_i(x^*) - \frac{\mu_i^*}{\rho} \right)^2 + 4 \frac{\xi}{\rho}} + \left(g_i(x^*) - \frac{\mu_i^*}{\rho} \right) \right];$$

(2) 若 (x^*, λ^*, μ^*) 满足方程组 (2.5)–(2.7), 且 $\xi = 0$ 和 $\rho > 0$, 则 (x^*, λ^*, μ^*) 是原始优化问题 (1.1) 的一个 KKT 对.

注意到 s 是 x 和 μ 的函数, 并依赖于参数 ξ 和 ρ . 称满足方程组

$$\nabla f(x) - \nabla h(x) \lambda - \nabla g(x) \mu = 0, \quad h(x) = 0, \quad g(x) - s(x, \mu; \rho, \xi) = 0$$

的点 (x, λ, μ) 为极大极小问题 (1.8) 的稳定点.

引理 2.4 (等价性) 给定 $\xi > 0$ 和 $\rho > 0$, 则 (x^*, λ^*, μ^*) 是极大极小问题 (1.8) 的稳定点当且仅当 $s^* = g(x^*)$, $\mu^* = \xi(S^*)^{-1}e$, 且 $((x^*, s^*), (\lambda^*, \mu^*))$ 是对数障碍问题 (1.5) 的 KKT 对.

定理 2.1 (最优性) 令 x^* 是原问题 (1.1) 的一个局部极小点, 在 x^* 处线性无关约束规范条件满足, 且存在 $\lambda^* \in \mathbb{R}^p$ 和 $\mu^* \in \mathbb{R}_+^m$ 使得 (x^*, λ^*, μ^*) 满足问题 (1.1) 的二阶最优性充分条件. 那么对于 $\lambda = \lambda^*$ 和 $\mu = \mu^*$, 存在常数 $\tilde{\rho} > 0$ 使得对所有 $\rho \geq \tilde{\rho}$ 及充分小的 $\xi > 0$, x^* 是内点增广 Lagrange 函数 $L_B(x, \lambda, \mu; \rho, \xi)$ 的一个 $\sqrt{\rho\xi}$ -近似严格局部极小.

3 一个原始对偶内点松弛方法及收敛性

由引理 2.4, 如果每次迭代近似精确求解问题 (1.8), 则它与现有的许多非线性优化的原始对偶内点方法 [5, 6, 17, 18, 28–30] 类似. 本节提出的算法每次迭代对给定的参数仅求解问题 (1.8) 一个拟 Newton 步, 然后自适应地修正罚参数和对数障碍参数. 它与经典非线性优化原始对偶内点方法的其他显著区别在于, 在充分靠近解之前对偶迭代 μ_k 不必是正的.

定义残差函数

$$\phi_{(\xi, \rho)}(x, \lambda, \mu) = \frac{1}{2} \|\nabla f(x) - \nabla h(x)\lambda - \nabla g(x)\mu\|^2 + \frac{1}{2} \|h(x)\|^2 + \frac{1}{2} \|g(x) - s(x, \mu; \rho, \xi)\|^2.$$

下面给出基于求解极大极小问题 (1.8) 的原始对偶内点松弛方法 (见算法 1).

算法 1 求解非线性约束优化 (1.1) 的原始对偶内点松弛方法

给定 $(x_0, \lambda_0, \mu_0) \in \mathbb{R}^{n+p+m}$, $B_0 \in \mathbb{R}^{n \times n}$, $\rho_0 > 0$, $\eta > 1$, $\gamma_0, \delta, \tau, \sigma \in (0, 1)$, $\epsilon > 0$.

选择 $\xi_0 > 0$ 使得 $\xi_0 \leq \max\{\eta\phi_{(\xi_0, \rho_0)}(x_0, \lambda_0, \mu_0), \epsilon\}$. 令 $\gamma = \min\{\gamma_0, \xi_0/\phi_{(\xi_0, \rho_0)}(x_0, \lambda_0, \mu_0)\}$.

对于 $k = 0, 1, \dots$, 如果 $\xi_k \leq \epsilon$ 且 $\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k) \leq \epsilon$, 算法终止. 否则,

步 1. 计算 $\Delta\xi_k = -\xi_k + \gamma\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k)$;

步 2. 计算 $s_k = s(x_k, \mu_k; \rho_k, \xi_k)$ 和 $z_k = \frac{1}{\rho_k}(\mu_k - g(x_k) + s_k)$, 令 $S_k = \text{diag}(s_k)$ 和 $Z_k = \text{diag}(z_k)$, 解线性方程组

$$\begin{bmatrix} B_k & -\nabla h(x_k) & -\nabla g(x_k) \\ \nabla h(x_k)^T & 0 & 0 \\ (S_k + Z_k)^{-1} Z_k \nabla g(x_k)^T & 0 & \frac{1}{\rho_k} (S_k + Z_k)^{-1} S_k \end{bmatrix} \begin{bmatrix} d_x \\ d_\lambda \\ d_\mu \end{bmatrix} = \begin{bmatrix} -r_k^d \\ -r_k^h \\ -r_k^e + \frac{1}{\rho_k} \Delta\xi_k (S_k + Z_k)^{-1} e \end{bmatrix}$$

得出 $d_k = (d_{xk}, d_{\lambda k}, d_{\mu k})$, 其中

$$r_k^d = \nabla f(x_k) - \nabla h(x_k)\lambda_k - \nabla g(x_k)\mu_k, \quad r_k^h = h(x_k), \quad r_k^e = g(x_k) - s_k;$$

步 3. 选择步长 $\alpha_k \in (0, 1]$ 是 $\{1, \delta, \delta^2, \dots\}$ 的最大数使得

$$\phi_{(\xi_k + \alpha_k \Delta\xi_k, \rho_k)}(x_k + \alpha_k d_{xk}, \lambda_k + \alpha_k d_{\lambda k}, \mu_k + \alpha_k d_{\mu k}) \leq (1 - 2\tau\alpha_k)\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k); \quad (3.1)$$

步 4. 令 $\xi_{k+1} = \xi_k + \alpha_k \Delta\xi_k$, $x_{k+1} = x_k + \alpha_k d_{xk}$, $\mu_{k+1} = \mu_k + \alpha_k d_{\mu k}$, $\lambda_{k+1} = \lambda_k + \alpha_k d_{\lambda k}$;

步 5. 修正 ρ_k 为 $\rho_{k+1} = \max\{\rho_k, \sigma\|\mu_{k+1}\|_\infty / \max(\|x_{k+1}\|, 1)\}$, 从 $\{1, \eta^{-1}, \eta^{-2}, \dots\}$ 中选择最大数 σ' 使得 $\xi'_{k+1} = \sigma'\xi_{k+1}$ 满足 $\xi'_{k+1} \leq \max\{\eta\phi_{(\xi'_{k+1}, \rho_{k+1})}(x_{k+1}, \lambda_{k+1}, \mu_{k+1}), \epsilon\}$, 令

$$\xi_{k+1} = \xi'_{k+1}, \quad \gamma = \min\{\gamma_0, \xi_{k+1}/\phi_{(\xi_{k+1}, \rho_{k+1})}(x_{k+1}, \lambda_{k+1}, \mu_{k+1})\};$$

步 6. 修正 B_k 为 B_{k+1} .

注 3.1 对某个点 (x_k, λ_k, μ_k) 及参数 $\rho_k > 0$, 可能有 $\xi > \eta\phi_{(\xi, \rho_k)}(x_k, \lambda_k, \mu_k)$ 对所有 $\xi > 0$ 成立. 可以证明, 在这种情形下, (x_k, λ_k, μ_k) 是原问题 (1.1) 的一个 KKT 对. 因此, 不失一般性, 假定 $\xi_k \leq \eta\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k)$ 对所有 $k \geq 0$ 成立.

本节的收敛性理论主要来源于文献 [14]. 为了保证算法 1 是可运行的, 除了目标函数 f 和约束函数 h 及 g 的连续可微性假设外, 我们还需要下列关于迭代点的一般性假设.

假设 3.1 假定对每个 $k \geq 0$, 在 (x_k, λ_k, μ_k) 点处下列条件成立:

(1) Jacobi 矩阵 $\nabla h(x_k)$ 是列满秩的;

(2) 对称矩阵 $B_k + \sum_{i=1}^m \frac{\rho_k^2}{\xi_k} z_{ki} \nabla g_i(x_k) \nabla g_i(x_k)^T$ 在零空间 $\{d_x : \nabla h(x_k)^T d_x = 0\}$ 上是正定的.

在满足假设 3.1 条件下, 若取 $\epsilon = 0$, 则算法 1 将产生一个无穷序列 $\{(x_k, \lambda_k, \mu_k)\}$. 为了建立其全局收敛性结果, 需要进一步加强假设 3.1.

假设 3.2 假定下列条件成立:

- (1) 序列 $\{x_k\} \subseteq \Omega$, 其中 Ω 是 $\mathbb{R}^n$ 上的一个有界开子集;
- (2) 矩阵 $\nabla h(x_k)$ 是对所有 $k \geq 0$ 都是列满秩的;
- (3) 矩阵序列 $\{B_k\}$ 是有界的, 且存在常数 $\chi > 0$ 使得对所有 $k \geq 0$ 及 $\{d_x : \nabla h(x_k)^T d_x = 0\}$, 都有

$$d_x^T \left(B_k + \sum_{i=1}^m \frac{\rho_k^2}{\xi_k} z_{ki} \nabla g_i(x_k) \nabla g_i(x_k)^T \right) d_x \geq \chi \|d_x\|^2.$$

定理 3.1 在假设 3.2 成立条件下, 如果对充分大的 k , ρ_k 保持不变, 则下列结论之一成立:

- (1) 对某个 $k \geq 0$ 和所有 $\xi > 0$, 有 $\xi > \eta\phi_{(\xi, \rho_k)}(x_k, \lambda_k, \mu_k)$, (x_k, λ_k, s_k) 是原问题 (1.1) 的一个 KKT 对;
- (2) 对所有充分大的 k , $\xi_k \leq \eta\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k)$, 且 $\{\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k)\}$ 单调下降, 则

$$\phi_{(\xi_k, \rho_k)}(x_k, \lambda_k, \mu_k) \rightarrow 0, \quad \xi_k \rightarrow 0, \quad \text{当 } k \rightarrow \infty \text{ 时},$$

即无穷序列 $\{(x_k, \lambda_k, \mu_k)\}$ 的每个聚点都是原问题 (1.1) 的 KKT 对.

为了分析局部收敛性, 需要下列更强的假设, 其中 (3), (4) 和 (5) 分别对应于线性无关约束规范条件、严格互补松弛条件和强二阶充分条件. 这些假设对于非线性约束优化算法的局部收敛性分析是标准的, 但对算法作适当的修改后可以适当地放松. 为简单起见, 我们对此不作更多的讨论.

假设 3.3 假定下列条件成立:

- (1) 当 $k \rightarrow \infty$ 时, $(x_k, \lambda_k, \mu_k) \rightarrow (x^*, \lambda^*, \mu^*)$ 和 $\xi_k \rightarrow 0$;
- (2) 函数 f , h 和 g 在 x^* 的某个邻域里二次 Lipschitz 连续可微;
- (3) 积极约束梯度向量 $\nabla h_i(x^*)$ ($i = 1, \dots, p$) 和 $\nabla g_i(x^*)$ ($i \in \{i = 1, \dots, m : g_i(x^*) = 0\}$) 线性无关;
- (4) 等式 $g(x^*) + \mu^* > 0$ 成立;
- (5) 对所有非零向量 $d \in \{d : \nabla h(x^*)^T d = 0, \nabla g_i(x^*)^T d = 0, i \in \{i = 1, \dots, m : g_i(x^*) = 0\}\}$, Hessian 矩阵 B^* 满足 $d^T B^* d > 0$, 其中

$$B^* = \nabla^2 f(x^*) - \sum_{i=1}^p \lambda_i^* \nabla^2 h_i(x^*) - \sum_{i=1}^m \mu_i^* \nabla^2 g_i(x^*).$$

为简单起见, 令 $w_k = (x_k, \lambda_k, \mu_k)$, $w^* = (x^*, \lambda^*, \mu^*)$.

定理 3.2 在假设 3.3 成立的条件下, 存在充分小的数 $\epsilon > 0$, 使得对所有 $(\xi_k, w_k) \in \{(\xi, w) \in \mathbb{R}_{++} \times \mathbb{R}^{n+p+m} : \|(\xi, w) - (0, w^*)\| < \epsilon\}$, 下列结论成立:

- (1) 如果对每个 $d_x \in \mathbb{R}^n$ 有 $\|(B_k - B^*)d_x\| = O(\|d_x\|^2)$, 则

$$\|(\xi_k, w_k) + (\Delta\xi_k, d_k) - (0, w^*)\| = O(\|(\xi_k, w_k) - (0, w^*)\|^2),$$

即 $(\Delta\xi_k, d_k)$ 是一个二阶收敛步;

- (2) 如果对每个 $d_x \in \mathbb{R}^n$ 有 $\|(B_k - B^*)d_x\| = o(\|d_x\|)$, 则

$$\|(\xi_k, w_k) + (\Delta\xi_k, d_k) - (0, w^*)\| = o(\|(\xi_k, w_k) - (0, w^*)\|).$$

它表明 $(\Delta\xi_k, d_k)$ 是一个超线性收敛步.

定理 3.3 在假设 3.3 成立条件下, 如果对每个 $d_x \in \mathbb{R}^n$ 有 $\|(B_k - B^*)d_x\| = O(\|d_x\|^2)$, 且 $\tau < 1/2$, 则 $\xi_{k+1} = \xi_k$, 或者 $\xi_{k+1} = \gamma_0 \phi(\xi_k, \rho^*)(w_k)$, 且 $x_{k+1} = x_k + d_{xk}$, $\mu_{k+1} = \mu_k + d_{\mu k}$, $\lambda_{k+1} = \lambda_k + d_{\lambda k}$ 对所有充分大的 k 成立, 以及 $\|w_{k+1} - w^*\| = O(\|w_k - w^*\|^2)$.

算法 1 已被应用于求解二次凸优化 (含线性优化) 并被证明是有效的 (参见文献 [33]).

4 一个交替极小化内点方法

与极大极小问题 (1.8) 相对应, 考虑极小极大问题

$$\max_{\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} L_B(x, \lambda, \mu; \rho, \xi), \quad (4.1)$$

并称其为问题 (1.8) 的对偶问题. 由弱对偶定理可知, 问题 (1.8) 与 (4.1) 满足下列关系:

$$\max_{\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} L_B(x, \lambda, \mu; \rho, \xi) \leq \min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^m} L_B(x, \lambda, \mu; \rho, \xi).$$

基于求解问题 (4.1), Liu 等^[16] 提出了求解问题 (1.1) 的交替极小化方法, 也称为内点增广 Lagrange 乘子方法 (见算法 2).

算法 2 求解非线性约束优化 (1.1) 的交替极小化方法

给定 $(x_0, \lambda_0, \mu_0) \in \mathbb{R}^{n+p+m}$, $\xi_0 \in (0, 1)$, $\rho_0 \in [1, \infty)$, $\gamma \in (0, 1)$, $\epsilon \in (0, \mu_0)$. 计算 KKT 方程残差

$$E_0^{\text{dual}} = \frac{1}{\rho_0} \|\nabla f(x_0) - \nabla h(x_0)\lambda_0 - \nabla g(x_0)\mu_0\|_\infty, \\ E_0^{\text{compl}} = \frac{1}{\rho_0} \|\mu_0 \circ g(x_0)\|_\infty, \quad E_0^{\text{feas}} = \|(h(x_0), \max\{0, -g(x_0)\})\|_\infty,$$

以及不可行稳定性残差度量

$$E_0^{\text{ifs}} = \|\nabla h(x_0)h(x_0) + \nabla g(x_0) \max\{0, -g(x_0)\}\|_\infty.$$

对于 $k = 0, 1, \dots$, 如果 $\max\{E_k^{\text{dual}}, E_k^{\text{compl}}, E_k^{\text{feas}}\} < \epsilon$, 或者 $E_k^{\text{feas}} > \epsilon$ 且 $E_k^{\text{ifs}} < \epsilon$, 终止算法. 否则, 循环计算步 1–4.

步 1 (得到原始变量估计). 对给定的 $\lambda_k \in \mathbb{R}^m$ 及参数 μ_k 和 ρ_k , 以 x_k 为初始点, 寻找光滑无约束优化

$$\min_x \frac{1}{\rho_k} L_B(x, \lambda_k, \mu_k; \rho_k, \xi_k) \quad (4.2)$$

的近似解 x_{k+1} 直到满足

$$\|\nabla_x L_B(x_{k+1}, \lambda_k, \mu_k; \rho_k, \xi_k)\|_\infty (= \|\nabla f(x_{k+1}) - \nabla h(x_{k+1})\hat{\lambda}_{k+1} - \rho_k \nabla g(x_{k+1})\hat{z}_{k+1}\|_\infty) \\ \leq 0.95 \max\{\rho_k \epsilon, \rho_k^{(1-\gamma)} \xi_k\}, \quad (4.3)$$

其中 $\hat{z}_{k+1} = z(x_{k+1}, \mu_k; \rho_k, \xi_k)$, $\hat{\lambda}_{k+1} = \lambda_k - h(x_{k+1})$.

步 2 (得到对偶变量估计). 计算 $\hat{s}_{k+1} = s(x_{k+1}, \mu_k; \rho_k, \xi_k)$. 令

$$\tilde{\mu}_{k+1} = \rho_k \hat{z}_{k+1} (= \mu_k + \rho_k(\hat{s}_{k+1} - g(x_{k+1}))). \quad (4.4)$$

步 3 (修正参数). 计算 $\tilde{s}_{k+1} = s(x_{k+1}, \tilde{\mu}_{k+1}; \rho_k, \xi_k)$. 令 $\tilde{r}_{k+1} = \|\tilde{s}_{k+1} - g(x_{k+1})\|_\infty$. 若 $\tilde{r}_{k+1} > 0.95\xi_k$ 或者 $\|h(x_{k+1})\|_\infty > 0.95\|h(x_k)\|_\infty$, 则取 $\rho_{k+1} \geq 2\rho_k$, $\xi_{k+1} = \xi_k$ 及 $\mu_{k+1} = \mu_k$, $\lambda_{k+1} = \lambda_k$; 否则, 令 $\xi_{k+1} \leq 0.1\xi_k$, $\rho_{k+1} = \max\{\rho_k, \|(\hat{\lambda}_{k+1}, \tilde{\mu}_{k+1})\|_\infty\}$ 及 $\mu_{k+1} = \tilde{\mu}_{k+1}$, $\lambda_{k+1} = \hat{\lambda}_{k+1}$.

步 4 (计算残差). 计算 $s_{k+1} = s(x_{k+1}, \mu_{k+1}; \rho_{k+1}, \xi_{k+1})$, $r_{k+1} = \|s_{k+1} - g(x_{k+1})\|_\infty$, 以及

$$E_{k+1}^{\text{dual}} = \frac{1}{\rho_{k+1}} \|\nabla f(x_{k+1}) - \nabla h(x_{k+1})\lambda_{k+1} - \nabla g(x_{k+1})\mu_{k+1}\|_\infty, \quad E_{k+1}^{\text{compl}} = \frac{1}{\rho_{k+1}} \|\mu_{k+1} \circ g(x_{k+1})\|_\infty, \\ E_{k+1}^{\text{feas}} = \|(h(x_{k+1}), \max\{0, -g(x_{k+1})\})\|_\infty, \quad E_{k+1}^{\text{ifs}} = \|\nabla h(x_{k+1})h(x_{k+1}) + \nabla g(x_{k+1}) \max\{0, -g(x_{k+1})\}\|_\infty.$$

注 4.1 由于我们不预先假定极小化子问题 (4.2) 总是可达到最优解的, 也不假定保证极小化子问题 (4.2) 是可解的任何充分条件, 算法 2 可在极小化子问题 (4.2) 时终止. 对任意给定的 λ_k, μ_k, ξ_k 和 ρ_k , 考虑用线搜索方法求解问题 (4.2):

$$x_{k,j+1} = x_{k,j} + \alpha_{k,j} d_{k,j}, \quad (4.5)$$

其中

$$d_{k,j} = -\frac{1}{\rho_k} B_{k,j}^{-1} \nabla_x L_B(x_{k,j}, \lambda_k, \mu_k; \rho_k, \xi_k).$$

为简单起见, 假定 $B_{k,j} \in \mathbb{R}^{n \times n}$ 对所有子迭代 j 一致正定, 选取 $\alpha_{k,j} \in (0, 1]$ 满足 Armijo 线搜索条件

$$L_B(x_{k,j} + \alpha_{k,j} d_{k,j}, \lambda_k, \mu_k; \rho_k, \xi_k) \leq L_B(x_{k,j}, \lambda_k, \mu_k; \rho_k, \xi_k) + \sigma \alpha_{k,j} \nabla_x L_B(x_{k,j}, \lambda_k, \mu_k; \rho_k, \xi_k)^T d_{k,j},$$

其中 $\sigma \in (0, 1)$ 是一个常数. 参照著名的 Zoutendijk 全局收敛性定理^[20, 27], 有如下结果:

引理 4.1 假定函数 f, h 和 g 均在 $\mathbb{R}^n$ 上两次连续可微, 且 f 在 $\mathbb{R}^n$ 上有下界. 如果问题 (4.2) 由线搜索迭代方法 (4.5) 求解, 则迭代 (4.5) 将在有限步内满足条件 (4.3).

由引理 4.1, 若取 $\epsilon = 0$, 则类似于算法 1, 我们将得到一个无穷序列 $\{x_k\}$. 本节接下来将给出关于该序列的全局和局部收敛性结果, 详细证明可参见文献 [16].

定理 4.1 假定函数 f, h 和 g 均在 $\mathbb{R}^n$ 上两次连续可微, f 在 $\mathbb{R}^n$ 上有下界, $\epsilon = 0$, $\{x_k\}$ 是由算法 2 产生的无穷序列. 如果序列 $\{x_k\} \subseteq \Omega$, 其中 Ω 是 $\mathbb{R}^n$ 上的一个有界开子集, 则下列 3 个结论之一成立:

(1) 对所有充分大的 k , 有 $\xi_{k+1} = \xi_{k_0} > 0$, $r_{k+1} > 0.95\xi_k$ 或者 $\|h(x_{k+1})\|_\infty > 0.95\|h(x_k)\|_\infty$, 当 $k \rightarrow \infty$ 时 $\rho_k \rightarrow \infty$, 序列 $\{x_k\}$ 的任何聚点 x^* 是问题 (1.1) 的不可行点, 且满足

$$\nabla h(x^*)h(x^*) + \nabla g(x^*) \max\{-g(x^*), 0\} = 0,$$

即 x^* 是问题 (1.1) 的一个不可行稳定点.

(2) 在某些迭代点 x_{k+1} 处, $r_{k+1} > 0.95\xi_k$ 或者 $\|h(x_{k+1})\|_\infty > 0.95\|h(x_k)\|_\infty$; 在一些其他迭代点 $x_{k'+1}$ 处, 却是 $r_{k'+1} \leq 0.95\xi_{k'}$ 及 $\|h(x_{k'+1})\|_\infty \leq 0.95\|h(x_{k'})\|_\infty$ ($k' > k, k \rightarrow \infty, k' \rightarrow \infty$), 两种情形交替反复出现, 且当 $k \rightarrow \infty$ 时有 $\rho_k \rightarrow \infty$ 和 $\xi_k \rightarrow 0$, 存在序列 $\{x_k\}$ 的聚点 x^* 及相应的向量 $\lambda^* \in \mathbb{R}^p$ 和 $z^* \in \mathbb{R}^m$ 使得

$$\nabla h(x^*)\lambda^* + \nabla g(x^*)z^* = 0, \quad h(x^*) = 0, \quad z^* \geq 0, \quad g(x^*) \geq 0, \quad (z^*)^T g(x^*) = 0,$$

即 x^* 是问题 (1.1) 的一个可行点且是一个退化稳定点.

(3) 对所有充分大的 k , 有 $r_{k+1} \leq 0.95\mu_k$ 及 $\|h(x_{k'+1})\|_\infty \leq 0.95\|h(x_{k'})\|_\infty$. 当 $k \rightarrow \infty$ 时有 $\xi_k \rightarrow 0$. 若存在某个正整数 k_0 使得对所有 $k \geq k_0$ 有 $\rho_k = \rho_{k_0}$, 则序列 $\{x_k\}$ 的每个聚点都是问题 (1.1) 的 KKT 点. 若当 $k \rightarrow \infty$ 时有 $\rho_k \rightarrow \infty$, 则存在序列 $\{x_k\}$ 的一个聚点 x^* 是问题 (1.1) 的一个退化稳定点.

上述定理表明算法 2 具有强全局收敛性, 它是在极弱条件下的非线性约束优化算法所能具有的全局收敛性质^[3, 7, 17, 18, 32]. 为了分析算法的局部收敛性, 需要更多关于迭代序列的假设条件. 基于近年来非线性约束优化在快速探测不可行性算法的研究进展^[2, 4, 8], 我们不仅关注算法 2 关于 KKT 点的局部快速收敛性 (对应于定理 4.1(3) 中的前一种情形), 也关注该算法关于不可行稳定点的局部快速收敛性 (与定理 4.1(1) 的情形相对应). 首先需要下列一般性假设.

假设 4.1 假定下列假设条件成立:

- (1) 当 $k \rightarrow \infty$ 时有 $x_k \rightarrow x^*$;
- (2) 函数 f, h 和 g 在 x^* 的某个邻域内二次 Lipschitz 连续可微;
- (3) 梯度向量 $\nabla h_i(x^*)$ ($i \in \mathcal{I}_h$) 及 $\nabla g_i(x^*)$ ($i \in \mathcal{I}_g$) 线性无关, 其中 $\mathcal{I}_h = \{i : h_i(x^*) = 0, i = 1, \dots, p\}$, $\mathcal{I}_g = \{i : g_i(x^*) = 0, i = 1, \dots, m\}$.

定理 4.2 在假设 4.1 下, 假定 x^* 是一个不可行稳定点, 矩阵

$$B^* = \sum_{i=1}^p h_i(x^*) \nabla^2 h_i(x^*) - \sum_{i=1}^m \max\{0, -g_i(x^*)\} \nabla^2 g_i(x^*) + \sum_{i=1}^p \nabla h_i(x^*) \nabla h_i(x^*)^T \\ + \sum_{i \in \{i | g_i(x^*) < 0\}} \nabla g_i(x^*) \nabla g_i(x^*)^T$$

正定. 如果子问题 (4.2) 由 Newton 方法求解, 即在任何迭代点 x_{kj} , 新的迭代点

$$x_{k,j+1} = x_{kj} - \frac{1}{\rho_k} B_{kj}^{-1} \nabla_x L_B(x_{kj}, \lambda_k, \mu_k; \rho_k, \xi_k),$$

其中

$$B_{kj} = \frac{1}{\rho_k} \nabla_{xx}^2 L_B(x_{kj}, \lambda_k, \mu_k; \rho_k, \xi_k),$$

则

$$\|x_{k+1} - x^*\| = O\left(\frac{1}{\rho_k}\right) + O(\|x_k - x^*\|^2).$$

从而当 $\frac{1}{\rho_k} = O(\|x_k - x^*\|^2)$ 时, 序列 $\{x_k\}$ 的收敛速率是二次的; 否则, 若 $\frac{1}{\rho_k} = o(\|x_k - x^*\|)$, 则序列 $\{x_k\}$ 的收敛速率是超线性的.

定理 4.3 在假设 4.1 下, 假定 x^* 是一个 KKT 点, 对所有 k 充分大有 $\rho_k = \rho$, 当 $k \rightarrow \infty$ 时有 $(\lambda_k, \mu_k) \rightarrow (\lambda^*, \mu^*)$, 且 $\mu^* + g(x^*) > 0$, 矩阵

$$B^* = \nabla^2 f(x^*) - \sum_{i=1}^p \lambda_i^* \nabla^2 h_i(x^*) - \sum_{i=1}^m \mu_i^* \nabla^2 g_i(x^*) + \rho \sum_{i=1}^p \nabla h_i(x^*) \nabla h_i(x^*)^T \\ + \rho \sum_{i \in \{i | \mu_i^* > 0\}} \nabla g_i(x^*) \nabla g_i(x^*)^T$$

正定, 则存在正的常数 $\bar{\xi}, \epsilon$ 和 δ 使得对所有 $(\lambda, \mu; \xi) \in D = \{(\lambda, \mu; \xi) : \|(\lambda, \mu) - (\lambda^*, \mu^*)\| \leq \delta, \xi \in [0, \bar{\xi}]\}$, 问题

$$\min_{x \in \{x | \|x - x^*\| < \epsilon\}} L_B(x, \lambda, \mu; \rho, \xi)$$

有唯一解 $x(\lambda, \mu; \xi)$. 函数 $x(\cdot, \cdot; \cdot) : \mathbb{R}^{p+m} \times \mathbb{R}_+ \rightarrow \mathbb{R}^n$ 在 D 的内部连续可微, 且对所有 $(\lambda, \mu; \xi) \in D$, 有

$$\|x(\lambda, \mu; \xi) - x^*\| = O(\|(\lambda, \mu) - (\lambda^*, \mu^*)\|) + O(\xi).$$

而且, 如果 $\tilde{\mu}(\lambda, \mu; \xi) = \rho x(\lambda, \mu; \xi), \mu; \rho, \xi$, 则存在正的常数 $\bar{\xi}, \epsilon$ 和 δ , 使得对所有 $(\lambda, \mu; \xi) \in D = \{(\lambda, \mu; \xi) : \|(\lambda, \mu) - (\lambda^*, \mu^*)\| \leq \delta, \xi \in [0, \bar{\xi}]\}$, 有

$$\|\tilde{\mu}(\lambda, \mu; \xi) - \mu^*\| = O(\|(\lambda, \mu) - (\lambda^*, \mu^*)\|) + O(\xi).$$

定理 4.4 在定理 4.3 假设下, 如果 $\xi_k = O(\|x_k - x^*\|)$, 则

$$\max\{\|x_{k+1} - x^*\|, \|(\lambda_{k+1}, \mu_{k+1}) - (\lambda^*, \mu^*)\|\} = O(\|x_k - x^*\|) + O(\|(\lambda_k, \mu_k) - (\lambda^*, \mu^*)\|).$$

本节的理论表明算法 2 可快速探测问题 (1.1) 的不可行性, 它既能应用于可行优化, 也可应用于不可行优化, 相关证明可参见文献 [16].

5 内点增广 Lagrange 函数在求解凸优化中的应用

本节首先考虑算法 1 在求解线性半定规划和二阶锥规划中的应用, 然后基于内点增广 Lagrange 函数提出线性规划的交替极小化内点方法和加速交替极小化内点方法. 由于本节关于线性半定规划和二阶锥规划的算法的步骤与算法 1 完全相同, 限于篇幅, 我们不给出它们的算法的完整描述, 仅给出算法的关键部分, 即如何产生搜索方向及线搜索使用的效益函数.

5.1 求解线性半定规划的内点松弛方法

考虑线性半定规划问题:

$$\min_{X \in \mathcal{S}^n} C \bullet X \quad \text{s.t. } A_i \bullet X = b_i, \quad i = 1, \dots, p; \quad X \succeq 0, \quad (5.1)$$

其中 $\mathcal{S}^n$ 是 n 阶实对称矩阵空间, $\bullet$ 是矩阵的内积, C 和 A_i ($i = 1, \dots, p$) 是已知的 n 阶实对称矩阵, $b \in \mathbb{R}^p$ 是常数向量. 与问题 (1.1) 相对应, $f(X) = C \bullet X$, $h_i(X) = A_i \bullet X - b_i$, $g(X) = X$. 相应地,

$$L_B(X, \lambda, U; \rho, \xi) = C \bullet X - \sum_{i=1}^p \lambda_i (A_i \bullet X - b_i) + \frac{1}{2} \rho \sum_{i=1}^p (A_i \bullet X - b_i)^2 + \psi(X, U; \rho, \xi), \quad (5.2)$$

其中

$$\psi(X, U; \rho, \xi) = -\xi \ln \det S(X, U; \rho, \xi) - U \bullet (X - S(X, U; \rho, \xi)) + \frac{1}{2} \rho \|X - S(X, U; \rho, \xi)\|_F^2,$$

这里

$$S(X, U; \rho, \xi) = \frac{1}{2} \left[\sqrt{\left(X - \frac{U}{\rho}\right)^2 + 4 \frac{\xi}{\rho}} + \left(X - \frac{U}{\rho}\right) \right]$$

(这里矩阵的平方和开方是指, 按正交分解意义下的特征值的平方和开方运算得到的, 详见文献 [34]).

毫无疑问, 算法 1 可直接应用于求解问题 (5.1). 结合线性半定规划问题的结构, 我们仅需要关注非线性约束残差. 因此与算法 1 不同, 我们使用 $\phi_{(\xi, \rho)}(X, \lambda, U) = \frac{1}{2} \|S(X, U; \rho, \xi) - X\|_F^2$ 作为残差函数, 它也是算法线搜索中使用的效益函数. 为简单起见, 令 $S = S(X, U; \rho, \xi)$, $\rho Z = U - \rho X + \rho S$, $G = \rho(S + Z)$, 搜索方向由下列 Newton 方程组计算:

$$\begin{aligned} A_i \bullet \Delta X &= b_i - A_i \bullet X, \quad i = 1, \dots, p, \\ \sum_{i=1}^p \Delta \lambda_i A_i + \Delta U &= C - \sum_{i=1}^p \lambda_i A_i - U, \\ L_S(\Delta U) + L_{(\rho Z)}(\Delta X) &= 2\Delta \xi I + L_G(S - X), \end{aligned}$$

其中 $L_S(\Delta U) = S\Delta U + \Delta U S$, $L_{(\rho Z)}(\cdot)$ 和 $L_G(\cdot)$ 与它一样均为 Lyapunov 算子, $\Delta\xi = -\xi + \gamma\phi_{(\xi, \rho)}(X, \lambda, U)$. 通过矩阵向量化, 并引进半定矩阵优化中常用的符号 $\otimes_s$, $\mathcal{A}$ 和 svec , 令 $\Delta x = \text{svec}(\Delta X)$, $\Delta u = \text{svec}(\Delta U)$, 可将上述矩阵方程转化为下列线性方程组求解:

$$\begin{aligned}\mathcal{A}(S \otimes_s S) \mathcal{A}^T \Delta \lambda &= \xi r_p + \xi \mathcal{A} \mathcal{E}^{-1} (\mathcal{F} r_d - r_c), \\ \Delta u &= r_d - \mathcal{A}^T \Delta \lambda, \\ \Delta x &= \mathcal{E}^{-1} (r_c - \mathcal{F} \Delta u),\end{aligned}\tag{5.3}$$

其中 $\mathcal{E} = \rho Z \otimes_s I$, $\mathcal{F} = S \otimes_s I$, $r_p = b - \mathcal{A}x$, $r_d = \text{svec}(C - \mathcal{A}^T \lambda - U)$, $r_c = \text{svec}(\frac{1}{2} L_G(S - X) + \Delta \xi I)$ (详见文献 [34]).

与经典线性半定规划的内点方法如 AHO、HKM 和 NT 方法等不同, 我们所考虑的内点松弛算法 (IPRSDP) 不需要选择对称化算子. 表 1 给出了几个内点方法复杂度的对比. 从表中可见 IPRSDP 形成系数矩阵的复杂度与 NT 方法接近, 小于 AHO 方法及 HKM 方法.

5.2 求解二阶锥规划的内点松弛方法

二阶锥规划问题具有如下形式:

$$\min \sum_{i=1}^m c_i^T x_i \quad \text{s.t.} \quad \sum_{i=1}^m A_i x_i = b, \quad x_i \in K_i, \quad i = 1, \dots, m,\tag{5.4}$$

其中 $A_i \in \mathbb{R}^{p \times n_i}$, $c_i \in \mathbb{R}^{n_i}$, $b \in \mathbb{R}^p$, K_i 为维数为 n_i 的二阶锥, 即

$$K_i = \{x_i = (x_{i0}, \bar{x}_i) \in \mathbb{R} \times \mathbb{R}^{n_i-1} : \|\bar{x}_i\| \leq x_{i0}\},$$

这里 $\|\cdot\|$ 是 Euclid 范数. 令 $n = n_1 + n_2 + \dots + n_m$, $K = K_1 \times K_2 \times \dots \times K_m$, $A = (A_1 \ A_2 \ \dots \ A_m)$, $c = (c_1, c_2, \dots, c_m)$, $x = (x_1, x_2, \dots, x_n)$, 则 $K \subset \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, $c \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, 且问题 (5.4) 可以简写为

$$\min c^T x \quad \text{s.t.} \quad Ax = b, \quad x \in K.$$

类似地, 可得到与问题 (5.4) 相应的内点增广 Lagrange 函数

$$L_B(x, \lambda, \mu; \rho, \xi) = c^T x - \lambda^T (Ax - b) + \frac{1}{2} \rho \|Ax - b\|^2 + \sum_{i=1}^m \psi(x_i, \mu_i; \rho, \xi),\tag{5.5}$$

表 1 AHO, HKM, NT 和 IPRSDP 的线性方程组的系数矩阵以及形成系数矩阵的复杂度对比

方向	AHO	HKM	NT	IPRSDP
$\mathcal{E}$	$U \otimes_s I$	$I \otimes_s I$	$I \otimes_s I$	$\rho Z \otimes_s I$
$\mathcal{F}$	$X \otimes_s I$	$X \otimes_s U^{-1}$	$W \otimes_s W$	$S \otimes_s I$
$\mathcal{E}^{-1} \mathcal{F}$	$(U \otimes_s I)^{-1} (X \otimes_s I)$	$X \otimes_s U^{-1}$	$W \otimes_s W$	$\frac{1}{\mu} (S \otimes_s S)$
r_c	$\text{svec}(\xi I - \frac{XU+UX}{2})$	$\text{svec}(\xi U^{-1} - X)$	$\text{svec}(\xi U^{-1} - X)$	$\text{svec}(\frac{1}{2} L_G(S - X) + \Delta \xi I)$
复杂度	$6mn^3 + m^2n^2$	$4mn^3 + m^2n^2/2$	$2mn^3 + m^2n^2/2$	$2mn^3 + m^2n^2/2$

其中

$$\psi(x_i, \mu_i; \rho, \xi) = -\xi \ln \det s_i(x_i, \mu_i; \rho, \xi) - \mu_i^T (x_i - s_i(x_i, \mu_i; \rho, \xi)) + \frac{1}{2} \rho \|x_i - s_i(x_i, \mu_i; \rho, \xi)\|^2,$$

这里

$$s_i(x_i, \mu_i; \rho, \xi) = \frac{1}{2} \left[\left(\left(x_i - \frac{\mu_i}{\rho} \right)^2 + 8 \frac{\xi}{\rho} e_i \right)^{\frac{1}{2}} + \left(x_i - \frac{\mu_i}{\rho} \right) \right].$$

这里向量的 $(\cdot)^2$, $(\cdot)^{\frac{1}{2}}$ 和下面的 $(\cdot)^{-1}$ 是在 Jordan 代数意义下的运算^[1], 详细定义可参见文献 [35, (9)–(11)], 且 $\det(x) = x_0^2 - \|\bar{x}\|^2$ ($x = (x_0, \bar{x})$), $e_i = (1, 0) \in \mathbb{R}^{n_i}$. Zhang 等^[35] 已表明, 当 $\xi > 0$ 和 $\rho > 0$ 时, $s_i(x_i, \mu_i; \rho, \xi) \in K_i$.

同样地, 基于问题 (5.4) 的结构, 我们只需要关注函数

$$\phi_{(\xi, \rho)}(x, \lambda, \mu) = \frac{1}{2} \sum_{i=1}^m \|x_i - s_i(x_i, \mu_i; \rho, \xi)\|^2,$$

并由下列方程组计算搜索方向:

$$\begin{aligned} \Delta \xi &= -\xi + \gamma \phi_{(\xi, \rho)}(x, \lambda, \mu), \\ A \text{Arw}^{-1}(\rho z) \text{Arw}(s) A^T \Delta \lambda &= r_p + A \text{Arw}^{-1}(\rho z) (\text{Arw}(s) r_d - r_c), \\ \Delta \mu &= r_d - A^T \Delta \lambda, \\ \Delta x &= \text{Arw}^{-1}(\rho z) (r_c - \text{Arw}(s) \Delta \mu), \end{aligned} \quad (5.6)$$

其中 $\text{Arw}(\cdot)$ 是箭形矩阵, $s = s(x, \mu; \rho, \xi)$, $\rho z = \mu + \rho s - \rho x$, $r_p = b - Ax$, $r_d = c - A^T \lambda$, $r_c = \text{Arw}(\rho s + \rho z)(s - x) + \Delta \xi e$.

算法的主要计算量在于系数矩阵 $A \text{Arw}^{-1}(\rho z) \text{Arw}(s) A^T$ 的形成与分解. 令 $G = \text{Arw}^{-1}(\rho z) \text{Arw}(s)$. 即使 A 是稀疏的, 矩阵 AGA^T 仍然可能是一个稠密矩阵, 从而给问题的求解带来困难. 为了给出该系数矩阵高效的计算方式, 下面来分析矩阵 G 的结构. 首先考虑只有一个锥的情形, 即 $m = 1$.

定理 5.1 对于任意 $x, \mu \in \mathbb{R}^n$, 若 $s = s(x, \mu; \rho, \xi)$, $z = \frac{\mu}{\rho} + s - x$, 则

$$G \triangleq \text{Arw}^{-1}(\rho z) \text{Arw}(s) = \frac{1}{\xi} (\det(s) I + 2ss^T - 2\det(s)ee^T). \quad (5.7)$$

对于多个锥的情形, 由定理 5.1 有

$$AGA^T = \sum_{i=1}^m A_i \text{Arw}^{-1}(\rho z_i) \text{Arw}(s_i) A_i^T = \frac{1}{\xi} \sum_{i=1}^m A_i (\det(s_i) I_{n_i} + 2(s_i s_i^T - \det(s_i) e_i e_i^T)) A_i^T. \quad (5.8)$$

事实上, 为了计算 $\Delta \lambda$, 我们并不直接求解 (5.6), 而是通过定义 AGA^T 的一个所谓的稀疏扩展表示, 将它转化为求解一个线性方程组, 以避免形成稠密矩阵, 从而提高求解效率 (详见文献 [35]).

5.3 求解线性规划的交替极小化方法

考虑线性规划

$$\min_x c^T x \quad \text{s.t.} \quad Ax = b, \quad x \geq 0, \quad (5.9)$$

其中 $x \in \mathbb{R}^n$ 是决策变量, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ 及 $c \in \mathbb{R}^n$ 是常数矩阵和向量. 与问题 (5.9) 对应的另一个线性规划问题是

$$\min_y -b^T y \quad \text{s.t. } c - A^T y \geq 0. \quad (5.10)$$

众所周知, 问题 (5.9) 和 (5.10) 互为原始对偶问题, 有强对偶性成立. 且它们都是优化 (1.1) 在目标和约束函数是线性函数时的特例. 特别地, 在 $m < n$ 时, 与问题 (5.9) 相比, 问题 (5.10) 有更少的变量, 并且不包含等式约束. 若令 $s = c - A^T y$, 则 x 和 (y, s) 互为原始对偶变量, 即当 x^* 是问题 (5.9) 的最优解时, 相应的 Lagrange 乘子 (y^*, s^*) 中的 y^* 也是问题 (5.10) 的最优解, 且 $s^* = c - A^T y^*$, 反之亦然.

下面仅考虑求解问题 (5.10). 首先在问题 (5.10) 中引进松弛变量 $s \in \mathbb{R}_{++}^m$, 与它相应的对数障碍松弛问题为

$$\min_{y, s} -b^T y - \sum_{i=1}^m \xi \ln s_i \quad \text{s.t. } A^T y + s = c. \quad (5.11)$$

相应的内点增广 Lagrange 函数为

$$L_C(y, s, \nu; \rho, \xi) \triangleq -b^T y - \sum_{i=1}^m \xi \ln s_i + \nu^T (A^T y + s - c) + \frac{1}{2} \rho \|A^T y + s - c\|^2. \quad (5.12)$$

与问题 (4.1) 使用 $L_B(x, \lambda, \mu; \rho, \xi)$ 为目标函数不同, 这里使用 $L_C(y, s, \nu; \rho, \xi)$ 为目标函数, 并考虑下列极大极小问题的交替极小化方法 (见算法 3):

$$\max_{\nu \in \mathbb{R}^n} \min_{y \in \mathbb{R}^m, s \in \mathbb{R}^n} L_C(y, s, \nu; \rho, \xi). \quad (5.13)$$

算法 4 是结合文献 [9] 中的加速技巧提出的一个具有加速功能的内点交替极小化方法. 如果对所有 $k \geq 0$, 有 $\tau_k = 1$, 则它退化为算法 3.

算法 3 求解线性规划 (5.10) 的内点交替极小化方法

给定 $y_0 \in \mathbb{R}^m$, $s_0 \in \mathbb{R}_{++}^n$, $\nu_0 \in \mathbb{R}_{++}^n$, $\xi_0 > 0$, $\rho_0 > 0$, $\rho_{\max} > \rho_0$, $\gamma_1, \gamma_2 \in (0, 1)$, $\epsilon > 0$.

对于 $k = 0, 1, \dots$, 如果终止条件满足, 算法终止.

步 1 (极小化 y -子问题). 计算 $y_{k+1} = \operatorname{argmin}_y L_C(y, s_k, \nu_k; \rho_k, \xi_k)$.

步 2 (极小化 s -子问题). 计算 $s_{k+1} = \operatorname{argmin}_s L_C(y_{k+1}, s, \nu_k; \rho_k, \xi_k)$.

步 3 (修正变量 ν). 令 $\nu_{k+1} = \nu_k + \rho_k (A^T y_{k+1} + s_{k+1} - c)$.

步 4 (修正内点参数和罚参数). 若 $\|A^T y_{k+1} + s_{k+1} - c\|_\infty \geq \|A^T y_k + s_k - c\|_\infty \geq \epsilon$, 增加 ρ_k 为 $\rho_{k+1} = \min(2\rho_k, \rho_{\max})$; 若 $\|A\nu_{k+1} - b\| \leq \gamma_1 \|A\nu_k - b\|$ 和 $\xi_k \geq \epsilon$, 令 $\xi_{k+1} \leq \gamma_2 \xi_k$.

算法 4 求解线性规划的加速内点交替极小化方法

给定 $y_0 \in \mathbb{R}^m$, $s_0 \in \mathbb{R}_{++}^n$, $\nu_0 \in \mathbb{R}_{++}^n$, $\xi_0 > 0$, $\rho_0 > 0$, $\rho_{\max} > \rho_0$, $\gamma_1, \gamma_2 \in (0, 1)$, $\tau_k \in (0, 2)$, $\epsilon > 0$.

对于 $k = 0, 1, \dots$, 如果终止条件满足, 终止算法.

步 1. 计算 $y_{k+1} = \operatorname{argmin}_y -b^T y + \nu_k^T A^T y + \frac{1}{2} \rho_k \|A^T y + s_k - c\|^2$.

步 2. 计算 $s_{k+1} = \operatorname{argmin}_s -\xi_k \sum_{i=1}^n \ln s_i + \nu_k^T s + \frac{1}{2} \rho_k \|\tau_k (A^T y_{k+1} - c) - (1 - \tau_k) s_k + s\|^2$.

步 3. 令 $\nu_{k+1} = \nu_k + \rho_k (\tau_k (A^T y_{k+1} - c) - (1 - \tau_k) s_k + s_{k+1})$.

步 4. 若 $\|A^T y_{k+1} + s_{k+1} - c\|_\infty \geq \|A^T y_k + s_k - c\|_\infty \geq \epsilon$, 增加 ρ_k 为 $\rho_{k+1} = \min(2\rho_k, \rho_{\max})$; 若 $\|A\nu_{k+1} - b\| \leq \gamma_1 \|A\nu_k - b\|$ 和 $\xi_k \geq \epsilon$, 令 $\xi_{k+1} \leq \gamma_2 \xi_k$.

注 5.1 算法 3 和 4 的主要计算量在步 1, 需要求解一个无约束的凸二次优化问题, 且该问题的 Hessian 矩阵为常数矩阵 AA^T . 它等价于求解一个正定或半正定的线性方程组, 因此算法特别适合于求解大规模线性规划. 步 2 是按分量可分离的严格凸二次函数极小化, 因此具有解析解, 且在 ρ_k 和 ξ_k 均为正的情形下, 所有解析解是正的, 因此两个方法是内点交替极小化方法. 这也是至今已知的第三个内点交替方向方法 (前两个内点交替方向方法参见文献 [13, 15]).

6 结论与展望

由于本文中的所有算法对于原始不等式约束的松弛变量对所有 $k \geq 0$ 均有 $s_k > 0$ (即所有迭代点均为内点), 因此它们均可归结为内点方法. 另外, 它们都允许初始对偶变量 μ_0 (或 ν_0) 是任意的, 其中前 4 个关于非线性优化的算法中的对偶迭代 μ_k 是否是内点取决于不等式约束是否渐进满足, 这是它们被称为内点松弛方法的原因. 注意到算法 3 和 4 中的对偶迭代对所有 $k \geq 0$ 有 $\nu_k > 0$, 因此它们本质上是一类原始对偶内点方法. 这也是它们与现有绝大多数线性规划交替极小化方法的重要不同之处.

基于对数障碍问题的增广 Lagrange 函数, 本文首先提出了求解一般约束优化的内点方法 (算法 1). 与现有绝大多数非线性优化内点方法不同, 算法 1 不使用内外迭代框架 (即在内迭代近似求解一个对数障碍子问题, 同时在外迭代修正障碍参数), 且在迭代过程中不需要通过阻尼步长来保证迭代点是内点. 从而实现了内点方法与光滑化 Newton 方法, 以及内点障碍参数和光滑化参数之间的贯通. 该方法在特殊凸优化中的应用表明了算法的有效性. 然后从对偶的角度, 考虑了一般约束优化的增广 Lagrange 内点方法 (算法 2). 这些内点方法与现有内点方法和增广 Lagrange 方法完全不同, 且能够快速探测问题的不可行性. 特别是, 对于线性规划, 我们的交替极小化方法是一个原始对偶内点方法, 且每次只需要求解一个常数系数矩阵的线性方程组, 不会出现现有绝大多数原始对偶内点方法类似的坏条件现象.

对于内点方法, 一个自然产生的疑问是, 对于凸优化和线性规划, 本文的内点方法是否具有多项式迭代复杂性. 另有两个值得关注的问题: 一是关于线性规划的交替极小化方法与现有凸优化的交替极小化方法享有极其类似的算法框架, 因此可期待我们的方法与现有交替极小化方法一样, 具有非遍历意义下和点列意义下的收敛速率; 二是关于凸优化的交替极小化方法 (包括 ADMM 方法) 的广泛的研究结果和有效技术 (如预估校正技术和惯性邻近技术) 是否可应用来进一步改进我们的方法. 这些问题是我们下一步的研究方向.

致谢 感谢审稿人对本文的认真审阅.

参考文献

- 1 Alizadeh F, Goldfarb D. Second-order cone programming. *Math Program*, 2003, 95: 3–51
- 2 Burke J V, Curtis F E, Wang H. A sequential quadratic optimization algorithm with rapid infeasibility detection. *SIAM J Optim*, 2014, 24: 839–872
- 3 Burke J V, Han S P. A robust sequential quadratic programming method. *Math Program*, 1989, 43: 277–303
- 4 Byrd R H, Curtis F E, Nocedal J. Infeasibility detection and SQP methods for nonlinear optimization. *SIAM J Optim*, 2010, 20: 2281–2299
- 5 Byrd R H, Gilbert J C, Nocedal J. A trust region method based on interior point techniques for nonlinear programming. *Math Program*, 2000, 89: 149–185
- 6 Byrd R H, Hribar M E, Nocedal J. An interior point algorithm for large-scale nonlinear programming. *SIAM J Optim*, 1999, 9: 877–900

- 7 Chen L, Goldfarb D. Interior-point ℓ_2 -penalty methods for nonlinear programming with strong global convergence properties. *Math Program*, 2006, 108: 1–36
- 8 Dai Y H, Liu X W, Sun J. A primal-dual interior-point method capable of rapidly detecting infeasibility for nonlinear programs. *J Ind Manag Optim*, 2020, 16: 1009–1035
- 9 Glowinski R. On alternating direction methods of multipliers: A historical perspective. In: *Modeling, Simulation and Optimization for Science and Technology*, vol. 34. Dordrecht: Springer, 2014, 59–82
- 10 Gondzio J. Interior point methods 25 years later. *European J Oper Res*, 2012, 218: 587–601
- 11 Hestenes M R. Multiplier and gradient methods. *J Optim Theory Appl*, 1969, 4: 303–320
- 12 Karmarkar N. A new polynomial-time algorithm for linear programming. *Combinatorica*, 1984, 4: 373–395
- 13 Lin T, Ma S, Ye Y, et al. An ADMM-based interior-point method for large-scale linear programming. *Optim Methods Softw*, 2021, 36: 389–424
- 14 Liu X W, Dai Y H, Huang Y K. A primal-dual interior-point relaxation method with global and rapidly local convergence for nonlinear programs. *Math Methods Oper Res*, 2022, 96: 351–382
- 15 Liu X W, Dai Y H, Huang Y K. Convergence and iteration-complexity of a primal-dual majorization-minimization method for large-scale linear programming. *Pac J Optim*, 2024, 20: 513–535
- 16 Liu X W, Dai Y H, Huang Y K, et al. A novel augmented Lagrangian method of multipliers for optimization with general inequality constraints. *Math Comp*, 2023, 92: 1301–1330
- 17 Liu X W, Sun J. A robust primal-dual interior-point algorithm for nonlinear programs. *SIAM J Optim*, 2004, 14: 1163–1186
- 18 Liu X W, Yuan Y X. A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties. *Math Program*, 2010, 125: 163–193
- 19 Nesterov Y. *Lectures on Convex Optimization*, 2nd ed. Berlin: Springer, 2018
- 20 Nocedal J, Wright S. *Numerical Optimization*. New York: Springer-Verlag, 1999
- 21 Ortega J M, Rheinboldt W C. *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic Press, 1970
- 22 Powell M J D. A method for nonlinear constraints in minimization problems. In: *Optimization*. New York: Academic Press, 1969, 283–298
- 23 Qi L, Sun J. A nonsmooth version of Newton’s method. *Math Program*, 1993, 58: 353–367
- 24 Rockafellar R T. The multiplier method of Hestenes and Powell applied to convex programming. *J Optim Theory Appl*, 1973, 12: 555–562
- 25 Rockafellar R T. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math Program*, 1973, 5: 354–373
- 26 Rockafellar R T. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math Oper Res*, 1976, 1: 97–116
- 27 Sun W Y, Yuan Y X. *Optimization Theory and Methods: Nonlinear Programming*. New York: Springer, 2006
- 28 Vanderbei R J, Shanno D F. An interior-point algorithm for nonconvex nonlinear programming. *Comput Optim Appl*, 1999, 13: 231–252
- 29 Wächter A, Biegler L T. Failure of global convergence for a class of interior point methods for nonlinear programming. *Math Program*, 2000, 88: 565–574
- 30 Wächter A, Biegler L T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*, 2006, 106: 25–57
- 31 Ye Y. *Interior Point Algorithms: Theory and Analysis*. New York: John Wiley & Sons, 1997
- 32 Yuan Y X. On the convergence of a new trust region algorithm. *Numer Math*, 1995, 70: 515–539
- 33 Zhang R J, Liu X W, Dai Y H. IPRQP: A primal-dual interior-point relaxation algorithm for convex quadratic programming. *J Global Optim*, 2023, 87: 1027–1053
- 34 Zhang R J, Liu X W, Dai Y H. IPRSDP: A primal-dual interior-point relaxation algorithm for semidefinite programming. *Comput Optim Appl*, 2024, 88: 1–36
- 35 Zhang R J, Wang Z W, Liu X W, et al. IPRSOCP: A primal-dual interior-point relaxation algorithm for second-order cone programming. *J Oper Res Soc China*, 2025, in press, doi: 10.1007/s40305-024-00538-z

Interior-point augmented Lagrangian methods for constrained optimization

Xinwei Liu & Yu-Hong Dai

Abstract The interior-point approach is one kind of the most important and effective methods for nonlinear inequality-constrained optimization, whereas classic augmented Lagrangian methods are very efficient in solving nonlinear equality-constrained optimization and convex optimization. The new research has shown that the interior-point augmented Lagrangian function can be of very nice higher-order smoothness, and is convex when the original optimization problem is convex. These nice properties make the novel augmented Lagrangian applicable and effective not only for convex optimization but also for non-convex optimization problems. In this paper, we provide an overview of our recent research progress about the interior-point augmented Lagrange function methods on the non-convex and convex optimization, and point out prospects for future research directions.

Keywords constrained optimization, non-convex optimization, convex optimization, linear programming, augmented Lagrangian, interior-point method, alternating minimization method

MSC(2020) 90C25, 90C26, 90C30, 90C51

doi: 10.1360/SSM-2024-0132