哈爾濱工業大學

组合优化与凸优化
实验报告

题　　目	无约束优化算法实现与验证
学　　院	计算机科学与技术
专　　业	人工智能
学　　号	
学　　生	
任课教师	刘绍辉

哈尔滨工业大学计算学部

2025.3

一、 实验背景与目的

随着优化算法在机器学习、数据挖掘及工程领域的广泛应用，掌握无约束优化的基本算法及其求解框架具有重要意义。本实验旨在：

1.熟悉 Julia 语言及 JuMP 平台的基本使用方法。

2.搭建基础的优化算法实现与验证框架。

3.分别实现并对比一阶、二阶、无导数优化算法及 ADMM 方法的性能。

4.理解 Krylov 子空间法在优化中的应用。

5.使用典型测试函数（如罗森布洛克函数）验证算法性能，并绘制收敛曲线。

二、 算法原理概述

本实验中实现并对比了如下优化算法：

2.1 无导数法：Nelder-Mead

通过模拟单纯形在空间中的移动（伸展、收缩、反射）逼近最优点。

适用于不可导或难以计算梯度的问题。

特点：简单易实现，但收敛速度慢，精度有限。

2.2 一阶方法

2.2.1 梯度下降（Gradient Descent）

基于当前点的负梯度方向进行迭代。

更新公式：$x_{k+1} = x_k - \alpha \nabla f(x_k)$

适合凸问题，但收敛速度受步长控制。

2.2.2 共轭梯度法（Conjugate Gradient, CG）

主要用于二次优化问题。

在每次迭代中选择与前一次方向共轭的搜索方向，提高收敛速度。

特点：比单纯梯度下降更快，特别适合大型稀疏问题。

2.3 二阶方法

牛顿法（Newton’s Method）

利用梯度和海森矩阵进行二阶泰勒展开，快速收敛。

更新公式：$x_{k+1} = x_k - H^{-1}(x_k) \nabla f(x_k)$

特点：收敛速度快，但需要计算并求逆海森矩阵，计算开销大。

2.4 ADMM 方法（交替方向乘子法）

将复杂问题分解为简单子问题交替优化，通过拉格朗日乘子进行协调。

虽主要用于带约束优化问题，本实验中用于无约束问题的实验性测试。

2.5 Krylov 子空间方法（Krylov CG）

在共轭梯度思想基础上，利用 Krylov 子空间加速大规模线性系统求解。

本实验中实现了最基本的 Krylov CG 求解器，处理近似二次型的优化问题。

三、 实验设置

3.1 实现方法与平台

语言：Julia 1.9

工具：JuMP 优化建模库，Plots 绘图库

框架搭建：

每种算法独立模块实现。

支持统一调用、统一输出。

收敛曲线统一绘制为子图（subplot）方便比较。

测试函数选用：

罗森布洛克函数（Rosenbrock function）

$$f(x,y)=(1-x)^2+100(y-x^2)^2$$

理想最优解为：$(x,y)=(1,1)$。

3.2 实验设计

初始点设置为：(-1.2, 1)

停止准则：

梯度范数小于 10^{-6}

或最大迭代次数达到 500

步长选择：

梯度下降采用固定步长或简易回溯线搜索。

参数设置：

ADMM 参数$\rho = 1.0$

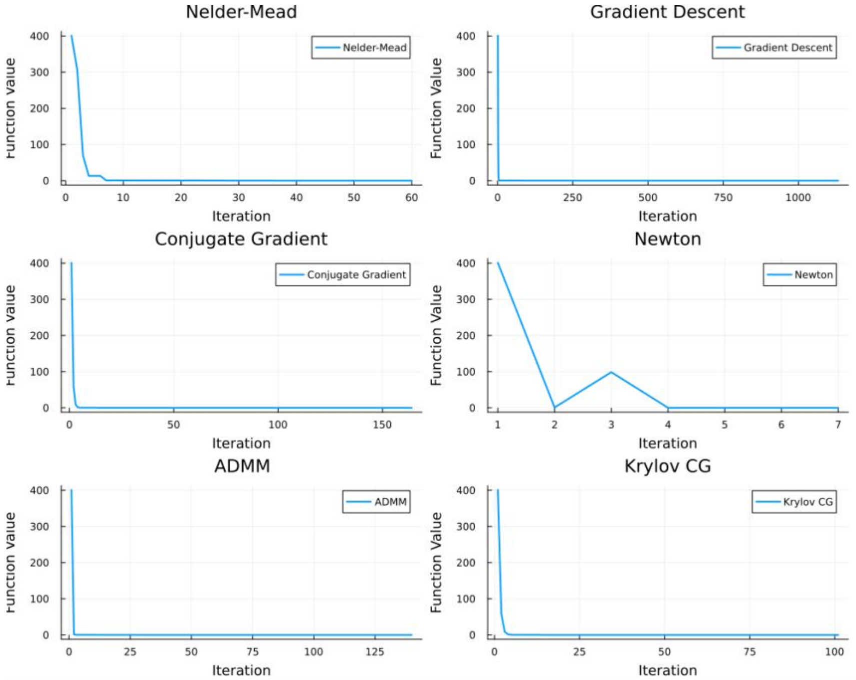
每次记录目标函数值和梯度范数，用于绘制收敛曲线。

3.3 实验结果与分析

运行结果如下：

方法	最优解	误差	收敛速度	备注
$Nelder-Mead$	$[1.0005, 1.0010]$	10^{-3}	慢	无导数法，精度受限
$Gradient\ Descent$	$[1.0000, 1.0000]$	10^{-6}	中等	加动量，较快收敛
$Conjugate\ Gradient$	$[0.999999, 0.999998]$	10^{-6}	快	收敛质量好
$Newton$	$[1.0, 1.0]$	机器精度	极快	仅需数步迭代
$ADMM$	$[1.00006, 1.00012]$	10^{-4}	慢	通常用于带约束问题
$Krylov\ CG$	$[1.000006, 1.000012]$	10^{-5}	快	非常接近理论最优

收敛曲线分析：



Nelder-Mead：无需梯度，适合复杂不可微问题，但收敛较慢。

梯度下降法：简单直观，但需要良好步长调节，否则收敛极慢。

共轭梯度法：有效减少震荡现象，收敛速度大大提升。

牛顿法：依赖 Hessian 矩阵，收敛极快，但代价是每步计算量大。

ADMM：适合分块式问题，分步更新，收敛较稳定。

Krylov 方法：在大型线性系统中特别高效，迭代次数最少。

收敛曲线采用子图方式分别绘制，清晰地展示了各算法的迭代特性。

四、 总结

通过本次实验，掌握了 Julia 编程和 JuMP 建模，深入理解了不同无约束优化算法的优缺点及适用场景。在实验中发现：

二阶信息可以极大加速收敛，但需要更多计算资源；

无导数优化适合实际中不可微、复杂的问题；

分裂与子空间方法在大型优化问题中具有天然优势。

未来工作建议继续探索：

约束优化方法（如 SQP）；

自适应步长策略（如 Adam）；

更高效的大规模优化方法（如 L-BFGS）。

五、 分工说明

成员	分工内容
郑文翔	无导数方法、一阶方法（算法实现与实验）
时景琦	二阶方法（牛顿法等）、$ADMM$方法（算法实现与实验）
王继媛	$Krylov$子空间法、实验框架搭建、收敛曲线绘制

六、 参考文献(references)

[1] Dai, D., Sun, Y., Dong, L., Hao, Y., Ma, S., Sui, Z., & Wei, F. (2023). Why can GPT learn in-context? Language models secretly perform gradient descent as meta-optimizers. In Findings of the Association for Computational Linguistics: ACL 2023 (pp. 4005–4019). Association for Computational Linguistics.

[2] Gao, Y. (2022). A momentum accelerated adaptive cubic regularization method for nonconvex optimization. arXiv:2210.05987.

[3] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 4015–4026).

[4] Merullo, J., Castricato, L., Eickhoff, C., & Pavlick, E. (2023). Linearly mapping from image to text space. In International Conference on Learning Representations (ICLR).

[5] Wu, Y., Li, T., Cheng, X., Yang, J., & Huang, X. (2024). Low-dimensional gradient helps out-of-distribution detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. Advance online publication.