

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

块映射视角下固定记忆谱梯度法的统一收敛理论及应用

杨书泰

校内导师：陈士祥 特任研究员　校外导师：袁亚湘 院士

中国科学技术大学 数学科学学院

2026 年 5 月

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

考虑无约束优化 $\min_{x\in\mathbb{R}^n} f(x)$，$x_{k+1} = x_k - \alpha_k g_k$。**BB 步长** (1988)：

$$\alpha_k^{\mathrm{BB1}} = \frac{s_k^T s_k}{s_k^T y_k}, \qquad s_k = x_{k+1} - x_k,\ y_k = g_{k+1} - g_k.$$

方法链：BB1/BB2 → Fletcher LMSD (2012) → harmonic-LMSD (Curtis–Guo 2018) → **最小二乘统一框架** (Ferrandi–Hochstenbach 2025)→ Yuan 2D/3D-BB (2026) 。

文献	算法	历史束假设	$\|g_k\|$ 速率
Dai–Liao 2002	BB (p=1)	无	R-线性
Fletcher 2012	LMSD	无（**声称** *）	仅全局收敛
Curtis–Guo 2018	LMSD	$\|R_k^{-1}\| \le \rho \|g_{k,1}\|^{-1}$	R-线性（条件性）
Li–Sun 2021	BB (p=1)	无	R-线性，速率 $1 - 1/\kappa$
本文	ω-Ritz 族 + BB	**无**	**无条件 R-线性 + BB 紧速率** $(\kappa-1)/(\kappa+1)$

*Fletcher 附录证明存在 **Ritz 值极限不连续**的漏洞，本文将予以修正。

动机：BB 只用最近**一对**步长信息（$p=1$），数值上常出现“锯齿”。
LMSD = **Limited Memory Steepest Descent**：把最近 p 步的梯度信息一次性用上，每个外循环产出 p 个步长。

算法（一个外循环 k）

1. 收集最近 p 个梯度 $G_k=[g_{k,0},\dots,g_{k,p-1}]\in\mathbb{R}^{n\times p}$；
2. QR 分解 $G_k=Q_kR_k$，定义 $p\times p$ **投影矩阵** $T_k=R_kJR_k^{-1}$（T_k 在二次情形下精确等于 H 在 $\mathcal{K}_p(H,g_{k,0})$ 上的**Galerkin 投影**）；
3. 求 T_k 的 p 个特征值 $\theta_1<\cdots<\theta_p$（**Ritz 值**）；
4. 用 $\alpha_{k,\ell}=1/\theta_\ell$ 作为接下来 p 步的步长，迭代 p 次。

要点：(i) $p=1$ 时退化为 BB；(ii) Ritz 值是 H 在 Krylov 子空间上的特征值近似 $\Rightarrow$ **自适应锁定谱信息**；(iii) Curtis–Guo 2018 把 $\omega(t)=1$ 替换 $\omega=t$ 得 **harmonic-LMSD**。

本文的 ω-加权族 $K^T\omega(H)HKv=\theta K^T\omega(H)Kv$ **一次性涵盖整族方法**。

本文六大主要贡献

贡献 1. 独立创造“**连续正齐次块映射 + 闭锥状态空间**”抽象工具，把“逐点收敛 $\Rightarrow$ R-线性收敛”写成统一定理；作为副产品给出 **BB 方法** R-**线性收敛**的一个不同于 Dai–Liao 的简洁证明。

贡献 2. 指出并**修正 Fletcher 原文**收敛证明的漏洞，借助**行列式 + Cauchy–Binet** 机制把结论推广为一族 ω-加权 Ritz-sweep 方法，证明严格凸二次模型上**逐点收敛**。

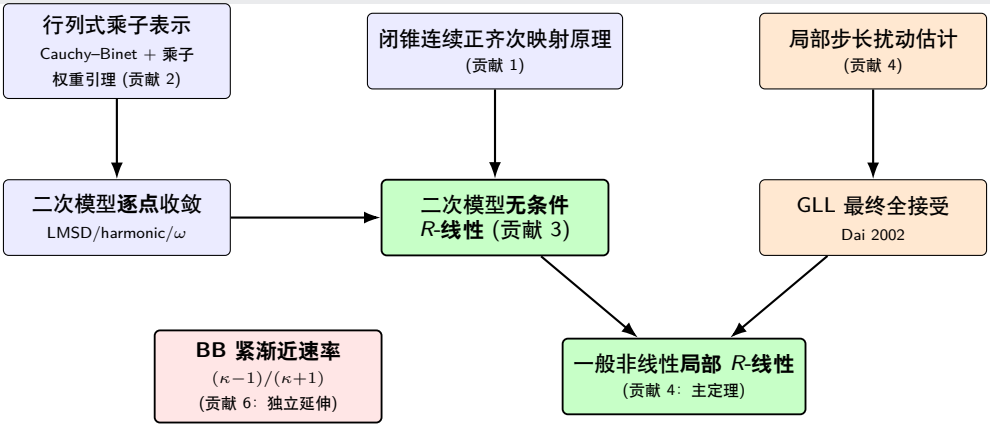
贡献 3. 结合贡献 1 与贡献 2：证明 LMSD、harmonic-LMSD 及整个 ω-族在严格凸二次模型上**无条件** R-**线性收敛**，去除 Curtis–Guo 的非退化假设。

贡献 4. 把闭锥工具配合扰动估计与 GLL 全接受传递到一般非线性，证明 pD(S)-BB 算法的**局部** R-**线性收敛**。

贡献 5. 系统的**数值实验**：有限终止、ω-族对比、Curtis–Guo 假设失效轨道、经验包络拟合等。

贡献 6. 借助**终止态紧化**与**遍历优化**工具，对经典 BB1 在严格凸二次问题上给出**紧渐近速率** $(\kappa-1)/(\kappa+1)$，定量改进 Li–Sun 2021 的 $1-1/\kappa$。

技术路线一览



两层主体 + 一条独立延伸

主体（蓝/绿）是 *R*-线性提升机制；独立延伸（红）是**同一动力系统视角下的平行实现**。

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

核心观察：谱梯度方法的步长函数对整体缩放**零次齐次** $\Rightarrow$ 块映射 $T:\Omega\to\Omega$ 是**连续、一次正齐次**的，且很多谱方法有**逐点收敛**性质。

引理（论文 Lemma 2.2）

设 $\Omega\subset X$ 闭锥，$T:\Omega\to\Omega$ 连续、一次正齐次。若 $T^k(x)\to 0,\ \forall x\in\Omega$，则存在**统一**的 $C>0$、$\rho\in(0,1)$ 使 $\left\|T^k(x)\right\|\le C\rho^k\|x\|,\ \forall x\in\Omega,\ k\ge 0$。

证明（论文 §2.2）：关键是把“逐点收敛”用紧覆盖升级为统一“减半步数 N_0”。

(i) 紧致 + 局部减半。$S_\Omega:=\Omega\cap\{\|x\|=1\}$ 紧致。$\forall x\in S_\Omega$ 由 $T^k(x)\to 0$ 取 n_x 使 $\|T^{n_x}(x)\|<1/2$；由 T^{n_x} 连续，存在 x 在 S_Ω 中开邻域 U_x 使

$$\|T^{n_x}(y)\|<1/2,\quad \forall y\in U_x.$$

(ii) 有限覆盖 $\to$ 统一 N_0。$\{U_x\}_{x\in S_\Omega}$ 取有限子覆盖 $U_{x_1},\dots,U_{x_r}$，令

$$N_0:=\max_{1\le i\le r} n_{x_i}.$$

则对每个 $y\in S_\Omega$，y 落入某 U_{x_i}，从而**存在**减半步数 $m(y):=n_{x_i}\in\{1,\dots,N_0\}$ 使 $\left\|T^{m(y)}(y)\right\|<1/2$。

续：(i)(ii) 给出 $\forall y \in S_\Omega$ 的减半步数 $m(y) \in \{1, \dots, N_0\}$。下面把“减半”推广到整个 Ω。

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

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

由正齐次性，$\left\|T^j(z)\right\| \le K\|z\|$ 对 $z \in \Omega$、$0 \le j \le N_0$ 成立。

(iv) 递归减半。固定 $x \neq 0$。递归 $z_0 := x$，$z_q := T^{m_q}(z_{q-1})$，$m_q := m(z_{q-1}/\|z_{q-1}\|) \in \{1, \dots, N_0\}$。正齐次性 + (ii) $\Rightarrow \|z_q\| \le \frac{1}{2}\|z_{q-1}\|$，从而 $\|z_q\| \le 2^{-q}\|x\|$。

(v) 桥接到任意 k。记 $s_q := \sum_{j \le q} m_j$。对 $k \ge 1$ 取 p 使 $s_p \le k < s_{p+1}$。由 (iii) 及 $k - s_p < N_0$，

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

由 $k < s_{p+1} \le (p+1)N_0 \Rightarrow p > k/N_0 - 1$，代回：

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

副产品：$\Omega_{\mathrm{BB}} = \{(u, (I - \alpha H)u) : \alpha \in [1/\lambda_n, 1/\lambda_1]\}$ + Raydan 1993 逐点收敛 $\Rightarrow$ BB 的 R-线性。

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

Fletcher 原证明的关键漏洞：试图从极限向量直接推出每个单个 Ritz 值的渐近位置，但 Krylov 维数下降时单个 Ritz 值并不连续：取 $H = \mathrm{diag}(1, 2)$，$p = 2$，$q_n = (1/n, 1)^T/\sqrt{1 + 1/n^2} \to e_2$。每个有限 n 两个 Ritz 值恒为 $\{1, 2\}$，极限处维数下降，仅得 2。

本文路径：不追踪单个 Ritz 值，转而控制**整个块乘子多项式**

$$\Psi_u(t) = \prod_{\ell=0}^{p-1}\Big(1 - \frac{t}{\theta_\ell(u)}\Big) = \frac{\det\big(K^T\omega(H)(H - tI)K\big)}{\det\big(K^T\omega(H)HK\big)}.$$

对 $\Psi_u(\lambda_r)$ 用 **Cauchy–Binet 展开成非负权重的加权平均** $\Rightarrow$ **乘子权重引理** $\Rightarrow$ LMSD、harmonic-LMSD 及整个 ω-加权 Ritz-sweep **族逐点收敛**。

所有 ω 族共享同一证明。

S 型逐点收敛：$A_r/B_r/C_r$ 集合归纳（贡献 2）

设置：记 $u_k := g_{k,0}$，标准形下 $u_{k+2}^{(i)} = \Psi_k(\lambda_i) u_{k+1}^{(i)}$。**第一分量**：$\theta_{k,\ell} \in [\lambda_1, \lambda_n] \Rightarrow u_k^{(1)} \to 0$。

归纳：设 $r > 1$ 是**最小指标使** $u_k^{(j)} \to 0\,(j < r)$ **但** $u_k^{(r)} \not\to 0$，$S_\varepsilon := \{k : |u_k^{(r)}| \geq \varepsilon\}$ **无限**。Cauchy–Binet 展开给出

$$\Psi_k(\lambda_r) = \frac{\sum_{I \in \mathcal{I}_p} \beta_I^{(S)}(u_k)\phi_I^{(r)}}{\sum_{I \in \mathcal{I}_p} \beta_I^{(S)}(u_k)}, \quad \beta_I^{(S)} = \det(V_I)^2 \prod_{i \in I} \lambda_i u_{k,i}^2,\ \phi_I^{(r)} = \prod_{i \in I}(1 - \lambda_r/\lambda_i).$$

把 $\mathcal{I}_p$ 三分（论文证明核心）：

$$A_r := \{I : r \in I\},\ B_r := \{I \subset \{r+1, \dots, n\}\},\ C_r := \{I : r \notin I,\ I \cap \{1, \dots, r-1\} \neq \emptyset\}$$

- A_r：$r \in I \Rightarrow \phi_I^{(r)}$ **含因子** $(1 - \lambda_r/\lambda_r) = 0 \Rightarrow$ **分子项为** 0。
- B_r：I **完全在** r **右侧** $\Rightarrow 0 \leq \phi_I^{(r)} \leq (1 - \lambda_r/\lambda_n)^p =: \rho_r < 1$。
- C_r：I **含某** $i(I) < r$。**令** $J(I) := (I \setminus \{i(I)\}) \cup \{r\} \in A_r$，**得** $\beta_I^{(S)}/\beta_{J(I)}^{(S)} = c_I (u_k^{(i(I))}/u_k^{(r)})^2$，**归纳假设** $+ |u_k^{(r)}| \geq \varepsilon$ $\Rightarrow$ C_r **相对权重** $\to 0$。

结论：**乘子权重引理给出** $\limsup_{k \in S_\varepsilon} |\Psi_k(\lambda_r)| \leq \rho_r < 1$，**配合粗界** $|u_{k+2}^{(r)}| \leq M_r |u_{k+1}^{(r)}|$ + Fletcher **两步引理** $\Rightarrow$ $u_k^{(r)} \to 0$，**矛盾**。 □

秩退化的处理（关键）：状态改为**连续两个块端点梯度对** $Z_k = (g_{k,0}, g_{k+1,0})$，**定义相容块端点闭锥**

$$\Omega_p = \Big\{(u,v) : \exists q(t) = \prod_{\ell=1}^{p}(1-\alpha_\ell t),\ \alpha_\ell \in \big[\tfrac{1}{\lambda_n}, \tfrac{1}{\lambda_1}\big],\ v = q(H)u\Big\}.$$

关键观察：精确秩退化 $\Longleftrightarrow$ 剩余活跃谱已 $\le p$ $\Longleftrightarrow$ 算法有限终止 $\Rightarrow$ 退化分支可纳入闭锥边界。

定理（论文 §4.3）

T_p^S 是 Ω_p **上连续、一次正齐次**自映射，$(T_p^S)^k(Z) \to 0$。结合贡献 1 引理 2.2 与贡献 2 逐点收敛：$\big\|(T_p^S)^k(Z)\big\| \le C\rho^k \|Z\|$。**对 harmonic-LMSD 与整个 ω-族成立**。

对比 Curtis–Guo 2018：他们需 $\|R_k^{-1}\| \le \rho \|g_{k,1}\|^{-1}$，本文**把秩退化吸收为闭锥内部边界**，此假设**不再必要**。

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

对称化步长（论文 §5.1）：$B_k = S_k^T Y_k$ **在非二次情形下不对称**，复特征值可能出现。取

$$C_k := \mathrm{sym}(S_k^T Y_k) = \tfrac{1}{2}(S_k^T Y_k + Y_k^T S_k), \qquad M_k := S_k^T S_k,$$

解对称广义特征值问题 $C_k z = \theta\, M_k z$，配合**秩压缩**（删 $\sigma < \varepsilon_{\mathrm{rk}}$ 的列）、**步长截断**（$[\underline{\alpha}, \overline{\alpha}]$）、**GLL 非单调线搜索**。

为何必须对称化？ 二次情形 C_k 自动对称（$Y_k = H_* S_k$）；非二次时 sym 保证特征值是实数且保证特征值**对历史数据的局部 Lipschitz 依赖**——这是后续扰动估计的关键。

GLL 必要性（论文命题 6.2）：反例 $F(x) = \varphi(x_1) + \frac{\gamma}{2}\sum_{i\geq 2} x_i^2$，预热点位于 $\mathrm{span}\{e_1\} \Rightarrow$ 历史矩阵**秩恒为** $1 \Rightarrow$ 秩压缩后**退化为 BB1**，Burdakov–Dai–Huang 2019 的四点循环反例**被原样提升**：纯实现可发散。**秩压缩 + 截断不足以替代全局化**。

全局收敛（命题 6.1，Raydan 1997）：GLL 非单调下降 + 强凸 $\Rightarrow g_k \to 0$，$x_k \to x^*$。

(i) 对称化 $\Rightarrow$ 特征值-步长对历史的局部 Lipschitz $\Rightarrow$ 步长扰动可控（引理 6.5 的基础）。
(ii) GLL 全局化 $\Rightarrow x_k \to x^*$ 打开局部分析入口，Dai 2002 的「最终全接受」让纯实现在尾轨道上还原。

假设：f 在 $x^*=0$ 附近 $C^{2,1}$，$\left\|\nabla^2 f(z)-H_*\right\|\le L\|z\|$，$H_*\succ 0$。令 $\rho:=\|E\|_\infty$。

(1) 二次逼近（引理 6.3，Taylor + L-Lip）：

$$\|\nabla f(x)-H_*x\|\le\tfrac{L}{2}\|x\|^2,\ \|(\nabla f(x)-\nabla f(y))-H_*(x-y)\|\le\tfrac{L}{2}(\|x\|+\|y\|)\|x-y\|.$$

(2) 割线差扰动：$R(E):=Y^f-Y^q=Y^f-H_*S$，$\|R\|\le c_1\rho\|S\|\Rightarrow C_f-C_q=\mathrm{sym}(S^TR)$，$\|C_f-C_q\|\le c_2\rho\|S\|^2$。

(3) 一致秩正则性（假设 6.3）：正则尾轨道上 $\sigma_{\min}(S)\ge\gamma_S\rho$，$\|S\|\le c_S\rho$；$W:=V\Lambda^{-1/2}$，$\|W\|_2\le 1/(\gamma_S\rho)$。

(4) 步长扰动（引理 6.4，Weyl 扰动）：$|\theta_i^f-\theta_i^q|\le c_2\rho$，$|\alpha_i^f-\alpha_i^q|\le\omega_1(\rho)\downarrow 0$。

(5) 块映射扰动（引理 6.5）：块内递推 $\eta_{i+1}^f=\eta_i^f-\alpha_i^f\nabla f(\eta_i^f)$ vs. $\eta_{i+1}^q=\eta_i^q-\alpha_i^qH_*\eta_i^q$，差分 δ_i 满足 $\delta_{i+1}\le M\delta_i+\gamma(\rho)\rho,\delta_0=0\Rightarrow$

$$\boxed{\|\mathcal{T}_f(E)-\mathcal{T}_q(E)\|_\infty\le\omega(\rho)\|E\|_\infty,\quad\omega(\rho)\to 0\,(\rho\downarrow 0).}$$

(6) 块映射 Lipschitz（引理 6.6）：$\left\|\mathcal{T}_f(E)-\mathcal{T}_f(\widetilde{E})\right\|_\infty\le L_{\mathcal{T}}\left\|E-\widetilde{E}\right\|_\infty$（在 $\mathcal{U}_{\mathrm{reg}}\cap\{\rho\le\delta_L\}$ 上）。

(1)(2) Hessian Lipschitz ⇒ 割线矩阵与二次模型偏差 $O(\rho)$；**(3) 秩正则**让标准化矩阵有控；**(4)** Weyl 扰动把矩阵偏差转成步长偏差；**(5)** 块内递推累积步长偏差 ⇒ **块映射偏差** $\omega(\rho)\|E\|$；**(6)** $\mathcal{T}_f$ 自身 Lipschitz。**核心结论**：真实块映射 $\mathcal{T}_f$ 是**二次块映射 $\mathcal{T}_q$ 的 $O(\rho)$ 扰动**。

贡献 4 主定理：局部 R-线性收敛

从扰动估计到 R-线性估计的传递：尾轨道一旦进入 $\mathcal{U}_{\mathrm{reg}} \cap \{\rho < \delta\}$：

- 引理 6.7 给出 $\left\|\mathcal{T}_f^N - \mathcal{T}_q^N\right\|_\infty \le \omega_N(\rho)\rho \Rightarrow \mathcal{T}_f^N$ 是 $\mathcal{T}_q^N$ 的小扰动；
- 贡献 1 紧子集版本 Prop. 2.5 + 贡献 3 二次 R-线性 $\Rightarrow \mathcal{T}_f$ 在尾轨道紧子集上保持 R-线性；
- Dai 2002 GLL 全接受（论文 §6.5）：GLL 从某时刻起**持续接受全试探步长**，与纯实现一致。

主定理（论文 Thm. 6.11）

设 f 强凸 + 正则假设 6.1–6.2。则 GLL 全局化 pD(S)-BB：

- (i) 全局收敛 $x_k \to x^*$（Raydan 1997 + GLL 强凸）；
- (ii) 尾轨道在 $\mathcal{U}_{\mathrm{reg}}$ 内**局部 R-线性**；
- (iii) GLL 从某时刻起**持续接受全试探步长**（Dai 2002）；
- (iv) GLL 全局化与纯实现尾轨道一致，继承同一 R-线性速率。

两层闭环：$\underbrace{\text{二次}R\text{-线性}}_{\text{贡献 3}} + \underbrace{\text{块映射扰动}\|\mathcal{T}_f - \mathcal{T}_q\| = o(\rho)}_{\text{贡献 4 引理链}} + \underbrace{\text{GLL 全接受}}_{\text{Dai 2002}} \Longrightarrow \text{局部 } R\text{-线性}.$

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

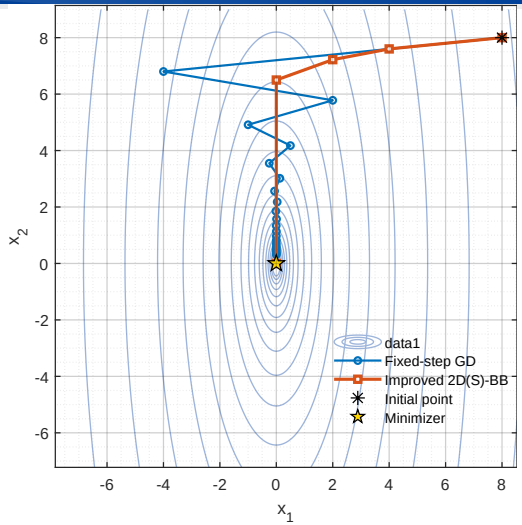
实验一：有限终止现象

二维可视化：$f(x) = \frac{1}{2}(10x_1^2 + x_2^2)$，$x_0 = (8,8)^T$。2D(S)-BB 应在 $p+n=4$ 步内终止，图中清晰可见。

高维（论文 Tab. 7.1）：

p	迭代数	理论上界 $p+n$
8	16	16
10	18	18
12	20	20

理论上界 $p+n$ 在 $\omega(t) = t^a$, $a \in \{-1, 0, 1\}$ 上一致精确取到 $\Rightarrow$ 有限终止定理是**正确**的。



论文图 7.1：2D 二次等高线与 2D(S)-BB 轨迹。

(实验二) 与 BB1/BB2/CG 对比：(A) 对角对数谱 $n{=}30, \kappa{=}500$；(B) 稠密 $n{=}20, \kappa{=}100$。

方法	(A) $n{=}30, \kappa{=}500$	(B) $n{=}20, \kappa{=}100$
CG（Krylov 最优基线）	45	21
BB1 / BB2	269 / 272	119 / 111
LMSD ($\omega \equiv 1$, $p{=}4$) / harm ($\omega{=}t$) / 逆权	187 / 222 / 183	88 / 85 / 85
LMSD ($p{=}10$)	**148**	—

观察：$p{=}4$ 时 LMSD 类比 BB 快约 30%；$p{=}10$ 再快约 20%；不同 ω 差异小（ω 族统一性印证）。

(实验三) Curtis–Guo 假设沿真实轨道失效（论文 §7.4）：$\mu_k := \sigma_{\min}(G_k)/\left\|g_{k,0}\right\|$，假设 $\Leftrightarrow \sup_k 1/\mu_k < \infty$。

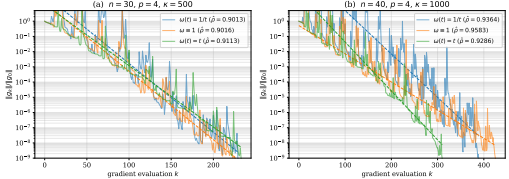
块 k	$\sigma_{\min}(G_k)$	$\left\|g_{k,0}\right\|$	$\lvert\mathcal{A}(w_k)\rvert$	$1/\mu_k$
1	3.22×10^{-4}	3.56	9	1.10×10^{4}
5	3.65×10^{-4}	0.086	9	2.35×10^{2}
10	$\mathbf{1.70\times10^{-12}}$	8.46×10^{-4}	6	$\mathbf{4.97\times10^{8}}$

$1/\mu_k$ **从** 10^4 **爆破至** 4.97×10^8 $\Rightarrow$ **贡献 3 去除该假设是必要的**。

实验四、五：R-线性包络 + 自适应 BB 变体对比

(实验四) R-线性包络拟合（论文 §7.7）：

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



半对数下原始轨迹非单调震荡，**上包络近似为直线**。

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

(实验五) 与自适应 BB 变体对比（论文 §7.8）：5 种谱实例，LMSD 类**全胜**。

谱	BB1	BB2	ABBmin	cyclic-BB	LMSD p=4	LMSD p=8	harm p=4
diag κ=100	123	120	108	125	96.6	**92.2**	94.5
diag κ=1000	397	419	320	389	273	**202**	294
n=100, κ=1000	424	432	346	407	379	**323**	369
two-cluster	52.6	50.2	38.7	47.3	40.8	44.4	**36.8**
low-cluster	110	118	89.8	99.9	65.5	**61.4**	66.6

LMSD p=8 在 4 行最优；harmonic p=4 在 two-cluster 最优；标准差极小。

对比基线：BB-Raydan、CG-PR、L-BFGS-m、pD(S)-BB-pure（无 GLL）。

方法	Trig. n=50	Trig. n=100	Ros. n=50	Convex2 n=10^3
pD(S)-BB-GLL (p=3)	2081	1044	3513	—
pD(S)-BB-GLL (p=5)	1081	858	2713	24
pD(S)-BB-GLL (p=7)	**708**	**676**	**1827**	—
pD(S)-BB-pure (p=5)	div†	div†	—	—
BB-Raydan	1461	803	2726	**8**
CG-PR	1570	2581	1431	14
L-BFGS-5	536	800	**574**	15

要点：

- Trigonometric n=100：$p=7$ 取 **676 略胜 L-BFGS-5 的** 800；
- 纯实现（无 GLL）在 Trig. **上发散** $\Longrightarrow$ **GLL 必要性的实证**；
- Rosenbrock：BB 类弱于 BFGS，但 pD(S)-BB 在 BB 类中最稳；
- Convex 2：可分良态强凸 $\Longrightarrow$ BB-Raydan 8 步即收敛，历史窗口反而劣。

† 超出 $N_{\max}=10000$ 仍未达停止准则。**Trigonometric** n=50 **经验包络** $\widehat{\rho}_{\text{peaks}}=0.9862$。

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

考虑严格凸二次 $q(x)=\frac{1}{2}x^THx-b^Tx$，$H\succ 0$。**BB1 迭代** $x_{k+1}=x_k-\alpha_k g_k$，最速下降预热 $\alpha_0=g_0^Tg_0/g_0^THg_0$，$\alpha_k=g_{k-1}^Tg_{k-1}/g_{k-1}^THg_{k-1}$ $(k\geq 1)$。
记 $c(H):=(\kappa-1)/(\kappa+1)$，$\kappa=\lambda_{\max}/\lambda_{\min}$。

定理（BB 最紧渐近速率，论文 §8.3）
对任意 $H\succ 0$、任意初值 x_0：$\limsup_{k\to\infty}\|g_k\|^{1/k}\leq c(H)$，且上界由**仅激活** $\lambda_{\min},\lambda_{\max}$ **两端点的对称轨道**恰好达到。
进一步，$\forall\varepsilon>0$ 存在 $C_\varepsilon(H)$ 使 $\|g_k\|\leq C_\varepsilon(H)(c(H)+\varepsilon)^k\|g_0\|$。
对比与意义：Dai–Liao 2002 给出 $(1-1/\kappa)^{1/n}$（粗略）、Li–Sun 2021 给出 $1-1/\kappa$；本文给出**理论最紧上界** $(\kappa-1)/(\kappa+1)$ 的定量改进，**且恰是最优固定步长梯度法的最坏速率** $\Rightarrow$ 解决 BB 方法 R-收敛最佳速率估计的**公开问题**。

梯度递推：由 $g_k = H(x_k - x^*)$，BB1 等价于

$$g_{k+1} = (I - \alpha_k H) g_k, \quad \alpha_k = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T H g_{k-1}} \text{（上一步 Rayleigh 商倒数）}.$$

关键观察 ①：α_k 依赖 $g_{k-1} \Rightarrow$ **真实状态是两步对** (g_k, g_{k-1})——这是要走“两步块映射”的根本。**谱压缩**：设 H 的不同特征值 $\mu_1 < \cdots < \mu_m$，正交谱投影 P_i。**激活谱** $I := \{i : P_i g_0 \neq 0\}$。令谱块能量 $E_i^k := \|P_i g_k\|^2$。由 $P_i g_{k+1} = (1 - \alpha_k \mu_i) P_i g_k$ **得标量递推**（论文引理 8.10）：

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

含义：BB 在 $\mathbb{R}^n$ 的动力学**完全等价于**激活谱上 $|I|$ 维谱块能量 $\{E_i^k\}$ 的标量递推。

归一化（消去整体尺度）：$w_i^k := E_i^k / \sum_j E_j^k \in \Delta_I$，$u(w) := \sum_i \mu_i w_i$，$\varphi_i(z) := (1 - \mu_i / u(z))^2$；$\alpha_k = 1/u(w^{k-1})$。代入并归一化：

$$w_i^{k+1} = \frac{\varphi_i(w^{k-1})\, w_i^k}{\langle \varphi(w^{k-1}), w^k \rangle}.$$

关键观察 ②：w^{k+1} 依赖 w^k **和** w^{k-1}——**自然要用两步对作状态**。

两步块映射 $\widehat{T}_I$、率函数 r_I 与终止态紧化（论文 §8.3）

两步块状态：$\xi_k := (w^k, w^{k-1}) \in X_I := \Delta_I \times \Delta_I$，$\xi_0 = (w^0, w^0)$。**块映射与率函数的显式公式**：

$$\boxed{T_I(w,z) := \left(\frac{\varphi(z)\odot w}{\langle \varphi(z), w\rangle},\ w \right), \qquad r_I(w,z) := \sqrt{\langle \varphi(z), w\rangle}.}$$

加性公式（$f_I := \log r_I$）：未终止时 $\xi_{k+1} = T_I(\xi_k)$、$\|g_{k+1}\| = r_I(\xi_k)\,\|g_k\| \Rightarrow$

$$\frac{1}{n}\log\frac{\|g_n\|}{\|g_0\|} = \frac{1}{n}\sum_{k=0}^{n-1} f_I(\xi_k).$$

核心转化：BB 渐近速率 **完全等价于** f_I 沿轨道时间平均的上极限——**将渐进速率紧上界转化为遍历优化问题**。

终止态紧化：$r_I = 0 \Leftrightarrow$ 下一步终止。引入孤立点 ∂，紧空间 $\widehat{X}_I := X_I \sqcup \{\partial\}$；$\widehat{T}_I$ 在 $\Sigma_I := \{r_I = 0\}$ 上送往 ∂；$\widehat{f}_I$ 在 $\Sigma_I \cup \{\partial\}$ 取 $-\infty$。

结果：$\widehat{f}_I$ **上半连续且有上界**（$\widehat{f}_I \le 0$），$\widehat{T}_I : \widehat{X}_I \to \widehat{X}_I$ 良定但**不连续**（仅在 $D(\widehat{T}_I) \subseteq \Sigma_I \cup \{\partial\}$ 上跳跃，且与 $\widehat{f}_I = -\infty$ 集重合）。

端点恒等式：H_J 与 ψ_J 的显式构造（论文 §8.3）

动机：要从加性公式 $\frac{1}{n}\sum f_l(\xi_k)$ 得到**对每个不变测度 ν 的上界** $\int \widehat{f}_l \,\mathrm{d}\nu \le \log c_I$，需要把 f_l 表达为"**临界斜率 + 可消去项 + 非负缺陷**"的形式。

记号：对每个非空 $J \subset I$，$a_J := \min J$，$b_J := \max J$，$c_J := \frac{b_J - a_J}{b_J + a_J}$。

注意 $c_J \le c_I = \frac{\kappa-1}{\kappa+1}$（$|J| = |I|$ 时取等）。

引理（端点恒等式，论文 §8.3.5）

显式构造

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

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

（其中 $u(z) = \sum_{i\in I} \mu_i z_i$）。则 $\psi_J \ge 0$（等号 $\Leftrightarrow u(z) = \frac{a_J+b_J}{2}$，由加权 AM–GM）且

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

对 ν 积分：$\int \widehat{f}_l \,\mathrm{d}\nu = \log c_J - \int \psi_J \,\mathrm{d}\nu + 0 \le \log c_J \le \log c_I$。**这就是上界。**

记 $\widehat{\beta}_I = \sup\{\int \widehat{f}_I \mathrm{d}\nu : \nu$ 为 $\widehat{T}_I$-不变 Borel 概率测度$\}$ ——**一个遍历优化问题**。

定理（最优常数，论文 §8.3.5）

设 $|I| \geq 2$。则 $\widehat{\beta}_I = \log c_I$，且**唯一极大遍历测度**是对称两端 Dirac $\delta_{x^I_{\mathrm{bad}}}$，

$x^I_{\mathrm{bad}} = (\frac{1}{2}e_{a_I} + \frac{1}{2}e_{b_I}, \frac{1}{2}e_{a_I} + \frac{1}{2}e_{b_I})$。

证明概要（Birkhoff + Poincaré + Portmanteau）：沿端点恒等式求和取 Birkhoff 平均，Poincaré 递归使边界项 $H_J(\widehat{T}_I^{n_\ell}x) - H_J(x) = o(n_\ell)$，得 $\int \widehat{f}_I \mathrm{d}\nu = \log c_J - \int \psi_J \mathrm{d}\nu \leq \log c_J \leq \log c_I$。等号迫使 $\psi_J \equiv 0 \Rightarrow$ 对称两端 Dirac。

从测度回到轨道：经验测度 $\mu_{N_j} \Rightarrow \mu$ 的弱极限 + 经验测度不变性引理 $\Rightarrow \mu$ 是 $\widehat{T}_I$-不变 $\Rightarrow$ Portmanteau 给出 $\limsup \frac{1}{N}\log\|g_N\| \leq \int \widehat{f}_I \mathrm{d}\mu \leq \log c_I$。

与第 2 章关系：闭锥用**紧覆盖**把逐点 $\to$ 速率；本节用**遍历测度极大化**把速率定到**精确常数**。同源，后者更定量。

1 研究背景与主要贡献

2 闭锥连续正齐次块映射工具

3 ω 族在二次模型上的应用

4 一般非线性的局部 R-线性收敛

5 数值实验

6 经典 BB 方法的紧渐近速率

7 总结与展望

两类提升范式：
(i) 全局——闭锥 Ω + 连续正齐次 T + 逐点收敛 $\Rightarrow$ **全局 R-线性**（覆盖 BB、LMSD、harmonic-LMSD、cyclic-BB、ω-族）；
(ii) 局部——紧集尾轨道 + 扰动估计 + GLL 全接受 $\Rightarrow$ **局部 R-线性**（一般非线性）。

动力系统视角的两个角度：定性的**速率存在**（闭锥工具，第 2 章）与定量的**精确速率**（终止态紧化 + 遍历优化，第 8 章）；$(\kappa-1)/(\kappa+1)$ 是后者的具体证据。

后续可推进方向：(i) 工程实现（删列、重启、变 p）的统一齐次块映射形式化；(ii) ω-族的局部 R-线性；(iii) 把终止态紧化 + 端点恒等式范式推广到 BB2、cyclic-BB、ω-族；(iv) 紧速率定理是否可去掉 ε。

把谱梯度法整个家族放进同一证明范式（齐次块映射），
并通过遍历理论给出 BB 的最优渐进速率——这是动力系统视角的两个成果。

1. *A determinantal multiplier framework for limited-memory steepest descent and weighted Ritz-sweep gradient methods.*

内容：行列式比 + Cauchy–Binet 展开 ⇒ LMSD/harm-LMSD/ω-族逐点收敛；修正 Fletcher 2012 漏洞。

对应章节：第 3 章 + §4.1–§4.4

目标期刊：*IMA J. Numer. Anal.*

2. *A closed-cone homogeneous mapping framework and unconditional R-linear convergence of fixed-memory spectral gradient methods.*

内容：闭锥连续正齐次自映射 + 逐点收敛 ⇒ 无条件 R-线性；BB 作 $p = 1$ 实例，重证 Dai–Liao 2002。

对应章节：第 2 章 + §4.5–§4.7

目标期刊：*Math. Program.*

3. *Symmetrized fixed-memory spectral gradient methods with rank compression and GLL globalization.*

内容：对称化步长 + 秩压缩 + GLL 检测；一般非线性局部 R-线性主定理。

对应章节：第 5–6 章 + §7 第二部分

目标期刊：*SIAM J. Optim.*

4. *A termination-state framework for the classical Barzilai–Borwein method: sharp asymptotic rate on strictly convex quadratics.*

内容：终止状态紧化 + 端点 cocycle 不等式 + 遍历优化；BB 紧收敛率 $(\kappa-1)/(\kappa+1)$，改进 Li–Sun 2021 的 $1-1/\kappa$。

对应章节：第 8 章

目标期刊：*Found. Comput. Math.*

综述：***A unified spectral state-space view of limited-memory steepest descent and the Barzilai–Borwein method.***

内容：提炼三条可迁移设计原则——加权端点不等式、闭锥状态空间、齐次提升。

请各位评审老师批评指正！

答辩人：杨书泰

加权端点不等式（论文 §8.3 引理）：对 $0 < a < b$ 与 $u \in [a, b]$，

$$\left(\frac{u-a}{u}\right)^{\frac{b}{a+b}} \left(\frac{b-u}{u}\right)^{\frac{a}{a+b}} \le \frac{b-a}{a+b} = c(a,b),$$

等号当且仅当 $u = (a+b)/2$。**证明**：令 $t = (u-a)/(b-u)$，对 $\{1, t\}$ 取权重 $(a/(a+b), b/(a+b))$ 的加权 AM–GM 即得。

端点恒等式中 H_J, ψ_J 的形式：

$$H_J(w,z) = -\tfrac{b_J}{2(a_J+b_J)} \log w_{a_J} - \tfrac{a_J}{2(a_J+b_J)} \log w_{b_J},$$
$$\psi_J(w,z) = \log c_J - \tfrac{b_J}{a_J+b_J} \log \tfrac{u(z)-a_J}{u(z)} - \tfrac{a_J}{a_J+b_J} \log \tfrac{b_J-u(z)}{u(z)}.$$

在 $P_J = \{(w,z) : w_i, z_i > 0, \forall i \in J\}$ 上：直接代入 T_I 的显式形式 $w_a^+ = \frac{(u-a)^2}{u^2 r_I^2} w_a$，$w_b^+ = \frac{(b-u)^2}{u^2 r_I^2} w_b$，计算 $H_J(\widehat{T}_I x) - H_J(x)$ 即得 $\widehat{f}_I = \log c_J - \psi_J + H_J \circ \widehat{T}_I - H_J$。

周期轨道初等证明（论文 §8.3）：对精确 p-周期轨道 $\xi^{(0)}, \dots, \xi^{(p-1)}$，沿周期乘起 $|\prod(1 - \mu_i/\nu_\ell)| = \rho^p \Rightarrow$ 取端点 $\mu_\pm$ 并对 F, G 用加权几何均值 $+$ 加权端点不等式 $\Rightarrow \rho \le c_I$。等号迫使 $\xi^{(j)} \equiv x^J_{\mathrm{bad}}$。这是最优常数定理的初等特例。

设激活谱 $\{a,b\}$，$0<a<b$。把 g_k 写成 $g_k = x_k v_a + y_k v_b$，定义 $q_k = x_k^2/y_k^2$。

命题（论文 §8.3）：最速下降预热使

$$q_{k+1} = q_k/q_{k-1}^2 \ (k \geq 1), \quad q_1 = 1/q_0.$$

对称轨道（$q_0 = 1$）：则 $q_k \equiv 1$，$\alpha_k \equiv 2/(a+b)$，$g_k = c(a,b)^k(x_0 v_a + (-1)^k y_0 v_b)$，$\|g_k\| = c(a,b)^k \|g_0\|$，$c(a,b) = (b-a)/(b+a)$。

这给出最优常数的下界（论文 §8.3 推论）：把 $H \succ 0$ 的初值仅激活在 $\lambda_{\min}, \lambda_{\max}$ 两端点且两端等能量 $\Rightarrow \|g_k\| = c(H)^k \|g_0\|$。

不稳定性：令 $\nu_k = \log q_k$，递推变成 $\nu_{k+1} = \nu_k - 2\nu_{k-1}$，特征方程 $r^2 - r + 2 = 0$ 的根模为 $\sqrt{2} > 1 \Rightarrow$ 除 $q_0 = 1$ 外 $|\nu_k|$ 以 $\sqrt{2}^k$ 量级振荡发散。Dai 2013 据此证明至多三步内必出现一次超线性步。

结论：$x_{\mathrm{bad}}^\star$ 是 $\widehat{T}_I$ 在射影意义下**唯一**不动点，且**动力学上不稳定**——与最优常数定理（测度层面唯一性）吻合。

对一般非线性问题，归一化轨道未必停留在整个闭锥，但可能停留在某紧子集中。

命题（论文 Prop. 2.5）

设 Ω 为开锥，$T:\Omega\cup\{0\}\to\Omega\cup\{0\}$ 连续正齐次，$K\subset\Omega\cap\mathbb{S}$ 紧。若

- (i) $z\in K$，$T(z)\neq 0\Longrightarrow T(z)/\|T(z)\|\in K$；
- (ii) 对每个 $z\in K$ 都有 $T^k(z)\to 0$，

则在锥 $\mathrm{cone}(K)$ 上有 $\left\|T^k(x)\right\|\le C_K\rho_K^k\|x\|$.

两条工具的分工：

- Lemma 2.3（闭锥版本）$\to$ 严格凸二次精确模型（贡献 3）；
- Prop. 2.5（紧子集版本）$\to$ 一般非线性局部分析（贡献 4）。

第 8 章紧速率证明与第 2 章工具的关系：二者同属**动力系统视角**下的不同实现。

- 第 2 章在闭锥上用**紧覆盖**把“逐点收敛”自动升级为“统一有限步收缩”；
- 第 8 章在终止态紧空间上用**端点恒等式 + 遍历测度极大化**把最坏渐近速率确定到精确常数 $c(H)$。

术语	定义
R-线性收敛	存在 $C>0,\rho\in(0,1)$ 使 $\|x_k-x^*\|\le C\rho^k$；同等价于 $\limsup \|x_k-x^*\|^{1/k}\le\rho<1$。*R* 表示 *root*（开根号）。
Krylov 子空间 & Ritz 值	$\mathcal{K}_p(H,g)=\operatorname{span}\{g,Hg,\dots,H^{p-1}g\}$。$H$ 在 $\mathcal{K}_p$ 上的 *Galerkin* 投影的 p 个特征值即 **Ritz 值**，是 H 谱的最优有限维近似。
harmonic Ritz	用 $\omega(t)=t$ 加权的变形：$K^TH^2Kv=\theta K^THKv$。比标准 Ritz 更能锁定 H 的小特征值。
ω-加权 Ritz	一般正权重 $\omega(t)>0$：广义特征问题 $K^T\omega(H)HKv=\theta K^T\omega(H)Kv$。$\omega\equiv1$ 得标准 LMSD；$\omega(t)=t$ 得 harmonic-LMSD。
GLL 非单调线搜索	Grippo–Lampariello–Lucidi 1986：接受步长当且仅当 $f(x_{k+1})\le\max_{0\le j\le M}f(x_{k-j})-\gamma\alpha\|g_k\|^2$，即与最近 $M+1$ 步的最大值比，而非单步严格下降。
闭锥 / 一次正齐次	$\Omega\subset X$ 闭锥：$x\in\Omega,t\ge0\Rightarrow tx\in\Omega$。$T:\Omega\to\Omega$ 一次正齐次：$T(tx)=tT(x)$ 对 $t\ge0$ 成立。是放缩对称性的代数形式。
Birkhoff 遍历定理	保测系统 $(X,T,\mu)+f\in L^1(\mu)$：μ-a.e. 有 $\lim_n\frac1n\sum_{k=0}^{n-1}f(T^kx)=\mathbb{E}[f\|\mathcal{I}]$。轨道平均 $\to$ 不变 σ-代数的条件期望。
Poincaré + Portmanteau	*Poincaré* 回复：保测系统中 μ-a.e. 点无限次回到任一正测集；*Portmanteau*：$\mu_n\rightharpoonup\mu$ 的若干等价刻画（USC 上极限不等式、闭集上半连续等）。

- 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.
- Dai, Liao. *R*-linear convergence of the BB method. *IMA J. Numer. Anal.*, 2002.
- Dai. On the nonmonotone line search. *J. Optim. Theory Appl.*, 2002.
- Dai. A New analyses 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 LMSD. *IMA J. Numer. Anal.*, 2018.
- Li, Sun. On a faster *R*-linear convergence rate of the BB method. arXiv:2101.00205, 2021.
- Huang, Dai, Liu, Zhang. On the asymptotic convergence and acceleration of gradient methods. *J. Sci. Comput.*, 2022.
- Ferrandi, Hochstenbach. Limited memory gradient methods. *Numer. Algorithms*, 2025.
- Walters. *An Introduction to Ergodic Theory*. Springer, 1982.
- Jenkinson. Ergodic optimization in dynamical systems. *Ergodic Theory Dynam. Systems*, 2019.
- 袁亚湘. 梯度法步长基于子空间逼近的选取. 待刊稿, 2026.

1. *A determinantal multiplier framework for limited-memory steepest descent and weighted Ritz-sweep gradient methods.*

内容：行列式比 + Cauchy–Binet 展开 ⇒ LMSD/harm-LMSD/ω-族逐点收敛；修正 Fletcher 2012 漏洞。

对应章节：第 3 章 + §4.1–§4.4

目标期刊：*IMA J. Numer. Anal.*

3. *Symmetrized fixed-memory spectral gradient methods with rank compression and GLL globalization.*

内容：对称化步长 + 秩压缩 + GLL 检测；一般非线性局部 R-线性主定理。

对应章节：第 5–6 章 + §7 第二部分

目标期刊：*SIAM J. Optim.*

2. *A closed-cone homogeneous mapping framework and unconditional R-linear convergence of fixed-memory spectral gradient methods.*

内容：闭锥连续正齐次自映射 + 逐点收敛 ⇒ 无条件 R-线性；BB 作 $p = 1$ 实例，重证 Dai–Liao 2002。

对应章节：第 2 章 + §4.5–§4.7

目标期刊：*Math. Program.*

4. *A termination-state framework for the classical Barzilai–Borwein method: sharp asymptotic rate on strictly convex quadratics.*

内容：终止状态紧化 + 端点 cocycle 不等式 + 遍历优化；BB 紧收敛率 $(\kappa-1)/(\kappa+1)$，改进 Li–Sun 2021 的 $1-1/\kappa$。

对应章节：第 8 章

目标期刊：*Found. Comput. Math.*

综述：***A unified spectral state-space view of limited-memory steepest descent and the Barzilai–Borwein method.***

内容：提炼三条可迁移设计原则——加权端点不等式、闭锥状态空间、齐次提升。

编号	论文主要内容	目标期刊
1	修正 Fletcher 2012 + LMSD/ω-族逐点收敛	*IMA J. Numer. Anal.*
2	闭锥抽象工具 + 无条件 *R*-线性	*Math. Program.*
3	对称化 *p*D-BB + GLL 全局化 + 非线性局部 *R*-线性	*SIAM J. Optim.*
4	BB 尖锐渐近率 $(\kappa-1)/(\kappa+1)$	*Found. Comput. Math.*
综述	谱状态空间统一观点 + 一些 open problems	--