

求解双层规划问题的光滑化牛顿方法

李晓夏

河北工业大学理学院, 天津

收稿日期: 2025年8月11日; 录用日期: 2025年9月5日; 发布日期: 2025年9月12日

摘要

考虑一类具有特殊结构的双层规划问题, 其上层问题无约束, 而下层问题为凸优化问题。首先, 通过引入下层问题的值函数, 将原问题转化为单层优化问题(VP)。在部分平稳性条件下, 通过将值函数约束惩罚到目标函数, 进一步得到问题(VP)的部分精确罚问题。随后, 在一定的约束规范条件下, 推导出该罚问题的最优性条件, 并利用光滑障碍增广拉格朗日函数方法处理其中的互补约束条件, 从而将问题转化为超定方程组进行求解。针对该方程组, 采用经典的高斯牛顿法进行求解, 并理论证明了当障碍参数趋近于零时算法的收敛性。最后, 基于BOLIB库中的小规模问题开展数值实验, 将所提算法与LM算法、拟牛顿法等方法的计算结果进行对比分析, 实验结果表明了该算法的有效性。

关键词

双层规划, 值函数, 约束规范, 光滑障碍增广拉格朗日函数

A Smoothing Newton Method for Bilevel Programming Problems

Xiaoxia Li

School of Science, Hebei University of Technology, Tianjin

Received: Aug. 11th, 2025; accepted: Sep. 5th, 2025; published: Sep. 12th, 2025

Abstract

Consider a class of bilevel programming problems with a special structure, where the upper-level problem is unconstrained and the lower-level problem is a convex optimization problem. First, by introducing the value function of the lower-level problem, the original problem is reformulated into a single-level optimization problem (VP). Under the partial calmness condition, an exact penalty problem for (VP) is further derived by penalizing the value function constraint into the objective function. Subsequently, under certain constraint qualifications, optimality conditions for the penalized problem are established. The complementary constraints in the optimality conditions are handled using a smoothing barrier augmented Lagrangian function method, transforming the problem into an

overdetermined system of equations. To solve this system, the classical Gauss-Newton method is employed, and the convergence of the algorithm is theoretically proven as the barrier parameter tends to zero. Finally, numerical experiments are conducted on small-scale problems from the BOLIB library. The proposed algorithm is compared with the LM algorithm, the Quasi-Newton method, and other approaches. The experimental results demonstrate the effectiveness of the proposed algorithm.

Keywords

Bilevel Programming, Value Function, Constraint Qualification, Smoothing Barrier Augmented Langrange Function

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

双层规划(Bilevel Programming)是指一类具有两级层次结构的系优化问题,其特点是上层问题和下层问题均具有独立的决策变量,且两层的约束条件与目标函数相互耦合。两层决策者在不同层次上进行决策。上层决策者制定某目标函数并对下层决策者做出一些限制条件,但不直接干涉下层的决策;而下层决策者则在上层给定的约束条件下,优化自身目标函数以确定最优策略。值得注意的是,下层决策会进一步影响上层目标函数的值,从而形成双向交互的双层决策机制。目前,双层规划在管理科学,工程优化及经济决策等领域具有广泛的应用(见[1]-[4])。同时,其理论研究促进了博弈论,最优化以及计算机等研究领域的发展,成为了解决复杂现实问题的重要工具。因此,开发高效的双层规划求解算法具有重要的理论与应用价值。

我们考虑以下双层规划模型:

$$(BP) \min F(x, y) \text{ s.t. } y \in S(x), \tag{1}$$

其中 $S(x)$ 是下层问题的最优解集

$$(P_x) \min_{y \in Y(x)} f(x, y), \tag{2}$$

即

$$S(x) := \arg \min_y \{f(x, y) \mid g(x, y) \leq 0\}. \tag{3}$$

这里我们假设 $Y(x) = \{y \in \mathbb{R}^m : g(x, y) \leq 0\}$, 函数 $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ 是连续可微的, $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^p$ 是连续可微函数,且对变量 y 是二阶连续可微的。定义 \mathcal{S} 是问题(BP)的可行域。我们的结果很容易推广到标准的双层规划模型,因此为方便分析结果,我们忽略了上层问题的约束条件。

众所周知,由于双层规划问题复杂的结构使得求解问题(1)较为困难。即使对于最简单的双层规划问题,也已被证明属于 NP-hard 问题(参见文献[5]),这主要是因为双层规划模型的解通常是近似最优解而非全局最优解。目前有多种方法可以求解双层规划问题(1),其中最常用的方法就是将其转化为一个单层问题。

一种方法是基于下层问题 KKT 条件,将双层规划问题转化为带有互补约束的优化问题:

$$\begin{aligned} (MPEC) \min_{x, y, \lambda} F(x, y) \\ \text{s.t. } \nabla_y f(x, y) + \lambda^T \nabla_y g(x, y) = 0, \\ \lambda \geq 0, g(x, y) \leq 0, \\ \lambda^T g(x, y) = 0. \end{aligned}$$

该方法为最常用的转化方法之一，上述问题被称为互补约束数学规划(Mathematical Program with Equilibrium Constraint, 简称 MPEC)。需要指出的是，即便在下层问题为凸的情形下，双层规划问题与对应的 MPEC 模型也未必等价[6]。为确保二者的等价性，需要添加一些额外的约束条件。Dempe 等人[7][8]研究了当下层问题满足 Slater 约束规范时，双层规划的全局解(局部解)与 MPEC 模型全局解(局部解)之间的关系。另一方面，转化之后的问题是非凸的，且理论上 MPEC 在任何可行点均不满足 MFCQ [9]。这会导致 MPEC 可能存在局部最优解不满足 KKT 条件的情况，使得一般的非线性规划理论不能直接应用于 MPEC 问题。为此，学者们引入了一些较弱的稳定点和约束规范(见[10][11])。此外，在数值求解方面，众多学者提出近似算法求解 MPEC 模型。[12]中设计了求解 MPEC 问题的序列二次规划方法，并建立了该方法的强收敛性理论。[13]中提出了求解一般约束优化问题的对偶内点方法，在不依赖正则性条件的假设下，证明了强收敛理论，并结合不等式松弛技术，用该方法求解 MPEC 模型。[14]中考虑一类上下层问题具有相同变量的问题，研究 MPEC 模型的最优性理论，通过构造对偶间隙函数将 MPEC 模型转化为非光滑问题进行研究。关于 MPEC 理论和算法的最新发展，请读者参阅文献[12]-[14]及其引用的相关研究工作。

另一种常用的方法是利用下层问题的值函数将原问题转化为单层问题。对于任意的 x ，假设下层问题的最优解集 $S(x) \neq \emptyset$ ，定义下层问题的值函数为

$$V(x) := \inf_{y \in Y(x)} f(x, y). \quad (4)$$

从值函数的定义易得 $y \in S(x)$ 当且仅当对于 $y \in Y(x)$ ， $f(x, y) - V(x) = 0$ ，令 $\varepsilon > 0$ ，我们考虑以下单层问题

$$\begin{aligned} (\text{VP}_\varepsilon) \quad & \min_{x, y} F(x, y) \\ & \text{s.t. } g(x, y) \leq 0, \\ & f(x, y) \leq V(x) + \varepsilon, \end{aligned} \quad (5)$$

若 $\varepsilon = 0$ ，问题 (VP_ε) 称为问题(VP)，此时为问题(BP)的一种重构；若 $\varepsilon > 0$ ，问题 (VP_ε) 则构成问题(BP)的近似模型。这一转化问题最初是由 Outrata [15]提出用于数值计算，随着被 Ye 和 Zhu [16]将其应用于推导双层规划问题的最优必要性条件。通常情况下，值函数是不可微的，且值函数约束的存在使得该问题很难满足一般的约束规范。针对这一挑战，文献[17]提出了积分熵函数来近似值函数，证明了积分熵函数是值函数的光滑逼近函数，并且在平稳性条件下，设计了一种光滑投影梯度法来求解该问题。除了用光滑函数近似值函数外，将原问题转化为方程组形式也是一种重要的求解方法。文献[18]-[20]中在部分平稳性条件和一定的约束规范下，推导出问题(VP)部分精确罚问题的最优性条件，并用光滑 FB 函数处理其中的互补约束条件，此时原问题就可转化为一个方程组。针对该系统，学者们运用了高斯牛顿法，半光滑牛顿法，LM 算法进行求解，从而近似求解了原问题。

本文中同样基于问题(VP)，在部分平稳性条件和一定的约束规范下，系统推导了该优化问题的最优性条件。针对其中的互补约束条件，我们采用光滑障碍增广拉格朗日函数方法进行处理，从而将其转化为超定方程组。在数值求解方面，文中提出了一种结合线搜索策略的高斯牛顿算法，并证明了当障碍参数趋近于零时算法的收敛性。

第二章系统介绍了相关约束规范的定义以及光滑障碍增广拉格朗日函数方法。第三章建立了在一定条件下将原问题转化为超定方程组的框架以及设计了高斯牛顿法和自适应线搜索法结合的算法，并证明了其收敛性。第四章中基于 BOLIB 库，进行了小规模数值实验，并与一些现有典型算法进行了对比分析。

2. 预备知识

2.1. 变分分析

首先介绍本文用到的变分分析[21]中的一些基础知识。给定集合 $C \subseteq \mathbb{R}^d$ ，对于 $\forall z \in C$ ，它的正则法锥定义为

$$\hat{N}_C(z) := \{v \in \mathbb{R}^d : \langle v, z' - z \rangle \leq o(\|z - z'\|) \quad \forall z' \in C\},$$

在该点的极限法锥定义为

$$N_C(z) := \left\{ \lim_{k \rightarrow \infty} v_k : v_k \in \hat{N}_C(z_k), z_k \in C, z_k \rightarrow z \right\}.$$

若集合 $C \subseteq \mathbb{R}^d$ 在 z 处局部闭且满足 $N_C(z) = \hat{N}_C(z)$ ，则称该集合在 z 处是 Clarke 正则的。若 $f: \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ 是一个下半连续的函数且在 $x \in \mathbb{R}^d$ 处是有限的，可定义该函数在点 x 的正则次微分为

$$\hat{\partial}f(x) := \left\{ \zeta \in \mathbb{R}^d : f(x') - f(x) - \langle \zeta, x - x' \rangle \leq o(\|x - x'\|) \right\},$$

在该点的极限次微分为

$$\partial f(x) := \left\{ \lim_{k \rightarrow \infty} \xi_k : \xi_k \in \hat{\partial}f(x_k), x_k \rightarrow x, f(x_k) \rightarrow f(x) \right\},$$

并且函数 f 在点 x 的 Clarke 次微分为

$$\partial^c f(x) := \text{co} \partial f(x).$$

令 $f: \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ 在点 x 是 Lipschitz 连续的，若 $\hat{\partial}f(x) = \partial f(x)$ ，我们称函数 f 在该点是 Clarke 正则的。

2.2. 约束规范

若优化问题的目标函数和约束条件均为凸函数，则称该问题为完全凸的。我们考虑以下系统

$$\begin{aligned} \text{(P)} \quad & g_i(x) \leq 0, \quad i = 1, \dots, p, \\ & h_j(x) = 0, \quad j = p+1, \dots, q, \end{aligned}$$

其中函数 $g_i (i = 1, \dots, p)$ 和 $h_j (j = p+1, \dots, q)$ 是连续可微的。

对于问题(P)的可行点 \bar{x} ，定义 $I(\bar{x}) := \{i = 1, \dots, p, g_i(\bar{x}) = 0\}$ 为 \bar{x} 处的积极集。下面介绍一些常用的约束规范[21]。

定义 2.1. (MFCQ) 在问题(P)的可行点 \bar{x} 处满足 MFCQ，如果存在向量 $d \in \mathbb{R}^n$ 使得

$$\begin{aligned} \nabla g_i(\bar{x})^\top d &< 0, \quad \forall i \in I(\bar{x}), \\ \nabla h_j(\bar{x})^\top d &= 0, \quad j \in \{p+1, \dots, q\} \end{aligned}$$

成立，并且等式约束梯度集 $\{\nabla h_j(\bar{x}), j \in \{p+1, \dots, q\}\}$ 是线性无关的。

2.3. 光滑障碍增广拉格朗日函数方法

考虑以下只有不等式约束的非线性优化问题

$$\begin{aligned} \min_{x,y} \quad & f(x, y) \\ \text{s.t.} \quad & g(x, y) \leq 0. \end{aligned}$$

若 $g(x, y) < 0$ 严格成立，用文献[22]中提出的对数障碍法重构该问题

$$\begin{aligned} \min_{x,y} f(x,y) - r \sum_{i=1}^p \ln z_i \\ \text{s.t. } z + g(x,y) = 0. \end{aligned}$$

对于 $\rho > 0$, $s \in \mathbb{R}^p$, 我们考虑上述问题的增广拉格朗日函数

$$L_r^\rho(x,y,z,s) := f(x,y) + \sum_{i=1}^p \left[-r \ln z_i + s_i (z_i + g_i(x,y)) + \frac{1}{2\rho} (z_i + g_i(x,y))^2 \right].$$

为确保乘子的可行性, 我们需要关于 s 对上述函数取最大值, 进而有以下问题:

$$\min_{x,y,z} \max_s L_r^\rho(x,y,z,s),$$

对于任意解 (x,y,s) , 一定有 $\nabla_z L_r^\rho(x,y,z,s) = 0$, 从而推导出 z 是关于 x,y,s,r,ρ 的函数:

$$z_i(x,y,s_i,r,\rho) := \frac{1}{2} \left(\sqrt{(\rho s_i + g_i(x,y))^2 + 4r\rho} - (\rho s_i + g_i(x,y)) \right), \quad (6)$$

同时定义函数

$$\begin{aligned} \kappa_i(x,y,s_i,r,\rho) &:= z_i(x,y,s_i,r,\rho) + g_i(x,y) + \rho s_i \\ &= \frac{1}{2} \left(\sqrt{(\rho s_i + g_i(x,y))^2 + 4r\rho} + (\rho s_i + g_i(x,y)) \right), \end{aligned} \quad (7)$$

其中 $i=1,\dots,p$. 为方便叙述, 我们将 $z_i(x,y,s_i,r,\rho)$ 和 $\kappa_i(x,y,s_i,r,\rho)$ 简写为 z_i 和 κ_i .

类似于刘等人[23], 我们列出函数 z_i 和 κ_i 的相关性质.

引理 2.1. 对于 $r \geq 0$, $\rho > 0$, $i=1,\dots,p$, 下列结论成立.

- a) $z_i \geq 0$, $\kappa_i \geq 0$, $z_i + g_i(x,y) = \kappa_i - \rho s_i$ 且有 $z_i \kappa_i = r\rho$;
- b) $z_i + g_i(x,y) = 0$ 当且仅当 $g_i(x,y) \leq 0$, $s_i \geq 0$, $-s_i g_i(x,y) = r$;
- c) 对于 $r > 0$, z_i 和 κ_i 关于变量 x,y,s_i 是可微的, 且有

$$\begin{aligned} \nabla_{(x,y)} z_i &= \frac{-z_i}{z_i + \kappa_i} \nabla g_i(x,y), \quad \nabla_{(x,y)} \kappa_i = \frac{\kappa_i}{z_i + \kappa_i} \nabla g_i(x,y), \\ \nabla_s z_i &= \frac{-\rho z_i}{z_i + \kappa_i} e_i, \quad \nabla_s \kappa_i = \frac{\rho \kappa_i}{z_i + \kappa_i} e_i. \end{aligned}$$

其中 e_i 是第 i 个元素为 1, 其余元素为 0 的单位向量.

3. 最优性条件

为求得问题(VP)的最优性条件, 我们还需引入部分平稳性条件.

定义 3.1. (部分平稳性条件) 设 (\bar{x}, \bar{y}) 是(VP)的局部最优解, 该问题在 (\bar{x}, \bar{y}) 满足部分平稳条件, 如果存在 $\lambda > 0$ 及 $(\bar{x}, \bar{y}, 0)$ 的邻域 \mathcal{U} 使得

$$F(x,y) - F(\bar{x}, \bar{y}) + \lambda |u| \geq 0, \quad \forall (x,y,u) \in \mathcal{U}: g(x,y) \leq 0, f(x,y) - V(x) - u = 0.$$

根据文献[24], 如果问题(VP)在局部最优解 (\bar{x}, \bar{y}) 满足部分平稳条件, 那么存在参数 $\lambda > 0$ 使得也是以下问题的局部最优解:

$$\min_{x,y} F(x,y) + \lambda (f(x,y) - V(x)) \quad \text{s.t. } g(x,y) \leq 0. \quad (8)$$

显然, 上述问题是将值函数约束作为目标函数的惩罚项. 因此, 我们通常称它为问题(VP)的部分精

确罚问题。

基于上述问题，可应用约束规范推导最优性条件，因此，我们有以下结论。

定理 3.1. 令 (x, y) 是问题(VP)的局部最优解。假设该问题中的所有函数均为二阶连续可微的，值函数 $V(x)$ 在 x 附近有限且下层问题完全凸。此外，问题(VP)在 (x, y) 处满足部分平稳性条件，下层问题在 (x, y) 处满足 MFCQ。那么，存在 $\lambda \in (0, +\infty)$ 和拉格朗日乘子 s, w 使得

$$\nabla_x F(x, y) + (w - \lambda s) \nabla_x g(x, y) = 0, \tag{9}$$

$$\nabla_y F(x, y) + \lambda \nabla_y f(x, y) + w \nabla_y g(x, y) = 0, \tag{10}$$

$$\nabla_y f(x, y) + s \nabla_y g(x, y) = 0, \tag{11}$$

$$s_i \geq 0, g_i(x, y) \leq 0, s_i g_i(x, y) = 0, \tag{12}$$

$$w_i \geq 0, g_i(x, y) \leq 0, w_i g_i(x, y) = 0, \tag{13}$$

其中 $i=1, \dots, p$ 。

在上述的最优性条件中，互补约束条件(9)~(13)的存在会使得常见的约束规范不成立从而导致 KKT 条件无法直接应用。同时，考虑到该系统的可微性，基于引理 2.1 (2)，我们用光滑函数(6)处理互补约束条件。对于 $\rho > 0$ 和充分小的 $r > 0$ ，我们构造如下光滑系统

$$\psi_{(r, \rho, \lambda)}(x, y, s, w) = \begin{pmatrix} \nabla_x F(x, y) + (w - \lambda s) \nabla_x g(x, y) \\ \nabla_y F(x, y) + \lambda \nabla_y f(x, y) + w \nabla_y g(x, y) \\ \nabla_y f(x, y) + s \nabla_y g(x, y) \\ z(x, y, s, r, \rho) + g(x, y) \\ z(x, y, w, r, \rho) + g(x, y) \end{pmatrix} = 0. \tag{14}$$

当 $r=0$ ，该系统与原系统(9)~(13)等价。根据文献[25]中提出的光滑化方法，对于固定参数 λ ，我们需要构造一个递减到 0 的序列 $\{r_k\}$ 来近似求解系统(9)~(13)，从而使得当 $k \rightarrow \infty$ ，

$$\psi_{(r_k, \rho, \lambda)}(x_k, y_k, s_k, w_k) \rightarrow 0,$$

显然，式(14)包含 $n+m+2p$ 个变量，而方程组个数为 $n+2m+2p$ ，这表明该系统是一个超定方程组。接下来，我们采用经典的高斯牛顿法求解该方程组：

算法 1

步骤 0: 给定 (x_0, y_0, s_0, w_0) ， $\lambda > 0$ ， $\rho_0 > 0$ ， $\varepsilon > 0$ ， $K > 0$ ， $r, \nu, \omega \in (0, 1)$ ，令 $k := 0$ 。

步骤 1: 若 $\|\psi_{(r, \rho_k, \lambda)}(x_k, y_k, s_k, w_k)\| \leq \varepsilon$ 或 $k \geq K$ ，算法结束。

步骤 2: 计算雅可比矩阵 $\nabla \psi_{(r, \rho_k, \lambda)}(x_k, y_k, s_k, w_k)$ 和迭代方向 d_k

$$d_k = -\left(\nabla \psi_{(r, \rho_k, \lambda)}^T \nabla \psi_{(r, \rho_k, \lambda)}\right)^{-1} \nabla \psi_{(r, \rho_k, \lambda)}^T \psi_{(r, \rho_k, \lambda)}. \tag{15}$$

步骤 3: 在 $\{1, \nu, \nu^2, \dots\}$ 中选择最大值作为迭代步长 $\alpha_k \in (0, 1]$ ，并且该步长需要满足不等式

$$\begin{aligned} & \left\| \psi_{(r, \rho_k, \lambda)}(x_k + \alpha_k d_{xk}, y_k + \alpha_k d_{yk}, s_k + \alpha_k d_{sk}, w_k + \alpha_k d_{wk}) \right\|^2 \\ & \leq \left\| \psi_{(r, \rho_k, \lambda)}(x_k, y_k, s_k, w_k) \right\|^2 + \omega \alpha_k \nabla \psi_{(r, \rho_k, \lambda)}^T(x_k, y_k, s_k, w_k) \psi_{(r, \rho_k, \lambda)}(x_k, y_k, s_k, w_k) d_k. \end{aligned} \tag{16}$$

步骤 4: 更新 $x_{k+1} = x_k + \alpha_k d_{xk}$ ， $y_{k+1} = y_k + \alpha_k d_{yk}$ ， $s_{k+1} = s_k + \alpha_k d_{sk}$ ， $w_{k+1} = w_k + \alpha_k d_{wk}$ ， $\rho_k \geq \rho_{k+1}$ 。

步骤 5: 令 $k := k + 1$ ，返回步骤 1。

在算法 1 中, ε 表示容差, K 是最大迭代次数。对于步长 α_k 的选取, 我们采用 Armijo 方法。由式(15)可知, 要使该算法良定, 矩阵 $\nabla \psi_{(r,\rho_k,\lambda)}^T \nabla \psi_{(r,\rho_k,\lambda)}$ 必须是非奇异的。又因为 $\nabla \psi_{(r,\rho,\lambda)}$ 的行数大于列数, 该矩阵 $\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)}$ 的列满秩可保证的非奇异性。接下来, 我们将给出确保这一条件成立的可行条件。

为方便叙述, 我们定义上下层问题的 Lagrange 函数

$$L^\lambda(x, y, s, w) := F(x, y) + (w - \lambda s)g(x, y), \quad \mathcal{L}^\lambda(x, y, s) := f(x, y) + sg(x, y),$$

以及这两个函数关于变量 (x, y) 的 Hessian 矩阵

$$\nabla^2 L^\lambda := \begin{bmatrix} \nabla_{xx}^2 L^\lambda & \nabla_{yx}^2 L^\lambda \\ \nabla_{xy}^2 L^\lambda & \nabla_{yy}^2 L^\lambda \end{bmatrix}, \quad \nabla(\nabla_y \mathcal{L}^\lambda) := \begin{bmatrix} \nabla_{xy}^2 \mathcal{L}^\lambda & \nabla_{yy}^2 \mathcal{L}^\lambda \end{bmatrix}.$$

此外, 令 $\nabla g(x, y)^T := \begin{bmatrix} \nabla_x g(x, y)^T \\ \nabla_y g(x, y)^T \end{bmatrix}$, 进而计算 $\psi_{(r,\rho,\lambda)}$ 的雅可比矩阵:

$$\nabla \psi_{(r,\rho,\lambda)}(x, y, s, w) = \begin{pmatrix} \nabla^2 L^\lambda & -\lambda \nabla g(x, y) & \nabla g(x, y) \\ \nabla(\nabla_y \mathcal{L}^\lambda) & \nabla_y g(x, y) & 0 \\ \mathcal{J} \nabla g(x, y)^T & \Gamma & 0 \\ \Theta \nabla g(x, y)^T & 0 & \mathfrak{B} \end{pmatrix},$$

其中 $\mathcal{J} := \text{diag}\{\tau_1, \dots, \tau_p\}$, $\Gamma := \text{diag}\{\gamma_1, \dots, \gamma_p\}$, $\Theta := \text{diag}\{\theta_1, \dots, \theta_p\}$, $\mathfrak{B} := \text{diag}\{\beta_1, \dots, \beta_p\}$, 这些矩阵的元素分别定义为

$$\tau_i = \frac{\kappa_i}{z_i + \kappa_i}, \quad \gamma_i = \frac{-\rho z_i}{z_i + \kappa_i}, \quad \theta_i = \frac{\tilde{\kappa}_i}{\tilde{z}_i + \tilde{\kappa}_i}, \quad \beta_i = \frac{-\rho \tilde{z}_i}{\tilde{z}_i + \tilde{\kappa}_i}, \quad i = 1, \dots, p. \tag{17}$$

其中 $z_i = z_i(x, y, s_i, r, \rho)$, $\tilde{z}_i = \tilde{z}_i(x, y, w_i, r, \rho)$, κ_i 和 $\tilde{\kappa}_i$ 同理定义。

类似于文献[25], 我们首先讨论上述元素的相关性质。

引理 3.1. 对于每一 $r > 0$, $\rho > 0$ 和使得 $\psi_{(r,\rho,\lambda)}(x, y, s, w) = 0$ 成立的点 (x, y, s, w) , 以下结论成立:

$$\begin{aligned} \tau_i &> 0, \quad \gamma_i < 0, \quad i = 1, \dots, p, \\ \theta_i &> 0, \quad \beta_i < 0, \quad i = 1, \dots, p. \end{aligned}$$

Proof. 根据引理 2.1 (1), 对于任意 $r > 0$, $\rho > 0$, 该引理显然成立。 ■

基于上述引理, 我们可以推导出保证矩阵 $\nabla \psi_{(r,\rho_k,\lambda)}^T \nabla \psi_{(r,\rho_k,\lambda)}$ 非奇异的充分条件。这将确保高斯牛顿迭代格式(15)是良定的。如同文献[25]所述, 我们只需证明是列满秩的即可。

定理 3.2. 对于某些 $r > 0$, $\rho > 0$, $0 < \lambda < \frac{\gamma_i \theta_i}{\tau_i \beta_i}$, $i = 1, \dots, p$ 和满足(14)的点 $(\bar{x}, \bar{y}, \bar{s}, \bar{w})$, 假设 $\nabla^2 L^\lambda(\bar{x}, \bar{y}, \bar{s}, \bar{w})$ 是正定的。那么, $\nabla \psi_{(r,\rho,\lambda)}(\bar{x}, \bar{y}, \bar{s}, \bar{w})$ 的列向量是线性无关的。

Proof. 考虑任意向量 $d = (d_1^T, d_2^T, d_3^T)^T$ 使得 $\nabla \psi_{(r,\rho,\lambda)}(\bar{x}, \bar{y}, \bar{s}, \bar{w})d = 0$, 其中 $d_1 \in \mathbb{R}^{n+m}$, $d_2 \in \mathbb{R}^p$, $d_3 \in \mathbb{R}^p$ 。因此我们有

$$\nabla^2 L^\lambda(\bar{x}, \bar{y}, \bar{s}, \bar{w})d_1 - \lambda \nabla g(\bar{x}, \bar{y})^T d_2 + \nabla g(\bar{x}, \bar{y})^T d_3 = 0, \tag{18}$$

$$\nabla(\nabla_y \mathcal{L}^\lambda(\bar{x}, \bar{y}, \bar{s}, \bar{w}))d_1 + \nabla_y g(\bar{x}, \bar{y})^T d_2 = 0, \tag{19}$$

$$\tau_i \nabla g_i(\bar{x}, \bar{y})^T d_1 + \gamma_i d_{2i} = 0, \tag{20}$$

$$\theta_i \nabla g_i(\bar{x}, \bar{y})^T d_1 + \beta_i d_{3i} = 0, \tag{21}$$

其中 $i=1, \dots, p$ 。根据引理 3.1, 我们可将式(20)和式(21)写为

$$\nabla g_i(\bar{x}, \bar{y})^T d_1 = -\frac{\gamma_i}{\tau_i} d_{2i}, \quad \nabla g_i(\bar{x}, \bar{y})^T d_1 = -\frac{\beta_i}{\theta_i} d_{3i}, \tag{22}$$

其中 $-\frac{\gamma_i}{\tau_i} > 0$, $-\frac{\beta_i}{\theta_i} > 0$ 。因此, 我们对式(18)左乘 d_1^T :

$$d_1^T \nabla^2 L^{\lambda}(\bar{x}, \bar{y}, \bar{s}, \bar{w}) d_1 - \lambda d_1^T \nabla g(\bar{x}, \bar{y})^T d_2 + d_1^T \nabla g(\bar{x}, \bar{y})^T d_3 = 0. \tag{23}$$

将式(22)带入上式有

$$d_1^T \nabla^2 L^{\lambda}(\bar{x}, \bar{y}, \bar{s}, \bar{w}) d_1 - \lambda \sum_{i=1}^p \left(-\frac{\gamma_i}{\tau_i} \right) d_{2i}^2 + \sum_{i=1}^p \left(-\frac{\beta_i}{\theta_i} \right) d_{3i}^2 = 0. \tag{24}$$

更多地, 由式(22)得

$$d_{2i} = -\frac{\tau_i}{\gamma_i} \nabla g_i(\bar{x}, \bar{y})^T d_1 = \frac{\tau_i}{\gamma_i} \frac{\beta_i}{\theta_i} d_{3i}, \tag{25}$$

并将该式带入式(24)有

$$d_1^T \nabla^2 L^{\lambda}(\bar{x}, \bar{y}, \bar{s}, \bar{w}) d_1 + \sum_{i=1}^p \left(1 - \lambda \frac{\tau_i \beta_i}{\gamma_i \theta_i} \right) \left(-\frac{\beta_i}{\theta_i} \right) d_{3i}^2 = 0. \tag{26}$$

那么, 根据引理 3.1, 由于 $\nabla^2 L^{\lambda}(\bar{x}, \bar{y}, \bar{s}, \bar{w})$ 是正定的, $0 < \lambda < \frac{\gamma_i \theta_i}{\tau_i \beta_i}$, $i=1, \dots, p$, 式(26)是非负项的求和, 则 d_1 和 d_3 只能是零向量。根据式(25), 进而推出 $d_{2i} = 0$, $i=1, \dots, p$ 。证毕。 ■

需要注意的是, $\lambda < \frac{\gamma_i \theta_i}{\tau_i \beta_i}$ 这一假设并不与 λ 严格为正相冲突, 因为根据引理 3.1, 有 $\frac{\gamma_i \theta_i}{\tau_i \beta_i} > 0$ 。

基于上述定理, 我们还需做出以下假设, 以证明 $\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)}$ 的非奇异性。

假设 3.1. 以下矩阵的每一行都是非零向量:

$$\left[\nabla^2 L^{\lambda}(x, y, s, w)^T \quad \nabla(\nabla_y \mathcal{L}(x, y, s, w))^T \quad \nabla g(x, y)^T \right].$$

假设 3.2. 对于 $\lambda > 0$, $r > 0$, $\rho > 0$, 矩阵 $\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)}$ 是严格对角占优的。

引理 3.2. 令假设 3.1 在点 (x, y, s, w) 成立, 那么对于任意的 $\lambda > 0$, $r > 0$, $\rho > 0$, 矩阵 $\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)}$ 的对角线元素是严格正的。

Proof. 定义 a_{ij} , $i, j=1, 2, 3$ 为矩阵的组成元素, 那么该矩阵写为

$$\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

显然, 我们只需考虑元素 a_{11} , a_{22} , a_{33} 我们将这三个元素的表达式列出:

$$\begin{aligned} (a_{11})_{jj} &= \sum_{k=1}^{n+m} (\nabla_{j,k}^2 L^{\lambda})^T \nabla_{j,k}^2 L^{\lambda} + \sum_{k=1}^m \nabla_j (\nabla_{y_k} \mathcal{L})^T \nabla_j (\nabla_{y_k} \mathcal{L}) \\ &\quad + \sum_{k=1}^p \nabla_j g_k(x, y)^T \nabla_j g_k(x, y) (\tau_k)^2 + \sum_{k=1}^p \nabla_j g_k(x, y)^T \nabla_j g_k(x, y) (\beta_k)^2, \end{aligned}$$

其中 ∇_j 是 $\nabla := (\nabla_{x_1}, \dots, \nabla_{x_n}, \nabla_{y_1}, \dots, \nabla_{y_m})$ 的第 j 个元素和 $\nabla_{j,k}^2$ 是以下矩阵的第 j 行第 k 列元素:

$$\begin{pmatrix} \nabla_{x_1 x_1} & \dots & \nabla_{x_1 x_n} & \nabla_{x_1 y_1} & \dots & \nabla_{x_1 y_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \nabla_{x_n x_1} & \dots & \nabla_{x_n x_n} & \nabla_{x_n y_1} & \dots & \nabla_{x_n y_m} \\ \nabla_{y_1 x_1} & \dots & \nabla_{y_1 x_n} & \nabla_{y_1 y_1} & \dots & \nabla_{y_1 y_m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \nabla_{y_m x_1} & \dots & \nabla_{y_m x_n} & \nabla_{y_m y_1} & \dots & \nabla_{y_m y_m} \end{pmatrix}.$$

结合假设 3.1 和引理 3.1, 易得 $(a_{11})_{jj} > 0, j = 1, \dots, n + m$ 。类似地, 其他对角线元素为

$$\begin{aligned} (a_{22})_{ii} &= \nabla g_i(x, y) \nabla g_i(x, y)^T + (\gamma_i)^2, \\ (a_{33})_{ii} &= \nabla g_i(x, y) \nabla g_i(x, y)^T + (\beta_i)^2, \quad i = 1, \dots, p. \end{aligned}$$

根据引理 4.1, 易得这两个元素也均大于 0。证毕。 ■

定理 3.3. 对于某些 $\lambda > 0, r > 0, \rho > 0$, 令 (x, y, s, w) 是满足系统(14)的稳定点。若假设 3.1 和假设 3.2 在该点成立, 则 $\nabla \psi_{(r, \rho, \lambda)}^T \nabla \psi_{(r, \rho, \lambda)}$ 是非奇异的。

Proof. 该结果可参考文献[13]中定理 4.8 类似证明。 ■

为证明算法 1 的收敛性, 我们还需证明步长的存在性。令 (x_k, y_k, s_k, w_k) 是第 k 次迭代点, 则我们有以下结果。

引理 3.3. 对于 $\lambda > 0, r > 0, \rho > 0$, 假设函数 F, f 和 g 是二阶连续可微的, 那么存在步长 $\alpha_k \in (0, 1]$ 使得式(16)成立。

定义残差函数 $\phi_{(r, \rho, \lambda)}(x, y, s, w) = \frac{1}{2} \|\psi_{(r, \rho, \lambda)}(x, y, s, w)\|^2$, 下列结论成立。

引理 3.4. 令 $z_{k+1}(\rho) = z(x_{k+1}, y_{k+1}, s_{k+1}, r, \rho), \tilde{z}_{k+1}(\rho) = z(x_{k+1}, y_{k+1}, w_{k+1}, r, \rho)$ 。假设 $\|z_{k+1}(\rho_k) + g(x_{k+1}, y_{k+1})\| \neq 0, \|\tilde{z}_{k+1}(\rho_k) + g(x_{k+1}, y_{k+1})\| \neq 0$ 。若 $\rho_k \geq \rho_{k+1} > 0$ 则有

$$\phi_{(r, \rho_{k+1}, \lambda)}(x_{k+1}, y_{k+1}, s_{k+1}, w_{k+1}) \leq \phi_{(r, \rho_k, \lambda)}(x_k, y_k, s_k, w_k).$$

Proof. 根据 z 和 \tilde{z} 表达式和引理 2.1, 易计算出

$$\frac{\partial z_i}{\partial \rho} = \frac{1}{\rho} \frac{z_i}{z_i + \kappa_i} (z_i + g_i), \quad \frac{\partial \tilde{z}_i}{\partial \rho} = \frac{1}{\rho} \frac{\tilde{z}_i}{\tilde{z}_i + \tilde{\kappa}_i} (\tilde{z}_i + g_i),$$

其中 $i = 1, \dots, p$, 进一步可得

$$\begin{aligned} & \left. \frac{D\phi_{(r, \rho, \lambda)}(x_{k+1}, y_{k+1}, s_{k+1}, w_{k+1})}{D\rho} \right|_{\rho=\rho_k} \\ &= \frac{1}{\rho_k} (\hat{z}_{k+1} + g(x_{k+1}, y_{k+1}))^T (Z_{k+1} + K_{k+1})^{-1} Z_{k+1} (\hat{z}_{k+1} + g(x_{k+1}, y_{k+1})) \\ & \quad + \frac{1}{\rho_k} (\tilde{z}_{k+1} + g(x_{k+1}, y_{k+1}))^T (\tilde{Z}_{k+1} + \tilde{K}_{k+1})^{-1} Z_{k+1} (\tilde{z}_{k+1} + g(x_{k+1}, y_{k+1})) \\ & > 0, \end{aligned}$$

其中 $\hat{z}_{k+1} = z(x_{k+1}, y_{k+1}, s_{k+1}, r, \rho_k), \tilde{z}_{k+1} = z(x_{k+1}, y_{k+1}, w_{k+1}, r, \rho_k), Z_{k+1} = \text{diag}(\hat{z}_{k+1}), K_{k+1} = \text{diag}(\hat{\kappa}_{k+1}),$

$\tilde{Z}_{k+1} = \text{diag}(\tilde{z}_{k+1})$, $\tilde{K}_{k+1} = \text{diag}(\tilde{\kappa}_{k+1})$ 。因此, 根据上式, 可推出当 $\rho > 0$ 时, $\phi_{(r,\rho,\lambda)}(x_{k+1}, y_{k+1}, s_{k+1}, w_{k+1})$ 关于 ρ 是单调递增的, 从而得出最终结论。证毕。 ■

以下是算法 1 收敛性定理。

定理 3.4. 若定理 3.1 中的假设和假设 3.1~3.2 在 (x_k, y_k, s_k, w_k) 处成立, 那么

$$\lim_{k \rightarrow \infty} \phi_{(r,\rho_k,\lambda)}(x_k, y_k, s_k, w_k) = 0.$$

备注 在上述假设下, 我们提出的光滑化的高斯牛顿法在理论上表现出良好的收敛性。但在实际生活中, 这些关键假设可能难以完全满足, 从而对算法的适用性带来挑战。接下来, 我们对这些假设的局限性以及潜在影响进行系统分析: 关于惩罚参数 λ 的后验条件对于算法的收敛以及后续数值实验的结果具有重要的影响, 若惩罚参数过大, 可能会导致惩罚过度从而数值不稳定, 若惩罚参数过小, 则可能无法保证约束从而导致解不可行。严格对角占优假设的成立保证了高斯牛顿法中雅可比矩阵的非奇异性, 从而保证了算法的良好性, 但在实际问题中该条件可能过于严格, 在高维或非凸问题中可能难以验证, 所以我们数值实验考虑解决小规模问题, 若该条件不成立, 可能导致迭代矩阵病态, 进而导致算法收敛到非稳定点, 对于该问题实际中可通过正则化方法解决, 但这会增加算法的复杂度。

4. 数值实验

本节中数值实验聚焦于求解系统(14)。基于该系统, 我们比较了高斯牛顿法[25], LM (Levenberg-Marquardt)算法[25], 信赖域算法[26], 拟牛顿法[27]以及 Nelder-Mead 单纯形法的测试结果。实验所用案例均来自双层优化测试问题库(BOLIB) [28], 该库包含 124 个非线性测试案例。本实验是在配置 Intel(R) Core(TM) Ultra 5 125H 3.60 GHz 处理器, Windows11 系统的笔记本电脑上进行, 实验平台为 Matlab2023b。

在算法 1 的第 0 步及其他算法中, 我们将容差设为 $\varepsilon = 1e-6$, 最大迭代次数设为 $K = 1000$ 。所有实验均采用六个不同的惩罚参数值, 即 $\lambda \in \{0.01, 0.1, 1, 10, 100, 1000\}$ 。采取不同的 λ 值的原因是为了避免对下层目标值与最优值的过度惩罚或惩罚不足。本节后续部分将展示各算法在 CPU 时间, 相对误差以及残差等方面的性能比较。默认设置下, 我们以 $x_0 = 1_n$, $y_0 = 1_m$ 作为初始点。若默认起始点不满足至少一个约束条件时, 则选择可行起始点(详见文献[28])。对于拉格朗日乘子, 按以下方式进行初始化: $s_0 = w_0 = 1_p$ 。

4.1. 时间性能曲线

性能曲线图[29]被广泛用于比较不同算法的特性。本小节我们用性能曲线图分析各算法的 CPU 时间, $t_{i,j}$ 表示算法 j 求解问题 i 所需的平均时间。若问题 i 的最优解已知但无法被算法 j 求解(即上层目标函数误差 $> 60\%$), 设定 $t_{i,j} = \infty$ 。定义时间性能比率:

$$r_{i,j} := \frac{t_{i,j}}{\min\{t_{i,j} : j \in J\}},$$

其中 J 表示算法集合。性能比是指算法 j 在解决问题 i 时的时间与集合 J 中表现最佳的算法所用时间的比值。定义算法 j 性能比的累积分布函数:

$$\rho_j(\tau) := \frac{|\{i \in I : r_{i,j} \leq \tau\}|}{n_i},$$

其中 I 表示问题集合。性能剖面图 $\rho_j(\tau)$ 用于统计算法 j 的性能比优于 τ 的问题数量。该性能剖面函数 $\rho_j: \mathbb{R} \rightarrow [0,1]$ 关于 τ 是非递减函数, 其中 $\rho_j(1)$ 表示算法 j 表现最优的问题数与全部问题的比例。

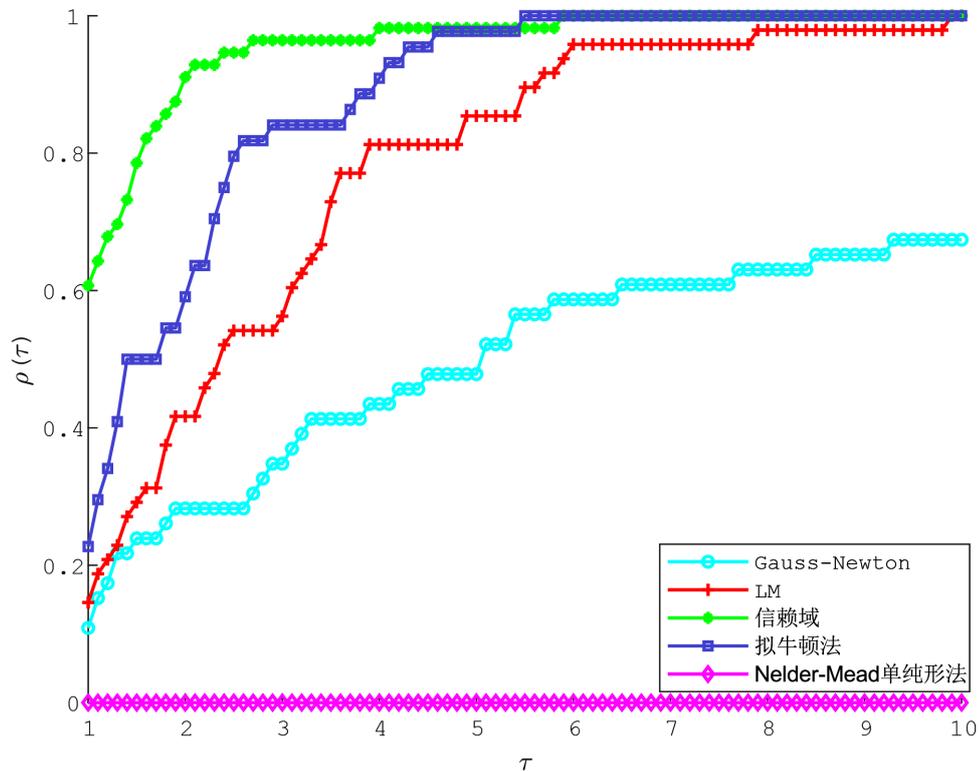


Figure 1. CPU time performance curves of different algorithms
图 1. 不同算法的 CPU 时间性能曲线图

上图 1 为不同算法性能比的折线图，曲线越高表示算法性能越优。纵轴表示算法 CPU 时间性能比率小于 τ 的问题占比。由图 1 可得，除单纯形法，高斯牛顿法的表现弱于其他算法。根据 $\rho_j(1)$ 的数值可知：高斯牛顿法在约 10% 的问题最快，LM 算法约 15% 的问题最快，拟牛顿法约 22% 的问题最快，而信赖域算法约在 60% 的问题中最快。图中还显示信赖域算法的 $\rho_j(2)$ 就已经达到了 90%，而其他算法最多的也就只有约 60%，这说明在 CPU 时间指标上，信赖域算法的性能明显优于其他算法。但是我们的实验针对的是小规模问题，还需通过以下表 1 进一步分析：

Table 1. Average execution time of algorithms
表 1. 各算法的平均运行时间

算法	高斯牛顿法	LM	信赖域	拟牛顿法	单纯形法
平均时间(秒)	6.68E-02	1.61E-02	8.16E-03	1.04E-02	4.02E+00

上述表格，高斯牛顿法，LM 算法和拟牛顿法在时间上量级相同，说明这三种算法的时间性能相当。除此之外，无论从性能曲线还是平均时间上，单纯形法显著差于其他算法，而信赖域算法优于其他算法。

4.2. 目标函数的相对误差

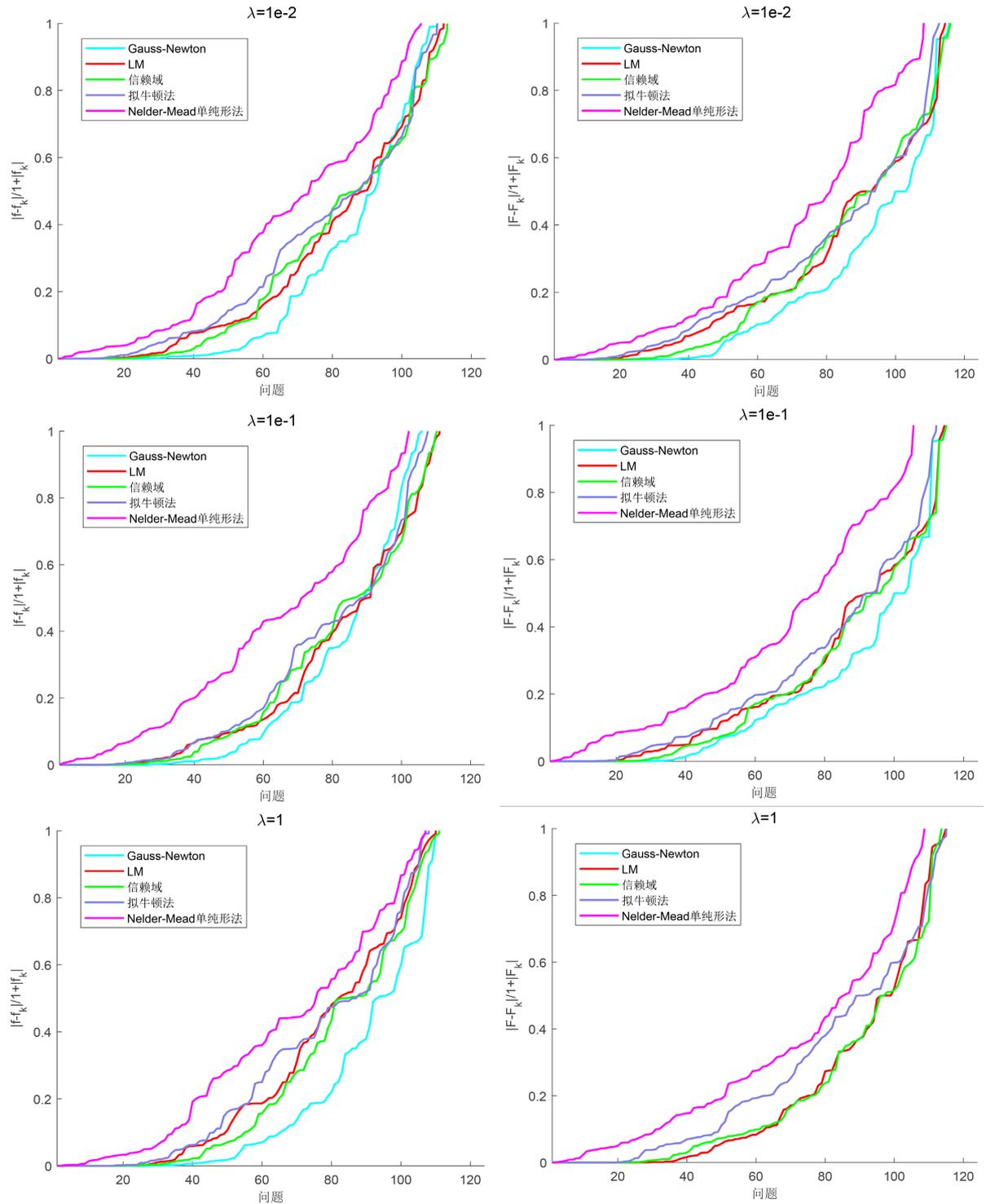
本小节我们将比较不同算法计算所得的上下层目标函数值。为进行此项比较，我们仅关注 116 个 BOLIB 中的案例，因为有部分算法在其他 7 个案例中未求出明确的解，例如高斯牛顿法在 NieEtal2017e 案例中可能因矩阵的奇异性而导致发散(详见文献中[30])。

定义 F ， f 为通过某算法求得的上下层目标函数值， F_k ， f_k 为文献[28]中 BOLIB 库中记录的最优解

对应的最优值，以及相对误差的公式定义为

$$uppererror = \frac{|F - F_k|}{1 + |F_k|}, \quad lowererror = \frac{|f - f_k|}{1 + |f_k|}.$$

以下图像分别展示了不同参数下不同算法的相对误差对比结果，图像按误差升序排列。



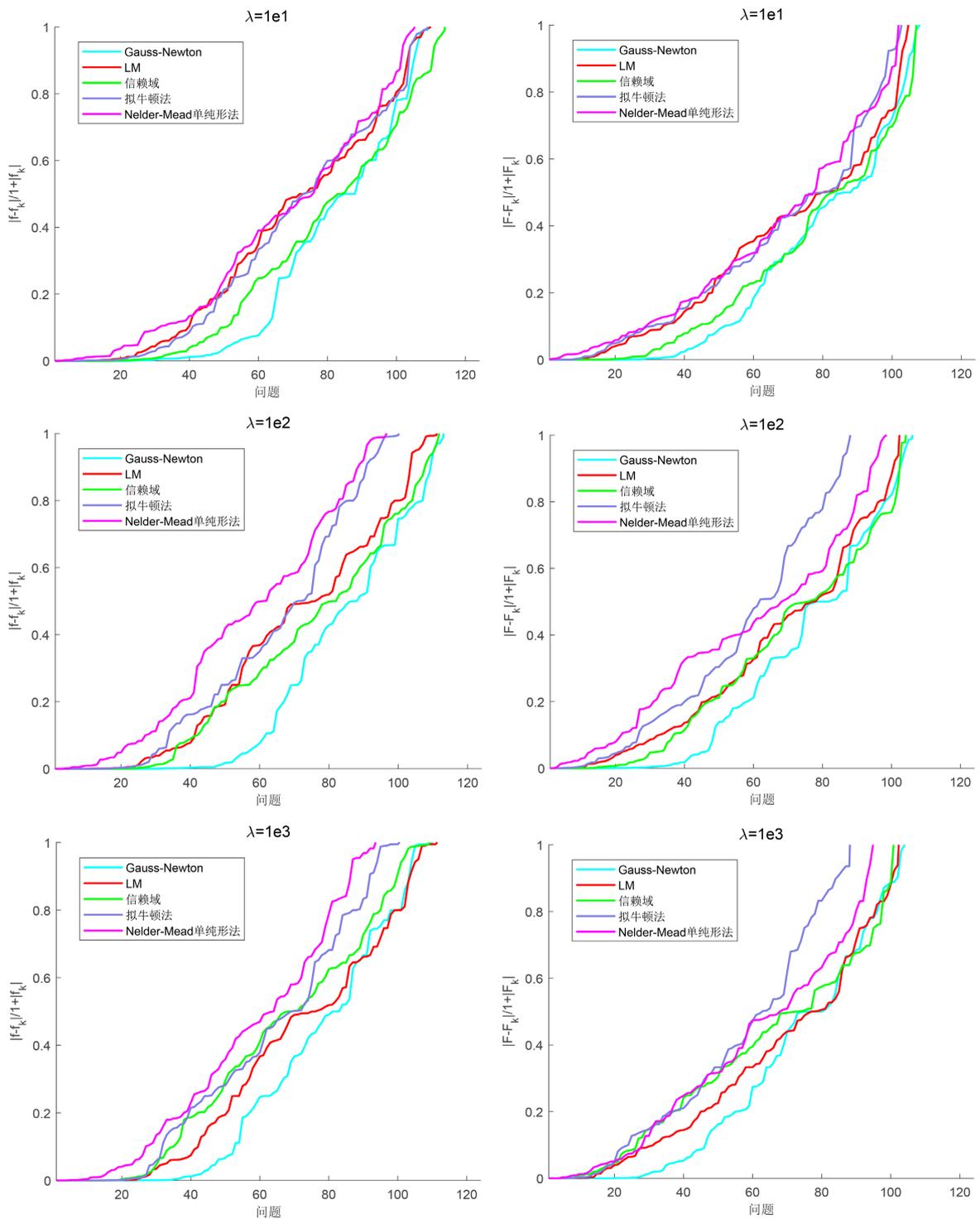


Figure 2. Comparison chart of the relative errors in the upper-level and lower-level objective functions across algorithms under different parameters

图 2. 不同参数下各算法上下层目标函数相对误差对比图

从图 2 中可以看出，对于约 80% 的问题，高斯牛顿法产生的误差小于其他算法，说明该算法比其他算法表现好，尤其是在 $\lambda = 100$ 时明显能看出高斯牛顿法上层目标函数的误差曲线低于其他曲线。具体

而言，在约 50% 测试问题中，高斯牛顿法所得解产生的误差可忽略不计(误差小于 5%)。在 6%~30% 的误差范围内，当罚参数分别为 0.01, 0.1, 1 时，除单纯性法外，该算法与其他算法差别不大，还原了约 70% 的解。在其他参数时，高斯牛顿法还原了近 60% 的解，而其他算法表现最佳的也仅还原了约 50% 的解。

4.3. 停止准则中的残差变化

在本小节中我们将评估算法 1 在放宽停止条件中容差时的表现。具体而言，在实际的数值实验中，当容差为 $1e-8$ 时，多数案例无法实现该精度。因此，当算法达到最大迭代次数或步长过小时，算法终止。将不同算法在不同参数条件下终止时的最小残差画图如下：

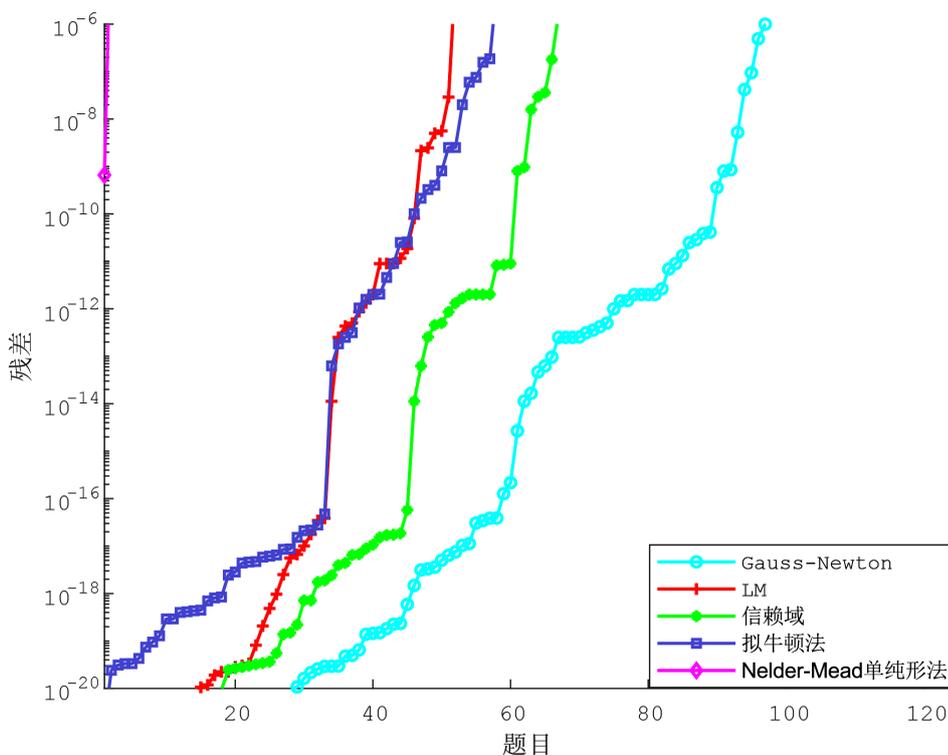


Figure 3. Residual performance comparison of algorithms
图 3. 各算法残差性能对比图

上图 3 显示的是各算法最终生成的 KKT 残差数值按升序排列结果。具体来看，从图 3 可得高斯牛顿法在 95 个案例中表现较好，在这些案例中残差值均小于 $1e-8$ 。这表明在这些案例中，当问题可被近乎求到精确解时，该算法的停止条件更为严格，从而能得到较优的数值解。整体来看，高斯牛顿法明显优于其他算法，在同一容差下，其他算法仅能求解约 50%~60% 的问题，因此，该算法更易达到稳定点。

5. 结论

本文中，基于对数障碍法与增广拉格朗日函数，构造了下层问题近似值函数，进而求得问题(VP)的近似问题。基于光滑化方法，我们将该近似问题重新表述为一个非线性方程组。理论分析表明，本文中所采用的高斯牛顿法具有良定性，并在一定条件下可证明该算法的收敛性。在数值实验，我们分别从时间，相对误差和残差三方面对比了我们的算法与 LM 算法，信赖域算法以及拟牛顿法的结果。实验结果表明，在时间性能方面，当 $\nabla \psi_{(r,\rho,\lambda)}^T \nabla \psi_{(r,\rho,\lambda)}$ 在迭代中不存在病态条件时，除单纯形法外，四种算法均展

现出极快的运算速度,在不同的罚参数下平均运行时间均低于 0.1 秒,但根据时间性能曲线,我们的算法在运算速度上慢于其他三种算法,这可以作为我们今后的改进工作;在相对误差方面,该模型适用于多数双层规划问题,高斯牛顿法在不同参数下可还原 60%~70%的问题的已知最优解,尤其当惩罚参数为 100 时,我们的算法的精度明显比其他算法要高,但是我们的实验只列举了六种参数下的结果,对于惩罚参数 λ 的选取规则一直是双层规划中的一大难题,后续我们将其作为一个重要的研究方向;在残差方面,高斯牛顿法明显优于其他算法,这表明我们的算法在求解双层规划问题上相较于几种经典算法具有更高的精度。综上,在 BOLIB 测试库的 124 个非线性案例中,高斯牛顿法展现出显著优势:不仅成功求解率最高,其残差下降速度也明显优于其他对比算法。

参考文献

- [1] Luo, Z.Q., Pang, J.S. and Ralph, D. (1996) *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511983658>
- [2] 吴昌龙, 罗华伟, 秦正斌, 等. 基于双层规划的电网侧储能配置算法研究[J]. 低压电器, 2021(4): 85-93.
- [3] 邵举平, 周将军, 孙延安. “碳中和”背景下基于演化博弈与双层规划的供应链减排动力研究[J]. 运筹与管理, 2023, 32(12): 79-85.
- [4] Zhang, Y., Sun, L., Sun, W., Ma, F., Xiao, R., Wu, Y., et al. (2024) Bilevel Optimal Infrastructure Planning Method for the Inland Battery Swapping Stations and Battery-Powered Ships. *Tsinghua Science and Technology*, **29**, 1323-1340. <https://doi.org/10.26599/tst.2023.9010138>
- [5] Vicente, L., Savard, G. and Júdice, J. (1994) Descent Approaches for Quadratic Bilevel Programming. *Journal of Optimization Theory and Applications*, **81**, 379-399. <https://doi.org/10.1007/bf02191670>
- [6] Dempe, S. and Dutta, J. (2012) Is Bilevel Programming a Special Case of a Mathematical Program with Complementarity Constraints. *Mathematical Programming*, **131**, 37-48. <https://doi.org/10.1007/s10107-010-0342-1>
- [7] Dempe, S. and Zemkoho, A.B. (2013) The Bilevel Programming Problem: Reformulations, Constraint Qualifications and Optimality Conditions. *Mathematical Programming*, **138**, 447-473. <https://doi.org/10.1007/s10107-011-0508-5>
- [8] Dempe, S. and Franke, S. (2019) Solution of Bilevel Optimization Problems Using the KKT Approach. *Optimization*, **68**, 1471-1489. <https://doi.org/10.1080/02331934.2019.1581192>
- [9] Ye, J.J., Zhu, D.L. and Zhu, Q.J. (1997) Exact Penalization and Necessary Optimality Conditions for Generalized Bilevel Programming Problems. *SIAM Journal on Optimization*, **7**, 481-507. <https://doi.org/10.1137/s1052623493257344>
- [10] Scheel, H. and Scholtes, S. (2000) Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity. *Mathematics of Operations Research*, **25**, 1-22. <https://doi.org/10.1287/moor.25.1.1.15213>
- [11] Ye, J.J. and Zhang, J. (2014) Enhanced Karush-Kuhn-Tucker Conditions for Mathematical Programs with Equilibrium Constraints. *Journal of Optimization Theory and Applications*, **163**, 777-794. <https://doi.org/10.1007/s10957-013-0493-3>
- [12] Liu, X., Perakis, G. and Sun, J. (2006) A Robust SQP Method for Mathematical Programs with Linear Complementarity Constraints. *Computational Optimization and Applications*, **34**, 5-33. <https://doi.org/10.1007/s10589-005-3075-y>
- [13] Lampariello, L. and Sagratella, S. (2020) Numerically Tractable Optimistic Bilevel Problems. *Computational Optimization and Applications*, **76**, 277-303. <https://doi.org/10.1007/s10589-020-00178-y>
- [14] Dempe, S., Dinh, N., Dutta, J. and Pandit, T. (2021) Simple Bilevel Programming and Extensions. *Mathematical Programming*, **188**, 227-253. <https://doi.org/10.1007/s10107-020-01509-x>
- [15] Outrata, J.V. (1990) On the Numerical Solution of a Class of Stackelberg Problems. *Zeitschrift für Operations Research*, **34**, 255-277.
- [16] Ye, J.J. and Zhu, D.L. (1995) Optimality Conditions for Bilevel Programming Problems. *Optimization*, **33**, 9-27. <https://doi.org/10.1080/02331939508844060>
- [17] Lin, G., Xu, M. and Ye, J.J. (2014) On Solving Simple Bilevel Programs with a Nonconvex Lower Level Program. *Mathematical Programming*, **144**, 277-305. <https://doi.org/10.1007/s10107-013-0633-4>
- [18] Sun, D. and Qi, L. (1999) On NCP-Functions. *Computational Optimization and Applications*, **13**, 201-220. <https://doi.org/10.1023/a:1008669226453>
- [19] Fischer, A., Zemkoho, A.B. and Zhou, S. (2022) Semismooth Newton-Type Method for Bilevel Optimization: Global Convergence and Extensive Numerical Experiments. *Optimization Methods and Software*, **37**, 1770-1804. <https://doi.org/10.1080/10556788.2021.1977810>

-
- [20] Jolaoso, L.O., Mehlitz, P. and Zemkoho, A.B. (2024) A Fresh Look at Nonsmooth Levenberg-Marquardt Methods with Applications to Bilevel Optimization. *Optimization*, **2024**, 1-48. <https://doi.org/10.1080/02331934.2024.2313688>
- [21] Rockafellar, R.T. and Wets, R.J-B. (1998) Variational Analysis. Springer.
- [22] Liu, X.W. and Dai, Y.H. (2020) A Globally Convergent Primal-Dual Interior-Point Relaxation Method for Nonlinear Programs. *Mathematics of Computation*, **89**, 1301-1329. <https://doi.org/10.1090/mcom/3487>
- [23] Liu, X.W., Dai, Y.H. and Huang, Y.K. (2022) A Primal-Dual Interior-Point Relaxation Method with Global and Rapidly Local Convergence for Nonlinear Programs. *Mathematical Methods of Operations Research*, **96**, 351-382. <https://doi.org/10.1007/s00186-022-00797-7>
- [24] Fliege, J., Tin, A. and Zemkoho, A. (2021) Gauss-Newton-Type Methods for Bilevel Optimization. *Computational Optimization and Applications*, **78**, 793-824. <https://doi.org/10.1007/s10589-020-00254-3>
- [25] Tin, A. and Zemkoho, A.B. (2023) Levenberg-Marquardt Method and Partial Exact Penalty Parameter Selection in Bilevel Optimization. *Optimization and Engineering*, **24**, 1343-1385. <https://doi.org/10.1007/s11081-022-09736-1>
- [26] 唐江花. 求解非线性方程组的信赖域算法[J]. 吉林化工学院学报, 2021, 38(5): 85-89.
- [27] Vater, N. and Borzi, A. (2025) Convergence of a Quasi-Newton Method for Solving Systems of Nonlinear Underdetermined Equations. *Computational Optimization and Applications*, **91**, 973-996. <https://doi.org/10.1007/s10589-024-00606-3>
- [28] Zhou, S.L., Zemkoho, A.B. and Tin, A. (2020) BOLIB: Bilevel Optimization Library of Test Problems. In: *Springer Optimization and Its Applications*, Springer, 563-580. https://doi.org/10.1007/978-3-030-52119-6_19
- [29] Dolan, E.D. and Moré, J.J. (2002) Benchmarking Optimization Software with Performance Profiles. *Mathematical Programming*, **91**, 201-213. <https://doi.org/10.1007/s101070100263>
- [30] Fliege, J., Tin, A. and Zemkoho, A.B. (2020) Supplementary Material for Gauss-Newton-Type Methods for Bilevel Optimization. School of Mathematical Sciences, University of Southampton.