

中国科学技术大学
University of Science and Technology of China

R-linear convergence of the Barzilai and Borwein gradient method

Y.-H. Dai & L.-Z. Liao, IMA J. Numer. Anal., 22(1): 1–10, 2002

杨书泰

最优化算法汇报展示

中国科学技术大学 数学科学学院

2026 年 5 月

1. 基本信息
2. 问题与动机
3. 算法设计与核心证明思路
4. 假设的必要性讨论
5. 数值实验: 复现与解读
6. 个人见解: 亮点、缺陷、与我的两个改进
7. 参考文献与延伸阅读
8. 总结

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

论文信息

题目 R-linear convergence of the Barzilai and Borwein gradient method
作者 Yu-Hong Dai (戴🄵虹), Li-Zhi Liao (廖立志)
单位 中科院计算数学所 (戴); 香港浸会大学 (廖)
期刊 *IMA Journal of Numerical Analysis*, 22(1): 1–10
年份 2002 (投稿 1999.09, 终稿 2001.04)
引用数 600+ (Google Scholar); 本领域里程碑文献之一

汇报人

姓名 杨书泰

学院 数学科学学院
年级 大四
分工 选题、文献阅读、理论梳理、数值复现、Beamer 制作 (独立完成)

选题动机: 论文是 BB 法 *R*-线性收敛的奠基性结果, 也是我个人毕设的主要前置工作之一。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

优化问题: 严格凸二次最小化 $\min_{x\in\mathbb{R}^n} f(x)=\frac{1}{2}x^TAx-b^Tx, A=A^T\succ 0, A\in\mathbb{R}^{n\times n}$。

梯度迭代: $x_{k+1}=x_k-\alpha_k^{-1}g_k, g_k=Ax_k-b$。问题归结到:**如何选择步长 α_k?**

Barzilai & Borwein (1988) 的拟牛顿动机: 把 $D_k=\alpha_k I$ 视作 Hessian 的标量近似, 要求满足割线条件 $D_k s_{k-1}\approx y_{k-1}$, 即

$$D_k=\arg\min_{D=\alpha I}\|Ds_{k-1}-y_{k-1}\|^2 \quad\Longrightarrow\quad \boxed{\alpha_k=\frac{s_{k-1}^Ts_{k-1}}{s_{k-1}^Ty_{k-1}}}$$

其中 $s_{k-1}=x_k-x_{k-1}, y_{k-1}=g_k-g_{k-1}$。

为什么有意思? 在二次模型下,$y_{k-1}=As_{k-1}$, 所以

$$\alpha_k=\frac{s_{k-1}^Ts_{k-1}}{s_{k-1}^TAs_{k-1}}=\frac{g_{k-1}^Tg_{k-1}}{g_{k-1}^TAg_{k-1}}\in[\lambda_1,\lambda_n]$$

是 A 在前一步梯度处的 **Rayleigh 商** ——α_k **扫过** $[\lambda_1,\lambda_n]$, **自适应锁定不同尺度的特征值**。

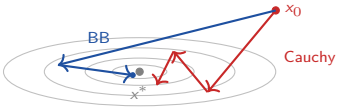
实验观察 (BB 1988, Fletcher 1990): BB 在数值上远比 Cauchy 最速下降快, 与 CG 可比。

反直觉的特点:

- **无单调下降**:$f(x_{k+1})$ 经常**大于** $f(x_k)$, 看上去“违背常识”。
- **步长出乎意料地大**:$\alpha_k \approx 1/\lambda_i$ 对某些 i, 不一定是 $1/\lambda_n$ (最速下降)。
- **算法本身极简**: 每步只多算一个内积。

直觉图像: 在二维 $f(x) = \frac{1}{2}(10x_1^2 + x_2^2)$ 上,

- Cauchy SD: 被大特征值方向“拖死”, 出现 **Z 字形锯齿**;
- BB: 步长在两个方向间**自适应交替**, 几步内突破。



已有的全局化与推广:

- Raydan (1997): BB + Grippo–Lampariello–Lucidi 非单调线搜索 $\Rightarrow$ **一般非线性全局收敛**, 数值上与 CG 可比;
- Birgin–Martínez–Raydan (2000): **推广到凸约束** (SPG, Spectral Projected Gradient);
- Liu–Dai (1999): **与随机近似结合处理强噪声问题**。

已有收敛结果与本文要解决的痛点

文献	结果	假设/局限
Barzilai–Borwein 1988	$n = 2$: R-超线性	仅二维
Raydan 1993	n 维: $g_k \to 0$	无收敛速率
Molina–Raydan 1996	n 维: Q-线性收敛	要求 $\lambda_n < 2\lambda_1$ (即 $\kappa < 2$)
Dai–Liao 2002 (本文)	**n 维: R-线性收敛**	**无附加条件**

为什么 $\kappa < 2$ 的限制是“致命”的? 实际工程问题条件数动辄 $10^4 \sim 10^8$, 该假设几乎**从未被满足**。

R-线性 vs Q-线性:

- Q-线性: $\|g_{k+1}\| \le \rho \|g_k\|$, 要求**每步**严格收缩。
- R-线性: $\|g_k\| \le C\rho^k \|g_0\|$, 允许**震荡但有指数包络**。
- BB 在数值上明显震荡 ($\|g_k\|$ 非单调) $\Rightarrow$ **R-线性才是“正确”的描述**。

本文目标: 在最弱的假设 (严格凸二次, 无附加条件) 下, 把 Raydan 1993 的**定性收敛**加强为**定量速率**。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

算法 (BB Gradient Method)

Input: A, b, x_0, x_1 (两个初值); tol

For $k = 1, 2, \ldots$ **do**

1. $g_k = Ax_k - b$; 如果 $\|g_k\| < \text{tol}$, 返回 x_k
2. $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$
3. $\alpha_k = \dfrac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}$ (BB 步长)
4. $x_{k+1} = x_k - \alpha_k^{-1} g_k$

核心观察 (论文 §2):

由 $g_k = Ax_k - b$, 梯度递推:

$$g_{k+1} = \frac{1}{\alpha_k}(\alpha_k I - A) g_k$$

Rayleigh 商性质: $\alpha_k = \dfrac{g_{k-1}^T A g_{k-1}}{g_{k-1}^T g_{k-1}}$, 因而 $\lambda_1 \le \alpha_k \le \lambda_n$ 对所有 k。

谱分解视角: 设 $A = Q\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n) Q^T$, 把梯度按特征向量展开:

$$g_k = \sum_{i=1}^{n} d_i^k v_i, \qquad d_i^{k+1} = \left(1 - \frac{\lambda_i}{\alpha_k}\right) d_i^k.$$

关键: 每个特征向量方向的分量 d_i^k 满足独立的标量递推, 且乘子 $|1 - \lambda_i/\alpha_k|$ 由 $\alpha_k \in [\lambda_1, \lambda_n]$ 控制。

引理 2.1 (单步 δ-放大)

对所有 $i = 1, \ldots, n$ 与 $k \geq 1$:

$$(d_i^{k+1})^2 \leq \delta^2 (d_i^k)^2, \qquad \|g_{k+1}\|^2 \leq \delta \, \|g_k\|^2, \qquad \delta = \frac{\lambda_n}{\lambda_1} - 1 = \kappa - 1.$$

证明的关键不等式: $|1 - \lambda_i/\alpha_k| \leq \max(1 - \lambda_1/\lambda_n,\ \lambda_n/\lambda_1 - 1) = \delta.$

- 若 $\lambda_i \geq \alpha_k$:$|1 - \lambda_i/\alpha_k| \leq 1 \leq \delta$ ($\delta \geq 1$ 时);
- 若 $\lambda_i < \alpha_k$:$\lambda_i/\alpha_k \leq \lambda_n/\lambda_1$, $\lambda_i/\alpha_k - 1 \geq 1 - \lambda_n/\lambda_1$.

两种局面:

- $\delta < 1$ 时 ($\kappa < 2$): $\|g_{k+1}\| \leq \sqrt{\delta}\, \|g_k\|$ 已是 Q-线性 $\Rightarrow$ **重现 Molina-Raydan 1996 的结果**。
- $\delta \geq 1$ 时 ($\kappa \geq 2$): 引理 2.1 **只给出非膨胀界**, 不直接给出收缩。**真正的难点在此**。

$\delta \geq 1$ **时怎么办?** 关键策略 (Dai-Liao 的核心想法):**逐个特征值方向归纳地证明分量衰减**。

引理 2.4 (统一减半步数)

存在仅依赖于 λ_1, λ_n 的正整数 m, 使得 $\|g_{k+m}\| \leq \frac{1}{2}\|g_k\|, \forall k \geq 1$.

归纳思路 (论文 §2 引理 2.2–2.3): 记 $D(k,l) = \sum_{i=1}^{l}(d_i^k)^2$, 对 l 归纳证明

$$D(k+j,l) \leq \varepsilon_l \|g_k\|^2, \quad \forall j \geq m_l, \quad \varepsilon_l = \tfrac{1}{4}(1+2\delta^4)^{l-n}, \ m_l \text{ 递归定义}.$$

为什么归纳能 work? 如果对 $i \leq l$ 已经有 d_i^k 足够小, 则

$$\alpha_{k+1} = \frac{\sum_i (d_i^k)^2\lambda_i}{\sum_i (d_i^k)^2} \approx \frac{\sum_{i>l}(d_i^k)^2\lambda_i}{\sum_{i>l}(d_i^k)^2} \geq \lambda_{l+1}$$

$\Rightarrow \alpha_{k+1}$ “被高频拽到至少 λ_{l+1}” $\Rightarrow |1-\lambda_{l+1}/\alpha_{k+1}| \leq c = 1-\lambda_1/\lambda_n < 1 \Rightarrow d_{l+1}^{k+1}$ **开始收缩!**

要点: Rayleigh 商**自适应地选取步长**, 一旦低频已经小, 步长就被推到对应高频的近 $1/\lambda$, 启动该频道的收缩。这就是 BB 法**自动谱适应**的精髓。

引理 2.2–2.3 的归纳细节

设置: $c = 1 - \lambda_1/\lambda_n \in [\frac{1}{2}, 1)$ (当 $\delta \geq 1$), $D(k,l) = \sum_{i \leq l} (d_i^k)^2$.

引理 2.2: 假设对前 l 个分量已有控制 $D(k+j, l) \leq \varepsilon_l \|g_k\|^2$ ($j \geq m_l$), 则 $\exists j_0 \in [m_l, m_l + \Delta_l + 1]$ 使

$$(d_{l+1}^{k+j_0})^2 \leq 2\varepsilon_l \|g_k\|^2, \quad \Delta_l = \left\lceil \frac{\log(2\varepsilon_l \delta^{-2(m_l+1)})}{2\log c} \right\rceil.$$

核心想法: 若 $(d_{l+1}^{k+j})^2 > 2\varepsilon_l \|g_k\|^2$ 对所有 $j \in [m_l, m_l + \Delta_l]$, 则

$$\alpha_{k+j+1} \geq \frac{\lambda_{l+1}(d_{l+1}^{k+j})^2}{\varepsilon_l \|g_k\|^2 + \sum_{i>l+1}(d_i^{k+j})^2} \geq \frac{2}{3}\lambda_{l+1}$$

$\Rightarrow |1 - \lambda_{l+1}/\alpha_{k+j+1}| \leq c < 1 \Rightarrow (d_{l+1}^{k+j+2})^2 \leq c^2 (d_{l+1}^{k+j+1})^2 \Rightarrow$ 连续 Δ_l 步收缩使 $(d_{l+1})^2$ 跌破 $2\varepsilon_l \|g_k\|^2$, 矛盾。

引理 2.3: 在 j_0 后, 一旦 $(d_{l+1}^{k+j+1})^2 > 2\varepsilon_l \|g_k\|^2$ 又发生, 立即触发 c^2-收缩, 在 j_2 步内回落到 $2\varepsilon_l$ 以下。中间最多被 δ^2 单步放大两次, 故 $(d_{l+1}^{k+j})^2 \leq 2\delta^4 \varepsilon_l \|g_k\|^2$ 对 $j \geq m_l + \Delta_l + 1$。

合成 (引理 2.4): $\varepsilon_l = \frac{1}{4}(1 + 2\delta^4)^{l-n}$, m_l 累加 Δ_j, 最终 $m = m_n$ 使 $\varepsilon_n = \frac{1}{4}$, 即 $\|g_{k+m}\|^2 \leq \frac{1}{2}\|g_k\|^2$。$m$ 与 c_2 都仅依赖 λ_1, λ_n, 与 n、初值都无关。

常数大小: 粗估 $m \sim n\Delta_l \sim n\kappa \log \kappa$, 极保守。这是论文末尾承认速率不紧的根源。

从引理 2.4 到 *R*-线性收敛: 有了 $\|g_{k+m}\| \le \frac{1}{2}\|g_k\|$, 配合引理 2.1 桥接 m 步内的非膨胀性:

定理 2.5 (Dai-Liao 2002, BB 的 *R*-线性收敛)

设 f 是严格凸二次, $\{x_k\}$ 是 BB 法生成的序列。则要么 $g_k = 0$ 对某有限 k 成立, 要么 $\{\|g_k\|\}$ ***R*-线性收敛到零**:

$$\|g_k\| \le c_1\, c_2^{k}\, \|g_1\|, \qquad c_1 = 2\delta^{m-1}, \quad c_2 = 2^{-1/m} \in (0,1).$$

推论 (论文 §3):

- **BB 法对一般非线性问题局部 *R*-线性收敛** (通过比较 x_k 与 "等价二次" $\hat{f}$ 的迭代轨道);
- **GLL 非单调线搜索的步长 α_k 最终被完全接受** (Dai 2002), 与 Raydan 1997 的数值观察吻合。

速率与条件数的关系 (论文 §3 后半): $\delta = \kappa - 1, m \approx O(\kappa \log \kappa), c_2 = 2^{-1/m} \to 1$ 当 $\kappa \to \infty$。

速率上界随 κ 增大而趋于 1 $\Rightarrow$ 解释 "BB 在病态问题上变慢", 但论文坦言此估计是否能定量改进至接近 BB 实际表现, **仍是开放问题**。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

为什么假设是必要的?

论文中的假设: $A = A^T \succ 0$ (严格凸二次)。

(i) 去掉“对称”: $y_{k-1} = As_{k-1}$, 分母 $s_{k-1}^T As_{k-1} = s_{k-1}^T \operatorname{sym}(A) s_{k-1}$ **只反映对称部分**。Rayleigh 商性质 $\alpha_k \in [\lambda_1, \lambda_n]$ 失效 $\Rightarrow$ 引理 2.1 的 δ-界塌陷, 证明从根上走不通。

(ii) 去掉“正定”: A 不定 $\Rightarrow$ f 无下界,x_k 可发散到 ∞;A 半定 $\Rightarrow$ α_k 的分母可为 0, **步长无定义**。

(iii) 去掉“二次”:

- 非线性时 $y_{k-1} \neq As_{k-1}$, Rayleigh 商身份不成立, α_k 可超出 $[\lambda_1, \lambda_n]$;
- **Burdakov-Dai-Huang 2019**: strongly convex 非二次函数上, 纯 BB 可陷入 **4 周期循环** $\Rightarrow$ 必须配合非单调线搜索。

论文的处理 (§3): Taylor 展开为等价二次 $\hat{f}(x) = f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(x^*)(x - x^*)$, 比较 $\{x_k\}$ 与 $\{\hat{x}_k\}$, 配合 GLL 全接受 (Dai 2002) $\Rightarrow$ **二次的 *R*-线性局部传递到一般非线性**。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

数值实验: 用 LMSD R-线性包络间接验证

论文本身无数值实验(纯理论)。Curtis-Guo 2018 把 Dai-Liao 2002 直接推广到 LMSD, 故 **LMSD 的 R-线性包络实验是本文结论在更大方法族上的间接经验验证**。下述实验来自毕设 §7。

实验设置 (毕设 §7.1):

- $H = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, 特征值在 $[1, \kappa]$ 上对数等距;
- 初值 $g_0^{(i)} = \sin(0.37i) + 0.3\cos(1.17i) + 0.5\sin(0.11i) + 1$;
- 终止 $\|g_k\| / \|g_0\| \le 10^{-8}$;$p \in \{4, 8\}$,$\omega(t) \in \{1/t, 1, t\}$。

三组实验:

(E1) 多配置下绘 $\log\|g_k\|$ 曲线 $\Rightarrow$ 可视化**上包络**;
(E2) 与 BB1/BB2、CG 对比 $\Rightarrow$ 实证 LMSD 速度优势;
(E3) 最小二乘拟合 $\log\|g_k\| \sim k\,\log\hat{\rho}$, 与 c_2、$1 - 1/\kappa$ 对比 (见下一页)。

(E2) 迭代数 (毕设表 7.2):

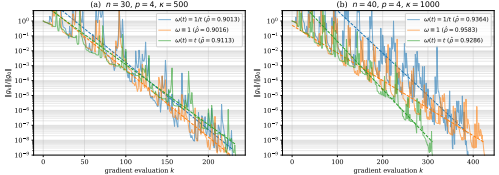
方法	对角 n=30 κ=500	稠密 n=20 κ=100
CG (Krylov 基线)	45	21
BB1 / BB2	269/272	119/111
LMSD ($\omega\equiv1$, p=4)	**187**	**88**
harm-LMSD (ω=t, p=4)	222	85
LMSD (p=10)	148	—

LMSD (p=4) 比 BB 快 $\sim 30\%$; p=10 再快 $\sim 20\%$。

(E1) 现象 (毕设 §7.7): 半对数轴下,$\|g_k\|$ **剧烈震荡**, 但**上包络近似为直线**。这正是 *R*-**线性 (而非 *Q*-线性)** 的图像。

(E3) 经验速率拟合 (毕设表 7.6):

配置 (p=4, $\omega\equiv$1)	$\hat{\rho}$	$1-1/\kappa$
n=30, κ=100	0.817	0.990
n=30, κ=500	0.902	0.998
n=30, κ=1000	0.941	0.999
n=40, κ=1000	0.956	0.999



毕设图 7.4 ($n = 30, \kappa = 500, p = 4$)。半对数下原始轨迹震荡, 上包络近似直线。

108 **次拟合:** $\hat{\rho}$ 中位数 ≈ 0.87, 全部 $\hat{\rho} \in [0.78, 0.96]$。

解读: 论文 *R*-线性证明对应的图象就是 **虚线包络**。**震荡 $\neq$ 不收敛, 而是 BB/LMSD 自适应谱信息的副产品**。**关键发现:** 所有 $\hat{\rho}$ 都远小于 Dai-Liao c_2 与 Li-Sun $1-1/\kappa$, 与下一节 "紧速率 $(\kappa-1)/(\kappa+1)$" 同阶。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

亮点与潜在缺陷

亮点:

- **证明极简**: 仅靠 Rayleigh 商区间 $[\lambda_1, \lambda_n]$ 与基本归纳法, 无高阶工具;
- **结论强**: 严格凸二次 + 无附加假设 $\Rightarrow$ R-线性, 移除 Molina-Raydan $\kappa < 2$ 限制;
- **开启 LMSD 系列**:Fletcher 2012 / Curtis-Guo 2018 / Ferrandi-Hochstenbach 2025 全部建立在本文之上;
- **为非线性局部分析铺路**:Taylor 比较 + GLL 全接受是后续 BB 局部 R-线性的范式。

潜在缺陷与公开问题:

- **速率不紧**:Dai-Liao 给 $c_2 = 2^{-1/m}$,$m \sim O(\kappa \log \kappa)$。论文末尾承认:*BB* 优于 *SD* 的理论证据, 仍是开放问题。
- **证明缺乏几何/动力学解读**: 四个引理的 ε_l、m_l 都是**递推定义**的**巨大常数**, 难以推广到 LMSD 等"块映射"方法;
- **依赖二次精确性**:Rayleigh 商身份在非二次情形不成立, 局部 R-线性需 GLL 全接受作**额外桥梁**。

基于以上, 我提出两个改进方向 (毕设独立完成的工作):

(A) 用**闭锥连续正齐次映射**给一个 **1 页证明**, 可推广到 LMSD 整个家族;

(B) 用**动力系统 + 遍历优化**解决"BB 最紧渐近速率"公开问题, 答案是 $(\kappa - 1)/(\kappa + 1)$。

核心观察: 谱梯度方法的步长函数对整体缩放是**零次齐次**的 ($\beta(tu) = \beta(u), t > 0$) $\Rightarrow$ 块映射 $T : \Omega \to \Omega$ **连续、一次正齐次**; 且许多谱方法已有**逐点收敛性** (Raydan 1993)。

引理 (毕设引理 2.2 / 统一有限步收缩原理)

设 X 有限维赋范, $\Omega \subset X$ 闭锥, $T : \Omega \to \Omega$ 连续、**一次正齐次**。若 $T^k(x) \to 0$ 对每个 $x \in \Omega$ 成立, 则**统一存在** $C > 0, \rho \in (0, 1)$ 使

$$\left\| T^k(x) \right\| \le C\rho^k \|x\|, \quad \forall x \in \Omega, \ \forall k \ge 0.$$

这个引理为何关键? 它把 "逐点收敛 $T^k(x) \to 0$" **自动**升级为 "**统一** R-线性 + 显式速率", 且**不依赖具体算法、不需逐项归纳**; 同一引理同时覆盖 **BB、LMSD、harmonic-LMSD、ω-加权 Ritz-sweep 全家族**。

证明的核心想法: 用**紧覆盖技巧**把 "逐点收敛" 升级为**统一减半步数** N_0, 再用正齐次性把 "减半" 传到整个 Ω。

证明分五步。设 $S_\Omega := \Omega \cap \{\|x\| = 1\}$ (**有限维 + 闭 $\Rightarrow$ S_Ω 紧**)。

(i) 紧致 + 局部减半. $\forall x \in S_\Omega$, 由 $T^k(x) \to 0$ 取 $n_x \in \mathbb{N}$ 使 $\|T^{n_x}(x)\| < \frac{1}{2}$。由 T^{n_x} 连续, x 在 S_Ω 中有开邻域 U_x 使 $\|T^{n_x}(y)\| < \frac{1}{2}$ 对所有 $y \in U_x$ 成立。

(ii) 有限覆盖 $\to$ 统一 N_0. $\{U_x\}_{x\in S_\Omega}$ 是 S_Ω 的开覆盖, 由紧性取**有限子覆盖** $U_{x_1}, \dots, U_{x_r}$, 令 $N_0 := \max_{1\le i\le r} n_{x_i}$. 对每个 $y \in S_\Omega, y$ 落在某 U_{x_i} 中, 故**存在**减半步数 $m(y) := n_{x_i} \in \{1, \dots, N_0\}$ 使 $\left\|T^{m(y)}(y)\right\| < \frac{1}{2}$.

(iii) 统一 K 界. 由 $T^j\,(j \le N_0)$ 连续 + S_Ω 紧, 定义

$$K := \max\left\{1,\ \sup_{y\in S_\Omega} \|T(y)\|,\ \dots,\ \sup_{y\in S_\Omega} \left\|T^{N_0}(y)\right\|\right\} < \infty.$$

由一次正齐次性 $T^j(tz) = tT^j(z)$ 扩展到整个 Ω: $\left\|T^j(z)\right\| \le K\|z\|, \forall z \in \Omega, 0 \le j \le N_0$.

(iv) 递归减半. 固定 $x \neq 0$, 递归构造 $z_0 := x$, $z_q := T^{m_q}(z_{q-1})$, 其中 $m_q := m(z_{q-1}/\|z_{q-1}\|) \in \{1, \dots, N_0\}$. 由一次正齐次性 + 步骤 (ii):

$$\|z_q\| = \|z_{q-1}\| \cdot \left\| T^{m_q}(z_{q-1}/\|z_{q-1}\|) \right\| \leq \tfrac{1}{2} \|z_{q-1}\| \implies \|z_q\| \leq 2^{-q} \|x\| .$$

(v) 桥接到任意 k. 记 $s_q := \sum_{j \leq q} m_j \leq qN_0$。对任意 $k \geq 1$, 取 p 使 $s_p \leq k < s_{p+1}$。由步骤 (iii) 及 $k - s_p < N_0$:

$$\left\| T^k(x) \right\| = \left\| T^{k-s_p}(z_p) \right\| \leq K \|z_p\| \leq K \cdot 2^{-p} \|x\| .$$

由 $k < s_{p+1} \leq (p+1)N_0 \Rightarrow p > k/N_0 - 1$, 代回得最终估计:

$$\boxed{\left\| T^k(x) \right\| \leq 2K \left(2^{-1/N_0}\right)^k \|x\|, \quad C = 2K,\ \rho = 2^{-1/N_0} \in (0,1).} \qquad \square$$

回顾: (i) 局部减半 + (ii) 紧覆盖统一 → (iii) 一致 K 界 → (iv) 正齐次递归 → (v) 桥接任意 k。整证只用了**紧致性 + 连续性 + 正齐次性**三件最朴素的工具。

状态空间 (两步梯度对). 定义闭锥

$$\Omega_{\mathrm{BB}} := \left\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n : \ \exists\, \alpha \in [1/\lambda_n, 1/\lambda_1],\ \ v = (I - \alpha H)u\right\}.$$

BB 块映射. $\mathcal{T}_{\mathrm{BB}}(u,v) := (v,\ (I-\beta(u)H)v),\ \beta(u) := u^T u/(u^T H u)$ 是 H 在 u 处的 **Rayleigh 商**, **零次齐次** (即 $\beta(tu) = \beta(u), t > 0$)。

验证引理的三个条件 (一一对应):

- **(i)** Ω_{BB} **是闭锥**: α 范围紧 + 极限封闭 + $0 \in \Omega_{\mathrm{BB}}$ + 关于正缩放封闭;
- **(ii)** $\mathcal{T}_{\mathrm{BB}}$ **连续、一次正齐次**: β 零次齐次 $\Rightarrow \mathcal{T}_{\mathrm{BB}}(tu,tv) = t\mathcal{T}_{\mathrm{BB}}(u,v)$, β 在 $u \neq 0$ 处连续 (原点处用 $\|v\| \le M_0 \|u\|$ 控制);
- **(iii)** $\mathcal{T}_{\mathrm{BB}}^k(Z) \to 0$: Raydan 1993 已证 $g_k \to 0 \Rightarrow Z_k = (g_k, g_{k+1}) \to (0,0)$ ——**现成结论, 无须重证**。

结论: 引理直接给出 $\|Z_k\| \le C\rho^k \|Z_1\|$, g_k 是 Z_k 的第一分量 $\Rightarrow$ ***R*-线性传递到** $\|g_k\|$。

$\Rightarrow$ **Dai-Liao 主定理是该抽象引理的直接推论**, 证明从原论文 **4 引理** + n **重归纳** + $\varepsilon_l, m_l, \Delta_l$ **巨大常数压缩到 3 行结构验证**。

改进 (A) 5/5:BB 应用的完整证明细节

Ω_{BB} **是闭锥:** 若 $(u_k, v_k) \in \Omega_{\mathrm{BB}}$, $\alpha_k \in [1/\lambda_n, 1/\lambda_1]$ 使 $v_k = (I - \alpha_k H)u_k$, $(u_k, v_k) \to (u, v)$. 区间紧 $\Rightarrow$ 子列 $\alpha_k \to \alpha \in [1/\lambda_n, 1/\lambda_1]$, 取极限 $v = (I - \alpha H)u$, 故 $(u, v) \in \Omega_{\mathrm{BB}}$。锥性显然。 □

$\mathcal{T}_{\mathrm{BB}}$ **自映射:** $(u, v) \in \Omega_{\mathrm{BB}}, u \neq 0 \Rightarrow \beta(u) = u^T u/(u^T H u) \in [1/\lambda_n, 1/\lambda_1]$ (Rayleigh 商), 故 $(v, (I - \beta(u)H)v)$ 同样满足闭锥定义中的 α 条件。 □

$\mathcal{T}_{\mathrm{BB}}$ **连续, 一次正齐次:**

- $\beta(tu) = \beta(u)$ $(t > 0) \Rightarrow \mathcal{T}_{\mathrm{BB}}(tu, tv) = t\mathcal{T}_{\mathrm{BB}}(u, v)$, 一次正齐次;
- $u \neq 0$ 处 β 连续, 故 $\mathcal{T}_{\mathrm{BB}}$ 连续;
- 原点处: $\|v\| \leq M_0 \|u\|$, $\|(I - \beta(u)H)v\| \leq M_0 \|v\| \leq M_0^2 \|u\| \to 0$, 其中 $M_0 = \max_{\alpha \in [1/\lambda_n, 1/\lambda_1]} \|I - \alpha H\| < \infty$。 □

$\mathcal{T}_{\mathrm{BB}}^k(Z) \to 0$: Raydan 1993 已证 $g_k \to 0 \Rightarrow Z_k = (g_k, g_{k+1}) \to (0, 0)$。 □

合并: 引理 (毕设 2.2) $\Rightarrow \exists\, C > 0, \rho \in (0, 1)$ 使 $\|Z_k\| \leq C\rho^k \|Z_1\|$。$g_k$ 是 Z_k 的第一分量, 故 R-线性传递。 □

与原证明的区别:

- **原证明** (Dai-Liao): 4 个引理 + n 重归纳 + 复杂常数 $\varepsilon_l, m_l, \Delta_l$;
- **闭锥证明**: 3 行验证 (Ω_{BB} 闭锥 + $\mathcal{T}_{\mathrm{BB}}$ 连续正齐次 + Raydan 1993 逐点) + 引用引理。

改进 (A) 的更广意义

推广能力 (毕设第 4 章): 把同样范式应用于 LMSD 等“块方法”：

- **LMSD (Fletcher 2012, p 步 Ritz)**: 状态为相容块端点梯度对 $Z_k = (g_{k,0}, g_{k+1,0})$, 闭锥 Ω_p 用乘子多项式 $\prod_{\ell=1}^{p}(1-\alpha_\ell t), \alpha_\ell \in [1/\lambda_n, 1/\lambda_1]$ 刻画;
- **harmonic-LMSD (Curtis-Guo 2018, $\omega(t)=t$); 整个 ω-加权 Ritz-sweep 族**: 权重 $\omega(t) > 0$ 任意。

与 Curtis-Guo 2018 的对比: 他们需**额外假设** $\left\|R_k^{-1}\right\| \le \rho \left\|g_{k,1}\right\|^{-1}$ (历史束最小奇异值控制), 称为“non-degenerate”。

- 毕设实验 §7.4 显示: 真实 BB 轨道上, $1/\mu_k := \|g_{k,0}\| / \sigma_{\min}(G_k)$ 可4 **个数量级爆破**, 假设几乎从**未满足**;
- **闭锥范式的关键改进**: 把**秩退化**吸收为闭锥的**边界**, 通过乘子多项式参数化绕过 R_k^{-1} 的不连续性。

定理 (毕设 §4.5, 无条件 R-线性)

LMSD、harmonic-LMSD 及任意 ω-加权 Ritz-sweep 在严格凸二次问题上**无条件 R-线性收敛**, 速率仅依赖于谱 $[\lambda_1, \lambda_n]$ 与历史长度 p。

设计哲学: Dai-Liao 用“分量归纳”刻画**某种特殊结构**; 闭锥范式表明**结构的本质是“放缩齐次性 + 连续闭锥”**, 自动给出 R-线性。

改进 (B):BB 的最紧渐近速率

问题陈述 (毕设 §8.3, 公开问题): 经典 BB1 + Cauchy 预热,

$$\alpha_0 = \frac{g_0^T g_0}{g_0^T H g_0}, \quad \alpha_k = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T H g_{k-1}} \ (k \geq 1).$$

求**最紧渐近速率** $\sup_{x_0} \limsup_{k\to\infty} \|g_k\|^{1/k} = ?$

既有最好结果: Li-Sun 2021 改进为 $1 - 1/\kappa$。**但实验上速率明显比这两个都小**

定理 (毕设 §8.3 定理 8.3, BB 最紧渐近速率)

对任意 $H \succ 0$ 与任意初值 x_0: $\limsup_{k\to\infty} \|g_k\|^{1/k} \leq c(H) := \dfrac{\kappa - 1}{\kappa + 1}$,

上界由**仅激活 $\lambda_{\min}, \lambda_{\max}$ 两端点且两端等能量**的对称轨道恰好达到。$\forall \varepsilon > 0$, $\exists\, C_\varepsilon(H)$ 使 $\|g_k\| \leq C_\varepsilon(H)(c(H) + \varepsilon)^k \|g_0\|$。

比较:

	速率上界	$\kappa \to \infty$ 渐进
Li-Sun 2021	$1 - 1/\kappa$	$\sim 1 - 1/\kappa$
毕设 (此处)	$\frac{\kappa-1}{\kappa+1}$	$\sim 1 - 2/\kappa$

注意:$(\kappa - 1)/(\kappa + 1)$ 恰好等于**最优固定步长梯度法的最坏速率**, 揭示 BB 与最优固定步长法的内在联系。

改进 (B) 的核心 (1/3): 从动力系统视角看 BB

步骤 1: 谱压缩. H 不同特征值 $\mu_1 < \cdots < \mu_m$, 谱投影 P_i, 块能量 $E_i^k := \|P_i g_k\|^2$:

$$\alpha_k = \frac{\sum_j E_j^{k-1}}{\sum_j \mu_j E_j^{k-1}}, \qquad E_i^{k+1} = (1-\alpha_k\mu_i)^2 E_i^k.$$

BB 在 $\mathbb{R}^n$ 的动力学完全等价于激活谱上的标量递推, 维数从 n 压到 $|I| \le m$.

步骤 2: 归一化与两步动力系统. 谱能量归一化 $w_i^k = E_i^k / \sum_j E_j^k \in \Delta_I, u(w) = \sum \mu_i w_i, \varphi_i(z) = (1-\mu_i/u(z))^2$。状态 $\xi_k = (w^k, w^{k-1}) \in \Delta_I^2$:

$$T_I(w,z) = \Big(\frac{\varphi(z)\odot w}{\langle \varphi(z), w\rangle}, w\Big), \quad r_I(w,z) = \sqrt{\langle \varphi(z), w\rangle}, \quad \|g_{k+1}\| = r_I(\xi_k)\,\|g_k\|.$$

$\Rightarrow \frac{1}{n}\log(\|g_n\| / \|g_0\|) = \frac{1}{n}\sum_{k=0}^{n-1} f_I(\xi_k)$, 渐近速率 $\equiv f_I$ 沿轨道时间平均的上极限 $\Rightarrow$ **遍历优化问题**。

步骤 3: 终止态紧化. $r_I = 0 \Leftrightarrow$ 下一步必终止 (病态分母)。加孤立点 ∂, 紧空间 $\widehat{X}_I = X_I \sqcup \{\partial\}, \widehat{T}_I$ 在 $\Sigma_I = \{r_I = 0\}$ 上送往 $\partial, \widehat{f}_I$ 在该集合上取 $-\infty$, 则 $\widehat{f}_I$ **上半连续且有上界**, 满足 Portmanteau 条件。

下页: 步骤 4 (端点恒等式 + 加权 AM-GM) 与步骤 5 (测度 $\to$ 轨道), 得到 $c(H) = (\kappa-1)/(\kappa+1)$。

步骤 4: 端点恒等式 + 加权 AM-GM. 对非空 $J \subset I$, 记 $a_J = \min J, b_J = \max J, c_J = (b_J - a_J)/(b_J + a_J)$。显式构造 $H_J(w,z)$, $\psi_J(w,z) \geq 0$:

$$\widehat{f}_I(x) = \log c_J - \psi_J(x) + \underbrace{H_J(\widehat{T}_I x) - H_J(x)}_{\text{coboundary}}$$

对任意 $\widehat{T}_I$-不变测度 ν 积分:

$$\int \widehat{f}_I \, d\nu = \log c_J - \int \psi_J \, d\nu + 0 \leq \log c_J \leq \log c_I. \qquad \text{这就是上界。}$$

步骤 5: 从测度回到轨道. 经验测度 $\mu_N = \frac{1}{N}\sum_{k=0}^{N-1} \delta_{\widehat{T}_I^k \xi}$ 的弱极限 μ 自动 $\widehat{T}_I$-不变;Portmanteau (上半连续 + 上界) $\Rightarrow$

$$\limsup_N \tfrac{1}{N} \log \|g_N\| \leq \int \widehat{f}_I \, d\mu \leq \log c_I.$$

下界 (尖锐性): 2D 对称轨道: 激活 $\{a, b\}$, 两端能量等 $\Rightarrow q_k \equiv 1, \alpha_k \equiv 2/(a+b), \|g_k\| = c(a,b)^k \|g_0\|$。$c(H)$ **上界恰被对称两端轨道实现**, 不可改进。

合成: 上下界一致 $\Rightarrow \limsup_{k\to\infty} \|g_k\|^{1/k} \leq c(H) = (\kappa - 1)/(\kappa + 1)$, 紧。

改进 (B) 的核心 (3/3): 遍历优化关键不等式

加权 AM–GM 端点不等式: 对 $0 < a < b$, $u \in [a, b]$,

$$\left(\frac{u-a}{u}\right)^{\frac{b}{a+b}} \left(\frac{b-u}{u}\right)^{\frac{a}{a+b}} \le \frac{b-a}{a+b}, \quad \text{等号} \Leftrightarrow u = (a+b)/2.$$

证明: 令 $t = (u-a)/(b-u)$, 加权 AM-GM 给出 $(a+tb)/(a+b) \ge t^{b/(a+b)}$, 整理即得。

端点恒等式 (cocycle 分解). 在 $P_J = \{w_i, z_i > 0,\ i \in J\}$ 上, 显式构造 $H_J, \psi_J \ge 0$:

$$H_J(w,z) = -\frac{b_J}{2(a_J+b_J)} \log w_{a_J} - \frac{a_J}{2(a_J+b_J)} \log w_{b_J},$$

$$\psi_J(w,z) = \log c_J - \frac{b_J}{a_J+b_J} \log \frac{u(z)-a_J}{u(z)} - \frac{a_J}{a_J+b_J} \log \frac{b_J-u(z)}{u(z)},$$

则 $\widehat{f}_I(x) = \log c_J - \psi_J(x) + H_J(\widehat{T}_I x) - H_J(x)$ (coboundary 形式)。

从 cocycle 到上界. Birkhoff 平均 + Poincaré 递归 ($H_J \circ \widehat{T}_I^{n_\ell} - H_J = o(n_\ell)$) $\Rightarrow \int \widehat{f}_I \, d\nu \le \log c_J \le \log c_I$. 等号迫使 $\psi_J \equiv 0$ + 必要支撑 $\Rightarrow$ **唯一极大遍历测度** $\delta_{x_{\text{bad}}^J}$.

2D 对称轨道的初等理解. 激活 $\{a, b\}$, $g_k = x_k v_a + y_k v_b$, $q_k = x_k^2/y_k^2$。Cauchy 预热 $\Rightarrow q_1 = 1/q_0$. 递推 $q_{k+1} = q_k/q_{k-1}^2$ 的唯一不动点 $q = 1$ (对称两端等能量), 给出 $\alpha \equiv 2/(a+b)$, $\|g_k\| = c(a,b)^k \|g_0\|$. **这就是 $c(H)$ 上界的下界轨道**。

两个范式的对照:

	Dai-Liao 2002 (本文)	改进 (B) (毕设 §8)
状态空间	梯度分量 d_i^k	归一化谱能量 $(w^k, w^{k-1}) \in \Delta_I^2$
工具	分量归纳 + Rayleigh 商	端点恒等式 + 遍历测度极大化
结论	R-线性, 速率 $c_2 = 2^{-1/m}$	紧速率 $c(H) = \frac{\kappa-1}{\kappa+1}$
紧性	不紧	紧 (有下界证据)

同源关系: 二者都属于**动力系统视角**:

- **闭锥工具 (改进 A)**: **紧覆盖**把 "逐点收敛" 升级为定性的 "R-线性存在";
- **遍历优化 (改进 B)**: **遍历测度极大化**把 "渐近速率" 定到**精确常数**。

对原论文的影响:

- Dai-Liao 2002 末尾**开放问题** —— "BB 是否优于 SD, 理论上有无明确证据?" $\Rightarrow$ **在最紧渐近速率层面,BB 上界 $\frac{\kappa-1}{\kappa+1}$ 等于最优固定步长梯度法的最坏速率**, 严格优于一般 SD;
- LMSD R-线性包络实验 (毕设 §7.7) 在 $\kappa \in \{100, 500, 1000\}$ 时拟合 $\hat{\rho} \in \{0.82, 0.90, 0.94\}$, 与 $c(H) = \{0.98, 0.996, 0.998\}$ 同阶, 远低于 $1 - 1/\kappa$,**为紧速率定理提供经验证据**。

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

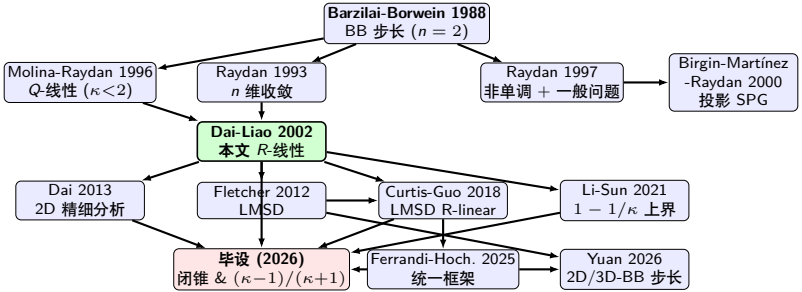
5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

参考文献与延伸: 谱梯度法的“家族树”



相关文献: **算法发展** Fletcher 2012 / Curtis-Guo 2018 / Ferrandi-Hochstenbach 2025 / Yuan 2026 (2D/3D-BB); **速率改进** Li-Sun 2021; **2D 精细分析** Dai 2013; **遍历优化** Walters 1982 / Jenkinson 2019; **非单调线搜索** GLL 1986 / Dai 2002。

- Barzilai, Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 1988.
- Raydan. On the BB choice of steplength. *IMA J. Numer. Anal.*, 1993.
- Molina, Raydan. Preconditioned BB method for PDEs. *Numer. Algorithms*, 1996.
- Raydan. BB method for the large scale unconstrained minimization. *SIAM J. Optim.*, 1997.
- **Dai, Liao. *R*-linear convergence of the BB gradient method. *IMA J. Numer. Anal.*, 2002.** (本文)
- Dai. On the nonmonotone line search. *J. Optim. Theory Appl.*, 2002.
- Dai. A new analysis of the BB gradient method. *J. Oper. Res. Soc. China*, 2013.
- Fletcher. A limited memory steepest descent method. *Math. Program.*, 2012.
- Curtis, Guo. *R*-linear convergence of limited memory steepest descent. *IMA J. Numer. Anal.*, 2018.
- Li, Sun. On a faster *R*-linear convergence rate of the BB method. arXiv:2101.00205, 2021.
- Ferrandi, Hochstenbach. Limited memory gradient methods. *Numer. Algorithms*, 2025.
- Yuan. 梯度法步长基于子空间逼近的选取. 待刊稿, 2026.
- Grippo, Lampariello, Lucidi. Nonmonotone line search for Newton's method. *SIAM J. Numer. Anal.*, 1986.
- Burdakov, Dai, Huang. Stabilized BB method. *J. Comp. Math.*, 2019.

1 基本信息

2 问题与动机

3 算法设计与核心证明思路

4 假设的必要性讨论

5 数值实验: 复现与解读

6 个人见解: 亮点、缺陷、与我的两个改进

7 参考文献与延伸阅读

8 总结

Dai-Liao 2002 的贡献:

- **建立 BB 在严格凸二次问题上无条件 R-线性收敛**, 去除 Molina-Raydan 1996 的 $\kappa < 2$ 限制;
- 把 Raydan 1993 的**定性收敛**加强为**定量速率**;
- 开启 LMSD 类方法 R-线性分析的范式 (Curtis-Guo 2018 直接建立在本文之上)。

我的两个个人见解 (来自毕设):

(A) **闭锥连续正齐次映射**给 BB 的 R-线性收敛一个 **1 页证明**, 同时统一覆盖 LMSD、ω-加权 Ritz-sweep 整个家族;

(B) **动力系统 + 遍历优化**解决 "BB 最紧渐近速率" 公开问题:$(\kappa-1)/(\kappa+1)$, 严格改进 Li-Sun 2021。

方法论启示: 在分析 "非单调震荡 + 自适应步长" 算法时,

- **放缩齐次性 (零次齐次步长)**是关键代数结构;
- **闭锥状态空间**是吸收退化、绕过病态分母的几何技巧;
- **遍历测度**把 "渐近速率" 概念精确化为可计算量。

把 BB 法放在动力系统的视角下看,
原本错综复杂的归纳证明, 在闭锥与遍历理论下变得透明而紧致。

感谢老师与同学的批评指正!

汇报人: 杨书泰　　数学科学学院

收敛速率

- Q-线性: $\|x_{k+1} - x^*\| / \|x_k - x^*\| \le \rho < 1$; "Q" =quotient.
- R-线性: $\|x_k - x^*\| \le C\rho^k$, 允许震荡; "R" =root; $Q \Rightarrow R$ 但反不然。

线性代数

- Rayleigh 商: $R(v; A) = \dfrac{v^T A v}{v^T v} \in [\lambda_{\min}, \lambda_{\max}]$.
- 拟牛顿条件: 割线方程 $B_{k+1} s_k = y_k$; BB 取 $B = \alpha I$ 的最小二乘解。

动力系统抽象工具

- 闭锥: $\Omega \subset X, 0 \in \Omega, x \in \Omega, t \ge 0 \Rightarrow tx \in \Omega, \Omega$ 闭。
- 一次正齐次映射: $T(tx) = tT(x)$ 对 $t \ge 0$; 放缩对称性。
- 零次齐次: $\beta(tu) = \beta(u)$, 如 Rayleigh 商。

遍历理论

- 不变测度: $\nu(T^{-1}A) = \nu(A)$; 遍历 $\Leftrightarrow$ 不变集测度 $\in \{0, 1\}$。
- 经验测度: $\mu_N = \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\widehat{T}^k x}$; 弱极限 μ 自动 $\widehat{T}$-不变。
- Birkhoff 遍历定理: 遍历 $\nu + \varphi \in L^1(\nu) \Rightarrow \frac{1}{n} \sum \varphi(T^k x) \to \int \varphi \, d\nu$, ν-a.e.
- Poincaré 递归: 紧空间 + 不变测度 $\Rightarrow$ ν-a.e. 点是递归的: $\exists n_\ell \to \infty$, $T^{n_\ell} x \to x$.
- Portmanteau: $\mu_n \Rightarrow \mu + \phi$ 上半连续有上界 $\Rightarrow \limsup_n \int \phi \, d\mu_n \le \int \phi \, d\mu$.

线搜索

- GLL 非单调 (Grippo-Lampariello-Lucidi 1986):

$$f(x_{k+1}) \le \max_{j \le M} f(x_{k-j}) - \gamma\alpha \|g_k\|^2 ,$$

M-窗口非单调下降。