谱梯度法的离散动力系统分析

收敛性、最坏收敛因子与 **Lyapunov** 函数

杨书泰

“实用最优化算法”课程读书报告　2026 年 10 月

摘要

Barzilai–Borwein（BB）方法利用相邻两点的割线信息确定步长, 不做精确线搜索, 迭代也不单调，这些都与最速下降法明显不同。本报告精读 Barzilai 与 Borwein 1988 年的论文 *Two-point step size gradient methods*：先详细推导文中的二维离散递推，再把这篇论文放到最速下降法等方法的谱分析、BB 方法的收敛理论和有限记忆最速下降法（LMSD）的研究脉络中加以考察。连续时间模型善于揭示加速机制和能量结构；谱梯度法的步长由历史梯度决定，对这类方法直接分析离散迭代更合适，由此可以讨论全局收敛、历史矩阵降秩时的处理以及最坏收敛速度等问题。结合我们的四篇预印本，本报告说明以下几点：BB 方法和 LMSD 都可以化为闭锥上一个连续正齐次映射的迭代，由每条轨道都收敛，可以得到常数与初值无关的 R-线性收敛；把 LMSD 一组迭代的乘子多项式写成若干固定多项式的凸组合，可以补全 Fletcher 原证明中取极限的一步；利用一个关于谱端点的不等式（下称端点不等式），结合 Poincaré 回复定理和半一致遍历定理，可以证明 BB 方法的最坏渐近收敛因子恰为 $(\kappa-1)/(\kappa+1)$；由同一个不等式，还能得到显式的单调 Lyapunov 函数。最后介绍向 Hilbert 空间的推广和非线性问题的局部结果，并讨论按数值秩截断、权随迭代变化，以及如何构造可以实际计算的 Lyapunov 函数等问题。

关键词：谱梯度法；BB 方法；有限记忆最速下降法；离散动力系统；正齐次映射；最坏渐近收敛因子；Lyapunov 函数

目录

1 引言与记号 **3**
1.1 为什么选择 BB1988 . . . 3
1.2 基本设定与收敛速度的几种描述 . . . 3

2 连续模型与离散模型 **4**
2.1 连续极限能说明什么 . . . 4
2.2 从连续能量回到离散算法 . . . 5

3 BB1988 原文精读 **5**
3.1 两点割线条件与 BB 步长 . . . 5
3.2 二维情形：比值的递推与范数的衰减 . . . 6
3.3 原文留下的问题 . . . 7

4 谱梯度法的研究脉络 **7**
4.1 步长只依赖当前梯度的方法 . . . 7
4.2 BB 方法的收敛理论 . . . 8
4.3 有限记忆最速下降法（LMSD） . . . 9
4.4 已有结论与我们的结果 . . . 10

5 闭锥上的齐次映射与 R-线性收敛 **12**
5.1 从轨道收敛到一致的几何收敛 . . . 12
5.2 BB 迭代的状态锥 . . . 13
5.3 加权 BB 方法与任意的首步步长 . . . 14

6 LMSD 的全局收敛与 R-线性收敛 **15**
6.1 历史矩阵的秩与有限步终止 . . . 15
6.2 原证明中的一个极限问题 . . . 16
6.3 乘子多项式的 Cauchy–Binet 表示 . . . 17
6.4 按特征值从小到大归纳 . . . 18
6.5 状态锥与映射的连续性 . . . 19
6.6 R-线性收敛 . . . 20
6.7 一个数值例子：满秩的历史矩阵也可能非常病态 . . . 22

7 BB 方法的最坏渐近收敛因子 **23**
7.1 问题与结论 . . . 23
7.2 分量比例与端点不等式 . . . 25
7.3 从不变测度到每一条轨道 . . . 26

8 BB 方法的 Lyapunov 函数 **28**
8.1 构造 . . . 28
8.2 指数的选取 . . . 29
8.3 端点分量为零时的处理 . . . 29
8.4 一个例子 . . . 31

9 比较、推广与展望 **32**
9.1 几种分析方法的比较 . . . 32
9.2 Hilbert 空间中的推广 . . . 32
9.3 非线性问题的局部结果 . . . 33
9.4 进一步的问题 . . . 33

1 引言与记号

1.1 为什么选择 BB1988

本报告精读的论文 [5] 是 Jonathan Barzilai 与 Jonathan M. Borwein 的 *Two-point step size gradient methods*，发表于 *IMA Journal of Numerical Analysis*，1988 年，第 8 卷第 1 期，141–148 页，DOI: 10.1093/imanum/8.1.141。选择它有两个原因。第一，它提出的步长只用相邻两点的信息，计算量很小，却明显改变了梯度法的表现：方法很有效，每一步却不一定下降，其中的道理值得弄清楚。第二，原文不只是提出步长、给出数值试验，还对二维问题推导出两个分量之比满足的精确递推。用动力系统的观点研究优化算法，这篇短文是较早的一个具体例子。

报告围绕一个问题展开：步长由过去的梯度决定时，应当怎样选取状态变量，才能既解释非单调的迭代过程，又能证明全局收敛、定量估计渐近收敛速度？我们会用到两种状态：不做归一化的状态保留了梯度的大小，适合研究收敛和稳定性；归一化的状态把大小和各分量所占的比例分开，适合研究长期的平均收缩速度。由第 7 节的端点不等式出发，还能构造一个用原变量表示、沿迭代单调下降的量。

报告中所说的“我们的结果”，出自作者与合作者的四篇论文 [34–37]，均按公开的预印本版本引用。全文以 BB1988 为出发点，讨论后续文献怎样回答它留下的问题，以及我们的工作补充了什么。第 4.4 小节的表 1 列出了已有结论和我们的结果，图 1 给出了证明路线，读者可以先看这两处，再读第 5–8 节的细节。

1.2 基本设定与收敛速度的几种描述

主要的研究对象是严格凸二次函数。记函数值误差为 $F(x)$，梯度为 $g(x)$：

$$F(x) := f(x) - f(x_*) = \frac{1}{2}(x-x_*)^{\mathsf{T}} H (x-x_*), \qquad g(x) = H(x-x_*), \quad H = H^{\mathsf{T}} \succ 0. \tag{1.1}$$

记 H 的最小、最大特征值为 $m = \lambda_{\min}(H)$、$L = \lambda_{\max}(H)$，条件数为 $\kappa = L/m$。一条迭代轨道实际涉及的特征值中，最小和最大的两个另记为 a、b（确切定义见各节），以免与整个谱的端点混淆。对 H 的每个不同的特征值 λ，记 P_λ 为相应特征子空间上的正交投影。

步长统一记为 η_k，其倒数记为 $\tau_k = 1/\eta_k$。从初始梯度 g_0 出发，第一步用任意正步长 η_0（下称首步步长）得到 $g_1 = (I - \eta_0 H) g_0$；此后 $k \ge 1$ 时的 BB 步长由 g_{k-1} 算出，却作用在 g_k 上。若迭代在有限步终止，此后的梯度都记为零。有限记忆最速下降法（LMSD）的记忆长度记为 p；它的迭代分组进行，每组至多 p 步（不终止时恰好 p 步），组的序号记为 j，以区别于 BB 方法的迭代序号 k。本报告所说的 LMSD 都是指这样的版本：上一组求出的全部 Ritz 值，其倒数在下一组中各作一次步长。

若误差序列满足 $\|e_{k+1}\| \le q\,\|e_k\|$，$q < 1$，则称该序列 Q-线性收敛，这时误差每一步至多变为原来的 q 倍；若只有 $\|e_k\| \le C\rho^k$，$\rho < 1$，则称该序列 R-线性收敛，后者允许误差在迭代中途反弹。本报告还要用到渐近收敛因子

$$\rho_g(g_0, \eta_0) = \limsup_{k\to\infty} \left(\frac{\|g_k\|}{\|g_0\|} \right)^{1/k}. \tag{1.2}$$

一条轨道的渐近收敛因子不超过 c，与“存在对所有初值都适用的常数 C，使 $\|g_k\| \le C c^k \|g_0\|$”含义不同。前者只保证对每个 $\gamma > c$，这条轨道有 γ^k 型的上界；常数 C 能否对所有初值一致地选取，以及 $\gamma = c$ 时这样的上界是否仍然成立，都需要另外证明。

对二次问题，梯度范数与误差范数等价，函数值误差与梯度范数的平方也等价：

$$m\left\|x_k-x_*\right\| \le \left\|g_k\right\| \le L\left\|x_k-x_*\right\|, \qquad \frac{1}{2L}\left\|g_k\right\|^2 \le F(x_k) \le \frac{1}{2m}\left\|g_k\right\|^2. \tag{1.3}$$

因此误差和梯度有相同的渐近收敛因子，函数值误差 $F(x_k)$ 的渐近收敛因子是它的平方。后文所说的最坏收敛因子都是针对梯度范数的；图中若画的是梯度范数的平方等其他量，会明确注明。

2 连续模型与离散模型

用动力系统的方法研究优化算法，第一步是把迭代看成某个状态的演化。连续模型便于看清阻尼、耗散和几何结构，离散模型则保留了相邻各步之间的依赖关系。两种模型都要回答两个问题：所选的状态是否足以确定下一步；所选的能量函数能否控制我们关心的误差。对动量法，位置和速度是自然的状态变量；对 BB 方法和有限记忆谱梯度法，决定步长的历史梯度也应当放进状态里。本报告采用离散的观点，是因为这类方法的步长本身就由这些历史信息决定。

2.1 连续极限能说明什么

Su、Boyd 与 Candès 研究了 Nesterov 加速梯度法（NAG）的小步长极限 [31]。记该算法的固定步长为 s，在 $t=k\sqrt{s}$ 的时间尺度下，凸光滑情形的极限方程是

$$\ddot{X}(t)+\frac{3}{t}\dot{X}(t)+\nabla f(X(t))=0, \tag{2.1}$$

初值为 $X(0)=x_0$、$\dot{X}(0)=0$。这里的 $3/t$ 不是随意加上的阻尼，而来自离散动量系数随迭代次数的变化。记 x_* 为最优解，则能量

$$\mathcal{E}(t)=t^2F(X(t))+2\left\|X(t)+\tfrac{t}{2}\dot{X}(t)-x_*\right\|^2 \tag{2.2}$$

满足

$$\dot{\mathcal{E}}(t)=2t\big(F(X(t))-\langle X(t)-x_*,\nabla f(X(t))\rangle\big)\le 0,$$

最后一步只用到凸性。于是 $F(X(t))\le 2\left\|x_0-x_*\right\|^2/t^2$：加速的收敛速度由一个简洁的耗散关系得到。这个推导告诉我们，函数值误差和动量变量应当怎样组合成能量。

但有限时间内的轨道逼近与长期的收敛速度是不同的问题。该文证明的是：在给定的初值下，对每个固定的 T，当 $s\downarrow 0$ 时，$0\le k\le T/\sqrt{s}$ 范围内的离散迭代趋于连续解。这是有限时间区间上的极限，不能直接用来推断 s 固定、$k\to\infty$ 时的渐近收敛因子；离散算法的加速收敛速度，仍需在适当的步长条件下另行证明。连续方程也可以按特征分量逐个分析二次函数，但要先弄清具体的极限过程丢掉了哪些信息。

低分辨率的极限方程会丢掉区分不同算法的信息。Shi、Du、Jordan 与 Su 提出的高分辨率微分方程，就是针对这一点的改进 [29]。在强凸情形，NAG 与文中用来比较的重球法有相同的低分辨率极限方程；保留 $\sqrt{s}$ 阶的项以后，NAG 对应的方程是

$$\ddot{X}+2\sqrt{\mu}\dot{X}+\sqrt{s}\,\nabla^2 f(X)\dot{X}+(1+\sqrt{\mu s})\nabla f(X)=0, \tag{2.3}$$

其中 $\mu>0$ 是强凸常数。式中含 Hesse 矩阵的项来自相邻两次梯度之差，原算法并不需要计算 Hesse 矩阵；这一项就是低分辨率模型丢掉的梯度校正项。作者随后重新构造离散能量，利用新增的耗散项控制离散化误差，得到了相应的步长范围和加速的线性收敛速度。即使保留了 s，连续模

型仍有截断误差，与原算法并不逐步对应。关于辛离散化的研究进一步说明，离散后能否保持加速，取决于方程、离散格式和误差估计三者是否配合得当 [28]。

另外两点也容易混淆。连续时间中的 $\exp(-ct)$，只有在时间与迭代次数的对应关系、步长的限制和每一步的计算量都明确以后，才能换算成算法的复杂度；对时间做变量替换，可以改变连续衰减的写法，却不会减少梯度的计算次数。对一般的系统，全局渐近稳定推不出指数稳定；要从前者得到后者，还必须利用线性、齐次性、紧性等足以控制收缩所需时间的结构。

2.2 从连续能量回到离散算法

Wibisono、Wilson 与 Jordan 从 Bregman 拉格朗日量出发，把多种加速模型纳入统一的变分框架 [32]。典型的能量形如

$$\mathcal{E}(t) = D_h(x_*, Z(t)) + A(t)F(X(t)),$$

其中 D_h 为由距离生成函数 h 确定的 Bregman 散度，Z 是位置与速度的组合，A 决定函数值误差衰减的快慢。选取不同的距离生成函数和时间权函数，可以得到不同的连续时间加速模型。不过，连续能量不增，并不保证直接用 Euler 格式离散后仍然稳定。该文构造离散算法时引入了辅助序列和相应的下降条件；有些辅助变量在小步长极限中合而为一，在离散证明中却提供了不可缺少的耗散。

Wilson、Recht 与 Jordan 进一步把连续和离散的 Lyapunov 分析对照起来 [33]。以其欧氏情形的一类离散能量为例，差分估计中含有

$$\begin{aligned}\mathcal{E}_{k+1} - \mathcal{E}_k \le{}& \frac{(A_{k+1}-A_k)^2}{2}\,\|\nabla f(x_{k+1})\|^2 \\ &+ A_{k+1}\bigl(f(y_{k+1}) - f(x_{k+1})\bigr).\end{aligned} \tag{2.4}$$

若取 $y_{k+1} = x_{k+1} - \nabla f(x_{k+1})/L$，由光滑性可以得到一个负的梯度平方项；条件 $(A_{k+1}-A_k)^2 \le A_{k+1}/L$ 保证它足以抵消第一项。可见离散分析要多做一步：用耗散项抵消差分带来的正误差项。

谱梯度法与时间积分之间也早有联系。Ascher 等把滞后的最速下降法解释为梯度流的变步长显式 Euler 格式 [2]。如果目的是准确跟踪瞬态，就必须顾及时间积分的误差；如果目的是尽快到达稳态，就可以采用别的步长策略。步长随迭代变化时，有时会超过常步长 Euler 格式的稳定性界限，算法却仍然收敛。每个特征分量在第 k 步的乘子是 $1-\eta_k\lambda$。本报告关心的是这些乘子长期累积的效果。为此，我们在状态中直接保留历史的 Rayleigh 商、一组迭代的乘子多项式，以及梯度在哪些特征子空间上有非零分量，而不先取小步长极限。

3 BB1988 原文精读

这篇论文 [5] 只有八页，既提出了两种步长，又对二维二次问题推导并分析了一个离散递推。后面介绍的研究，大体围绕它留下的三个问题展开：步长滞后为什么有效，非单调的迭代为什么收敛，以及一般维数下最坏能有多慢。

3.1 两点割线条件与 BB 步长

令 $s_{k-1} = x_k - x_{k-1}$，$y_{k-1} = g_k - g_{k-1}$。拟牛顿法用矩阵来满足割线条件，BB 方法则只用数量矩阵，在最小二乘意义下近似满足它。若用 τI 近似 Hesse 矩阵，则

$$\tau_k = \arg\min_{\tau\in\mathbb{R}} \|\tau s_{k-1} - y_{k-1}\|^2 = \frac{s_{k-1}^{\mathsf T} y_{k-1}}{s_{k-1}^{\mathsf T} s_{k-1}}, \qquad \eta_k^{\mathrm{BB1}} = \tau_k^{-1}.$$

若直接用 ηI 近似 Hesse 矩阵的逆，则

$$\eta_k^{\mathrm{BB2}} = \arg\min_{\eta\in\mathbb{R}} \|s_{k-1} - \eta y_{k-1}\|^2 = \frac{s_{k-1}^{\mathsf{T}} y_{k-1}}{y_{k-1}^{\mathsf{T}} y_{k-1}}.$$

原文先给出的是今天所说的 BB2，然后才是 BB1，不能按原文公式的顺序来对应今天的名称。在一维情形，两种步长相同，BB 方法就是割线法。它们只保存两点之差，却把曲率信息带进了步长，说明改进梯度法不一定要存储一个大矩阵。对一般的非线性函数，$s_{k-1}^{\mathsf{T}} y_{k-1}$ 不一定为正，原论文也没有证明不加保护措施的 BB 方法对一般函数全局收敛。

两种拟合用的误差度量不同，不能只看成同一个公式的代数变形。对严格凸二次函数，由 Cauchy–Schwarz 不等式得 $\eta_k^{\mathrm{BB2}} \le \eta_k^{\mathrm{BB1}}$，所以 BB2 和 BB1 也常分别称为短 BB 步长和长 BB 步长。这个大小关系只是在相同的历史信息下比较两个候选值，并不表示短步长在每个问题上都更稳定或更快。长期表现取决于所选步长怎样改变下一步各分量所占的比例，以及这一变化又怎样影响以后的步长。

对严格凸二次函数，记 Hesse 矩阵为 H，则 $y_{k-1} = H s_{k-1}$。利用 $s_{k-1} = -\eta_{k-1} g_{k-1}$，两个步长化为

$$\eta_k^{\mathrm{BB1}} = \frac{\|g_{k-1}\|^2}{g_{k-1}^{\mathsf{T}} H g_{k-1}}, \qquad \eta_k^{\mathrm{BB2}} = \frac{g_{k-1}^{\mathsf{T}} H g_{k-1}}{g_{k-1}^{\mathsf{T}} H^2 g_{k-1}}, \qquad k \ge 1. \tag{3.1}$$

这里第一步用正步长 η_0 由 g_0 得到 g_1，从第二步起才使用 BB 公式。精确线搜索的最速下降法用当前的 g_k 计算步长，BB1 则把同一个公式推迟了一步。此后的 BB 步长总在 $[1/L, 1/m]$ 中，但 κ 较大时，在某些特征值 λ 上，乘子的绝对值 $|1-\eta_k\lambda|$ 仍可能大于 1。因此步长有界不等于每步下降；收敛只能来自连续若干步乘子之间的相互制约，这种制约是由步长对历史梯度的依赖造成的。

3.2 二维情形：比值的递推与范数的衰减

二维情形可以完全算清楚。设 $H = \mathrm{diag}(a, b)$，$0 < a < b$，$g_k = (u_k, v_k)^{\mathsf{T}}$，暂设各分量都不为零，并令

$$\zeta_k = \frac{u_k^2}{v_k^2}.$$

将式 (3.1) 代入 $g_{k+1} = (I - \eta_k H) g_k$，对 BB1 有

$$\frac{u_{k+1}}{u_k} = \frac{(b-a)v_{k-1}^2}{a u_{k-1}^2 + b v_{k-1}^2}, \qquad \frac{v_{k+1}}{v_k} = -\frac{(b-a)u_{k-1}^2}{a u_{k-1}^2 + b v_{k-1}^2}.$$

两式相除再平方，复杂的分母消去，得到

$$\zeta_{k+1} = \frac{\zeta_k}{\zeta_{k-1}^2} \quad (\mathrm{BB1}), \qquad \zeta_{k+1} = \left(\frac{b}{a}\right)^2 \frac{\zeta_k}{\zeta_{k-1}^2} \quad (\mathrm{BB2}). \tag{3.2}$$

对 BB1 取 $z_k = \log \zeta_k$，对 BB2 取 $z_k = \log \zeta_k - \log(b/a)$，两者都化为

$$z_{k+1} = z_k - 2 z_{k-1}, \qquad r_\pm = \frac{1 \pm i\sqrt{7}}{2}, \qquad |r_\pm| = \sqrt{2}. \tag{3.3}$$

原论文的离散分析就是围绕这个线性递推展开的。只要 z_k 不恒为零，它的振幅就越来越大，两个分量轮流占优；两分量之比剧烈变化的同时，梯度的范数可以迅速减小。$\sqrt{2}$ 是递推特征根的模，也是超线性收敛阶的来源，它不是梯度范数的放大因子。

原文重点分析 BB2, 利用振荡的递推和数论估计证明了 R-超线性收敛; 作者特别说明, 这个分析要求第一步取最速下降步长。Dai 在 2013 年对 BB1 给出了更直接的梯度范数分析 [8]: 除下文所

说的平衡轨道外，两个分量都不为零时 BB1 是 $\sqrt{2}$ 阶 R-超线性收敛的，有 $\|g_k\| \le C\exp(-c(\sqrt{2})^k)$ 型的上界；而且在每连续三步中，至少有一步的比值 $\|g_{k+1}\| / \|g_k\|$ 渐近地趋于零（这并不是说所有相邻比值都趋于零）。若某个分量为零，迭代会在有限步内终止，要单独讨论，不能再用对数坐标。这是后人对原文的完善；两篇论文研究的算法和证明条件不尽相同，不宜简单合并成一个结论。

还有一类“平衡”轨道需要单独讨论。若对 BB1 有 $\zeta_0 = \zeta_1 = 1$，或对 BB2 有 $\zeta_0 = \zeta_1 = b/a$，则式 (3.3) 的解恒为零。首步步长取 $\eta_0 = 2/(a+b)$ 时，这样的迭代满足

$$\|g_k\| = c_{a,b}^{k}\,\|g_0\|\,, \qquad c_{a,b} = \frac{b-a}{b+a}. \tag{3.4}$$

原论文已经明确给出这类只有线性收敛的轨道（文中注明是袁亚湘向两位作者指出的），并指出它们在数值上不稳定。不稳定是指附近的轨道会逐渐远离这类轨道，并不是说这条精确的轨道不存在。二维的典型轨道超线性收敛，最坏的轨道却只有线性收敛，二者完全可以同时成立；研究最坏的渐近收敛因子时，不能因为这类轨道难以在浮点计算中长时间观察到，就把它们排除在外。

读这些结论时，还要分清三个概念：一步的比值 $\|g_{k+1}\| / \|g_k\|$ 可以大于 1；R-线性收敛要求有几何的上界；渐近收敛因子只刻画长时间的平均收缩。渐近收敛因子小于 1，并不意味着单步比值有极限，更不意味着每一步的比值都小于 1。函数值误差是梯度的二次型，所以对函数值误差而言，几何上界中的因子要平方。后文比较不同论文的结果时，会注明所用的范数和迭代次数的计法（按步还是按组）。

3.3 原文留下的问题

原论文的数值例子表明，同样沿负梯度方向，只改变步长规则就能明显减少迭代次数，函数值或梯度范数也可能出现反弹。这说明精确线搜索每一步最优，未必长期最优；但仅凭少数算例，不能说 BB 方法普遍优于共轭梯度法。原文的实质贡献是提出两点步长，并给出了一段可以逐步验证的二维分析。一般维数下的全局收敛、一致的 R-线性收敛和最坏的渐近收敛因子，都要由后来的工作解决。

从式 (3.2) 还可以看到，BB 方法的状态至少要包含相邻两步的信息：只给定 g_k 时，不同的 g_{k-1} 会给出不同的 η_k 和下一个梯度。二维时，取对数就把这种依赖关系化成了线性递推；高维时一般没有这样简单的显式递推。后面的分析沿用这一做法，即把相邻两步都放进状态。

4 谱梯度法的研究脉络

4.1 步长只依赖当前梯度的方法

用动力系统的观点研究梯度法，早在 BB 方法之前就开始了。Akaike 在 1959 年把精确线搜索的最速下降法化为概率分布的迭代 [1]。设 P_i 是 H 的第 i 个不同特征值对应的特征子空间上的正交投影，令 $w_i = \|P_i g\|^2 / \|g\|^2$、$\bar{\lambda} = \sum_i \lambda_i w_i$，则各分量比例的更新为

$$(Tw)_i = \frac{(\lambda_i - \bar{\lambda})^2 w_i}{\sum_{\ell}(\lambda_\ell - \bar{\lambda})^2 w_\ell}. \tag{4.1}$$

这个分布的方差在迭代中不减，增量可以写成矩的 Gram 行列式。利用单调性、紧性和等号成立的条件，可以说明极限分布只集中在两个特征值上；再比较乘子的大小，可知这两个特征值就是梯度的非零分量所对应的最小和最大特征值。奇数步和偶数步的极限一般是集中在这两个端点上

的两点分布，而不是分别集中在一个端点上。这一分析的思路今天仍然常用：在归一化的状态上找单调量，确定等号成立的集合，再控制其余分量所占的比例。

Forsythe 在 1968 年研究由当前梯度生成的 s 维最优梯度法，把正交多项式、归一化映射和极限集的结构联系起来 [15]；这种方法可以用每次从当前点重新开始的 s 步共轭梯度法实现。对一般的 s，论文证明了偶数步子序列的极限集是连通的，并具有两步不变性，但这一结论不能概括为“对任意的 s 都收敛到唯一的二周期轨道”。Pronzato、Wynn 与 Zhigljavsky 把这类只用当前梯度、权固定不变的方法推广到一般的有限维空间和 Hilbert 空间，用矩的单调性证明了极限分布集中在端点等性质 [24]。该文正式发表于 2006 年，常见的 2005 年是在线发表的时间。

这些结果提供了谱测度和正交多项式的工具，但不能直接用于 BB 方法。区别在式 (4.1) 中一看便知：把当前分布的均值换成上一步分布的均值后，下一个状态还依赖于前一个状态，原来关于单调量的证明也就不再适用。Pronzato 等在原文中也明确指出了 BB 方法与他们所研究方法的区别。Fletcher 在 2005 年的综述中已经写出了 BB 方法相邻两步分量比例的更新公式，并把解释其长期行为和多步平均收敛速度列为待解决的问题 [13]。因此，我们的工作是对这个已有问题的推进，并非首次把 BB 方法看成动力系统。

遍历方法也有直接的先例。Pronzato 与 Zhigljavsky 研究了渐近分布给定的步长倒数序列，用乘子绝对值的对数平均来计算长期的收缩速度 [23]。在那里，步长是由一个统计性质已知的动力系统给出的；BB 方法的步长则由前一个梯度决定，必须先研究 BB 方法自身状态空间上的不变测度和奇异集。两者都利用“乘积取对数变成平均”这一结构，需要验证的条件却不同。

4.2 BB 方法的收敛理论

Raydan 在 1993 年证明了有限维严格凸二次函数上 BB 方法的全局收敛 [25]。证明按特征值从小到大归纳：较小特征值上的分量先趋于零；若下一个分量始终保持一定大小，它会使随后的步长倒数变大，从而让自己进入收缩的范围。再说明分量离开小邻域时的反弹幅度有界，就能由子列的性质得到整个序列的收敛。这个机制说明了为什么不必要求每一步都下降。

Dai 与 Liao 在 2002 年把这种归纳推进到固定步数内的收缩，得到了 R-线性收敛 [10]。可见，在有限步内一致收缩这一思路并非我们首创；我们采用闭锥和齐次映射，是为了用统一的结构得到这种收缩，并处理算法的边界情形。Friedlander 等的带滞后的梯度法、Dai 的交替步长梯度法以及相应的统一分析框架说明，有界滞后和重复使用步长的规则也能严格分析 [7, 16]。但滞后的上界、候选步长的选取和保护条件都是定理条件的一部分，叙述结论时不能略去这些条件，笼统地说任意谱步长都收敛。

更精确的定量估计来自 Li 与 Sun 对 BB1 的分析 [19]。在首步步长也按 BB 公式由 g_0 计算的条件下，梯度的各分量有 θ^k 型的上界，$\theta = 1 - 1/\kappa$。而他们给出的平衡轨道，渐近收敛因子是 $c = (\kappa - 1)/(\kappa + 1) < \theta$。两者对条件数的依赖是同阶的，但精确常数仍有差距。Li 与 Huang 的 Property B 涵盖了带固定正权的步长和多种有限滞后的步长规则，并给出了各分量相应的 R-线性上界 [21]。这篇文章研究的是满足 Property B 的一类非单调梯度法；它涵盖的方法比全程使用同一种 BB 步长的方法多，所以 BB 方法的精确结论不能自动推广到整个 Property B 类。

二维的超线性收敛要和高维的现象分开看。Dai 与 Fletcher 用简化的递推解释了超线性与线性行为之间的转换，但对简化模型的结论，不能当作高维 BB 方法普遍超线性收敛的定理 [9]。对循环 BB 方法，已有三维的周期例子 [11]；Li、Jiang 与 Hong 最近的预印本则针对 BB 方法本身，对每个 $n \geq 4$ 给出了二次问题数据的一个开集，在这些数据下迭代有吸引的周期轨道（周期为 7），渐近收敛因子严格为正 [20]。这是存在性的结果，并不是对所有 Hesse 矩阵和几乎所有初值的分类。因此，对所有轨道一致成立的最坏速度、特定的吸引周期和典型初值下的表现，需要分开讨

论。

在 Hilbert 空间中也已有系统的研究。Azmi 与 Kunisch 对由有界、自伴、一致正定算子定义的二次问题，建立了 BB 方法的 R-线性收敛理论，并在后续工作中研究了非线性问题的局部收敛和网格无关性 [3, 4]。无限维的单位球不是紧的，连续谱的端点也不一定对应特征向量，所以证明不能照搬有限维中依赖紧性的论证。对一般的非线性问题，要保证全局收敛通常需要修改算法，例如 Raydan 把 BB 方法与非单调线搜索结合，证明了全局收敛性 [26]。

4.3 有限记忆最速下降法（LMSD）

Fletcher 的有限记忆最速下降法（LMSD）把前一个梯度的 Rayleigh 商推广为历史子空间上的多个 Ritz 值 [14]。记忆长度和组的序号仍记为 p 和 j。在满秩的情形，上一组的第一个梯度 u（每组的第一个梯度下称组首梯度）生成 Krylov 子空间 $\mathcal{K}_p(H,u)$，由此求出 Ritz 值 $\theta_{j,1},\dots,\theta_{j,p}$；记下一组的组首梯度为 v，下一组把这些 Ritz 值的倒数各用一次，组末的梯度为

$$v^{+}=\prod_{\ell=1}^{p}\left(I-\frac{H}{\theta_{j,\ell}}\right)v. \tag{4.2}$$

同一个多项式若作用在生成它的 u 上，就与共轭梯度法的 Galerkin 残差联系起来；LMSD 却把它作用在下一组的组首梯度 v 上，二者并不是同一个算法。Ritz 值、正交多项式、矩与 Gauss 求积之间的联系是熟知的 [17, 22]。需要新论证的是：多项式由 u 决定却作用在 v 上，两者之间应满足什么关系；历史子空间降秩时，映射又应当怎样定义。

几个名称相近的方法，用到的信息并不相同，相应的定理也就不能直接比较。下表列出了它们的区别。

方法	步长如何产生和使用	分析时不能忽略的特点
BB1/BB2	由前一个梯度得到一个 Rayleigh 商型的步长	步长滞后一步
循环 BB	一个 BB 步长连续使用若干步	周期长度及周期内的反弹
LMSD	由上一组梯度求出多个 Ritz 值，其倒数在下一组中各用一次	一组迭代的乘子多项式与历史矩阵的秩
Gu–Du 的修正 LMSD	每步重新计算候选值，从中选一个满足安全下界的值	选取规则与安全下界
实用的 LMSD 变体	可能筛选或截断 Ritz 值、改用其他步长，或提前丢弃旧梯度	子空间和每组的步数可能改变

Fletcher 原文附录中的全局收敛论证，在归一化的组首梯度趋于一个使 Krylov 子空间降秩的极限时，直接引用了 Ritz 值的连续性，而此时相应的 Krylov 子空间未必收敛到极限向量生成的子空间 [14]。阶数固定时，对称矩阵的特征值连续依赖于矩阵，这一点没有问题；困难在于 $u_n\to u$ 并不能保证 $\mathcal{K}_p(H,u_n)$ 趋于维数较低的 $\mathcal{K}_p(H,u)$。所以，不能只根据极限向量含有哪些分量，就给取极限之前的所有 Ritz 值一个共同的下界。后文补全的是这一步证明，并没有给出 LMSD 发散的反例。

Curtis 与 Guo 已经在附加条件下证明了记忆长度固定的 LMSD 是 R-线性收敛的 [6]。除了每个历史矩阵都列满秩外，他们还假设存在一个与迭代无关的常数 C，使 QR 分解中的 R_j 满足

$$\left\|R_j^{-1}\right\|\le\frac{C}{\|g_{j,0}\|}.$$

这个条件控制的是归一化以后的历史矩阵离退化有多远，比要求每组的历史矩阵都满秩更强。作者讨论了实际计算中删去旧列、缩短一组长度的做法，但没有据此证明所有变长版本的 R-线性收敛。Gu 与 Du 每步只选一个候选值，并给步长倒数加上安全的下界，证明了这种修改后的方法是 R-线性收敛的 [18]；它不需要关于历史矩阵的这类条件，但这并不是对原来的 LMSD 无条件地证明了 R-线性收敛。

满秩条件还与有限步终止有关。若组首梯度恰好在 s 个不同特征值的特征子空间上有非零分量，则 Krylov 子空间的维数为 $\min(p, s)$；当 $s \le p$ 时，它已经包含了这些分量的方向，下一组用完全部 Ritz 值的倒数后，这些分量就全被消去。特别地，$s = p$ 时历史矩阵仍可以列满秩，而迭代会在下一组内终止；所以不能由满秩推出 s 严格大于 p。另一方面，即使不终止的迭代每一组都满秩，归一化以后仍可能越来越接近降秩。这说明“每组都满秩”代替不了 Curtis 与 Guo 所要求的那种一致的定量条件。

Ferrandi 与 Hochstenbach 在 2025 年的工作中进一步研究了有限记忆方法的计算实现，并对其结构作了解释 [12]。他们比较了 Cholesky、列主元 QR 和 SVD 三种低维表示，研究了由普通 Ritz 值和调和 Ritz 值得到的步长，并用低维的最小二乘割线关系解释它们的来源；对非线性问题，还讨论了割线数据的对称化、带对称约束的拟合以及带保护的实现。按数值阈值截断，得到的可能不是 Krylov 子空间；对称拟合也不能自动保证正定。本报告按精确秩处理历史矩阵，所得的理论并不涵盖按任意阈值做 SVD 截断、提前丢弃旧梯度，以及非线性问题中带线搜索的实现。

袁亚湘最近从子空间逼近的角度，构造了基于二维和三维子空间的步长 [40]。其中基于位移子空间的步长，理论上与 $p = 2, 3$ 时的 LMSD 等价；二维情形在相邻两次位移几乎共线时改用 BB 步长，该文的收敛证明依赖于这一切换。历史梯度接近线性相关，是有限记忆方法分析的难点：该文用切换步长来处理，第 6 节则对不做切换、按精确秩计算的 LMSD 直接证明收敛。

Zou 与 Magoulès 的综述把滞后、交替、循环和有限记忆等方法放在统一的谱分析框架下讨论 [39]，其中 LMSD 的推广还包括把每个候选值重复使用多次的策略 [38]。本报告研究的有限记忆方法，则限定为每个 Ritz 值的倒数在下一组中各用一次。范围明确以后，问题就归结为以下三点：历史矩阵降秩时，一组迭代对应的映射是否仍然连续；在状态锥（第 5 节）上，每条轨道都收敛能否加强为一致的几何收敛估计；BB 方法乘子的对数沿轨道的长期平均，能否给出最坏渐近收敛因子的精确值。后文的主要结果分别回答这三个问题。

4.4 已有结论与我们的结果

表 1 把前面提到的主要收敛结论列在一起，横线以下是我们的结果。对 BB 方法，全局收敛和 R-线性收敛早已建立，缺少的是最坏渐近收敛因子的精确值：已知的上界 $1 - 1/\kappa$ 与平衡轨道的 $(\kappa - 1)/(\kappa + 1)$ 之间还有差距。κ 较大时后者约为 $1 - 2/\kappa$，按它估计，达到同样精度所需的步数大约只有按前者估计的一半。对 LMSD，Fletcher 的全局收敛证明在历史矩阵趋于降秩时，取极限的一步缺少依据；Curtis 与 Guo 的 R-线性收敛结果则要附加归一化历史矩阵一致非退化的条件，而这个条件无法事先验证。

图 1 给出了这两项结果的证明路线。图中上面一行处理 R-线性收敛：先把一组迭代的乘子多项式写成若干固定多项式的凸组合，由此证明 LMSD 从任意初值出发都收敛；再把 BB 方法和 LMSD 都化为闭锥上连续正齐次映射的迭代，利用紧性把“每条轨道都趋于零”加强为常数与初值无关的几何收敛。下面一行处理 BB 方法的最坏收敛因子：借助端点不等式，把 $\log(\|g_{k+1}\| \,/\, \|g_k\|)$ 精确地分解开，再用 Poincaré 回复定理和半一致遍历定理，把关于不变测度的结论推广到每一条轨道；把同一个分解式改写回未归一化的变量，就得到单调的 Lyapunov 函数。

表 1　严格凸二次函数上 BB 方法与 LMSD 的主要收敛结论。横线以下为我们的结果。

文献	方法	结论
Barzilai–Borwein 1988 [5]	BB（二维）	第一步取最速下降步长时，BB2 是 R-超线性收敛的；另有收敛因子为 $(\kappa-1)/(\kappa+1)$ 的平衡轨道
Raydan 1993 [25]	BB	任意维数全局收敛
Dai–Liao 2002 [10]	BB	R-线性
Li–Sun 2021 [19]；Li–Huang 2025 [21]	BB；一类非单调梯度法	R-线性，因子 $1-1/\kappa$
Fletcher 2012 [14]	LMSD	全局收敛；证明中取极限的一步缺少依据
Curtis–Guo 2018 [6]	LMSD	R-线性，需假设归一化的历史矩阵一致非退化
袁亚湘 2026 [40]	二维、三维子空间步长	全局收敛；R-线性
我们 [34, 35]	LMSD（含加权情形）	R-线性，无需上述假设，常数与初值无关；补全 Fletcher 的全局收敛证明
我们 [36, 37]	BB1，BB2	最坏渐近收敛因子恰为 $(\kappa-1)/(\kappa+1)$；给出显式的单调 Lyapunov 函数

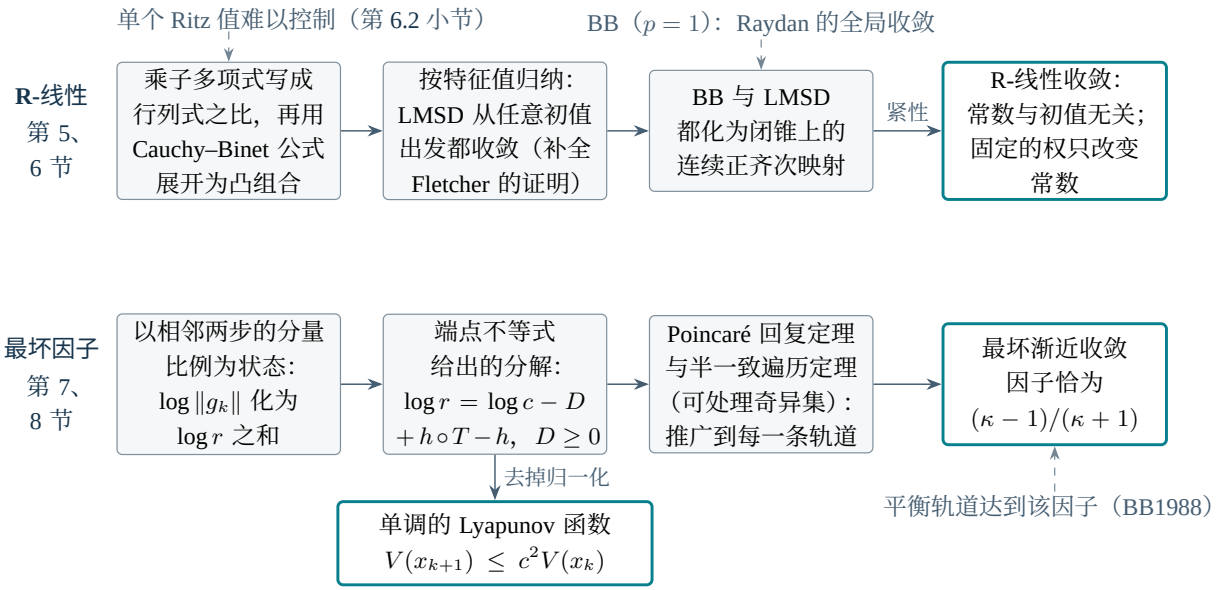


图 1　我们两项主要结果的证明路线。粗框为结论，虚线箭头表示用到的已有结果或例子。设 $a<b$ 分别是 g_0 的非零分量所对应的最小和最大特征值，图中的 $c=(b-a)/(b+a)$，它不超过 $(\kappa-1)/(\kappa+1)$。

5　闭锥上的齐次映射与 R-线性收敛

BB 方法是非单调的，只比较相邻两步的梯度范数，往往看不出迭代为什么收敛。另一种做法是把决定下一步所需的信息都放进状态：先证明从每个初始状态出发的轨道都趋于零，再利用映射的结构，把“每条轨道都趋于零”加强为对初值一致的几何收敛估计。我们的工作 [34] 采用的就是这一思路。其中用到的稳定性定理是已有的；我们的贡献是构造合适的状态空间：它要如实反映算法的迭代规则，要包含有限步终止的情形，并且其中每个状态都是算法从某个初值出发实际会经过的状态。

5.1　从轨道收敛到一致的几何收敛

设 X 为有限维赋范空间。若非空闭集 $\Omega \subset X$ 满足 $0 \in \Omega$，且对每个 $t \geq 0$ 有 $t\Omega \subset \Omega$，就称 Ω 为闭锥；这里不要求 Ω 是凸的。设映射 $T : \Omega \to \Omega$ 连续，并且是一次正齐次的，即

$$T(tz) = tT(z), \qquad z \in \Omega, \quad t \geq 0.$$

由齐次性，只需考虑范数为 1 的初值；有限维和闭性又保证这些初值构成紧集。两者结合起来，才能给不同方向上收缩所需的步数一个共同的上界。

定理 5.1（闭锥上齐次映射的稳定性）　设 Ω 是有限维赋范空间中的闭锥，$T : \Omega \to \Omega$ 连续且一次正齐次，则下列三个条件等价：

(i) 对每个 $z \in \Omega$，$T^k z \to 0$；

(ii) 存在与 z 无关的常数 $C \geq 1$ 及 $\rho \in (0,1)$，使得

$$\left\| T^k z \right\| \leq C\rho^k \left\| z \right\|, \qquad z \in \Omega, \quad k \geq 0; \tag{5.1}$$

(iii) 存在 $N \geq 1$ 及 $\theta \in (0,1)$，使得

$$\left\| T^N z \right\| \leq \theta \left\| z \right\|, \qquad z \in \Omega. \tag{5.2}$$

定理 5.1 是 Shen 与 Hu 的稳定性结果 [27, 定理 5.1] 在单个映射、一次齐次情形下的特例。下面给出一个基于有限子覆盖的证明，第 8 节处理端点分量为零的情形时，还会用到同样的思路。

证明　当 $\Omega = \{0\}$ 时结论显然。否则单位截面 $\Sigma = \{z \in \Omega : \|z\| = 1\}$ 非空且紧。先证 (i)⇒(ii)。对每个 $y \in \Sigma$，由 $T^k y \to 0$ 知，存在 $n_y \geq 1$，使得 $\|T^{n_y} y\| < 1/2$。由连续性，存在 y 在 Σ 中的相对邻域 U_y，使得对任意 $w \in U_y$ 都有 $\|T^{n_y} w\| \leq 1/2$。由 Σ 的紧性，可取有限子覆盖 $U_{y_1}, \dots, U_{y_s}$。令

$$N_0 = \max_{i} n_{y_i}, \qquad A = \max_{0 \leq r \leq N_0} \max_{y \in \Sigma} \|T^r y\| < \infty.$$

于是对每个 $y \in \Sigma$，存在不超过 N_0 的步数 n，使 $\|T^n y\| \leq 1/2$；但 n 可能随 y 而变，所以还不能断言 $\left\|T^{N_0} y\right\| \leq 1/2$。

固定非零的 z，令 $z_0 = z$，$t_0 = 0$。只要 $z_q \neq 0$，就对单位向量 $z_q / \|z_q\|$ 取相应的步数 $m_q \in \{1, \dots, N_0\}$，并令

$$t_{q+1} = t_q + m_q, \qquad z_{q+1} = T^{m_q} z_q.$$

由齐次性，$\|z_{q+1}\| \leq \|z_q\| / 2$，故 $\|z_q\| \leq 2^{-q} \|z\|$。若某个 $z_q = 0$，则此后各项均为零。否则 $t_q \to \infty$；对任意 k，取 q 使 $t_q \leq k < t_{q+1}$，则

$$\left\| T^k z \right\| \leq A \left\| z_q \right\| \leq A 2^{-q} \left\| z \right\|.$$

又因 $k < t_{q+1} \le (q+1)N_0$，有 $q > k/N_0 - 1$，从而

$$\left\|T^k z\right\| \le 2A\big(2^{-1/N_0}\big)^k \|z\|. \tag{5.3}$$

这里的常数与初值无关，且 $A \ge 1$，故上式对 $k = 0$ 和有限步终止的情形也成立。这就证明了 (i)⇒(ii)。

(ii)⇒(iii)：只需取 N 使 $C\rho^N < 1$。

(iii)⇒(ii)：令

$$A_N = \max_{0\le r<N} \max_{y\in\Sigma} \|T^r y\|.$$

对 $k = qN + r$，$0 \le r < N$，有

$$\left\|T^k z\right\| \le A_N \theta^q \|z\| \le \frac{A_N}{\theta}\big(\theta^{1/N}\big)^k \|z\|.$$

(ii)⇒(i) 是显然的。于是三个条件等价。 □

这里所说的“一致”有两层意思。第一，常数 C 和 ρ 对锥中所有初值都适用，这比每条轨道各有自己的常数要强；第二，步数 N 也与初值无关，但不要求 $N = 1$。对 BB 这样的非单调方法，允许在 N 步之内有增长，正好避开了“每一步都按固定比例下降”（即 Q-线性收敛）这一过强的要求。还要注意，这里的范数是状态的范数（对 BB 方法，状态由相邻两个梯度组成），而不是单个梯度的范数。

为了区分“存在一致的收敛因子”和“确定最优的一致收敛因子”这两个问题，对非平凡的锥引入增长率：令

$$M_k = \max_{z\in\Sigma} \left\|T^k z\right\|, \qquad M_0 = 1.$$

由齐次性，$M_{k+\ell} \le M_k M_\ell$，从而

$$r(T) = \lim_{k\to\infty} M_k^{1/k} = \inf_{k\ge 1} M_k^{1/k} = \inf\{\gamma > 0 : \text{存在 } C_\gamma, \text{使对一切 } k \text{ 有 } M_k \le C_\gamma \gamma^k\}.$$

事实上，对任意 $\gamma > r(T)$，当 k 充分大时已有 $M_k \le \gamma^k$，前面有限多步可以并入 C_γ；反过来的不等式开 k 次方即得。定理保证 $r(T) < 1$，但没有给出它的值，也不保证在 $\gamma = r(T)$ 时还有一致的常数。确定最优的一致因子属于第 7 节的最坏收敛因子问题；本节只证明它小于 1。

证明并不要求范数每一步都下降。不同方向达到收缩所需的步数可以不同，中间的增长由 A 控制；紧性保证这些步数有一致的上界。这也说明，借用动力系统中稳定性的概念时，必须保留齐次性和单位截面的紧性：一般情况下，渐近稳定并不蕴含指数稳定。例如，在 ℓ^2 上定义对角算子 T，使 $Te_j = (1 - 1/j)e_j$，则每条轨道都趋于零，但 $\|T^k\| = 1$，不存在小于 1 的一致几何因子。

5.2 BB 迭代的状态锥

沿用严格凸二次函数的设定，$H = H^{\mathsf T} \succ 0$，$m = \lambda_{\min}(H)$，$L = \lambda_{\max}(H)$。对固定的正定权矩阵 $W = \omega(H) \succ 0$，定义

$$\eta_W(u) = \frac{u^{\mathsf T} W u}{u^{\mathsf T} W H u}, \qquad u \ne 0.$$

它满足 $L^{-1} \le \eta_W(u) \le m^{-1}$。取 $W = I$ 得到 BB1，取 $W = H$ 得到 BB2；二次函数上的迭代统一写为

$$g_{k+1} = \big(I - \eta_W(g_{k-1})H\big) g_k, \qquad k \ge 1. \tag{5.4}$$

由于步长由 g_{k-1} 算出、却作用在 g_k 上，状态至少要包含相邻的两个梯度。

下面称区间 $[L^{-1}, m^{-1}]$ 中的步长为允许的步长。如果取整个乘积空间作为状态空间，就会遇到 $(0, v)$（$v \neq 0$）这样的状态：第一个分量为零时 Rayleigh 商没有定义，而以允许的步长作一步梯度迭代，也不会产生这样的状态。为此取

$$\Omega_{\mathrm{BB}} = \big\{(u, v) : \text{存在 } \eta \in [L^{-1}, m^{-1}],\ \text{使 } v = (I - \eta H)u\big\}, \tag{5.5}$$

配以乘积范数 $\|(u, v)\|_{\times} = (\|u\|^2 + \|v\|^2)^{1/2}$，并定义

$$T_W(u, v) = \big(v, (I - \eta_W(u)H)v\big) \quad (u \neq 0), \qquad T_W(0, 0) = (0, 0). \tag{5.6}$$

命题 5.2　Ω_{BB} 是闭锥，T_W 是 Ω_{BB} 到自身的连续、一次正齐次映射。

证明　锥的性质由定义直接得到。设 $(u_\ell, v_\ell) \in \Omega_{\mathrm{BB}}$ 且 $(u_\ell, v_\ell) \to (u, v)$，相应的步长 η_ℓ 都属于紧区间 $[L^{-1}, m^{-1}]$，不妨设 $\eta_\ell \to \eta$；在 $v_\ell = (I - \eta_\ell H)u_\ell$ 中取极限即得闭性。由定义还可知，$u = 0$ 时必有 $v = 0$。

记 $B = \max\{1, L/m - 1\}$，则对所有允许的步长有 $\|I - \eta H\| \le B$。在 $u \neq 0$ 处，Rayleigh 商连续；在原点处，映射第二个分量的范数不超过 $B\,\|v\|$。因此，尽管 $u \to 0$ 时 $\eta_W(u)$ 本身没有极限，映射在原点处仍然连续。当 $v = 0$ 时迭代已经终止，这样的状态被映到 $(0, 0)$。映射的像 $\big(v, (I - \eta_W(u)H)v\big)$ 所用的步长 $\eta_W(u)$ 是允许的步长，所以像仍在锥中；Rayleigh 商是零次齐次的，故映射是一次正齐次的。 □

锥的定义来自实际的迭代，并没有为了证明方便而去掉某些初值。在这个集合上，前一个梯度趋于零时，当前梯度一定也趋于零；所以即使步长在原点附近没有极限，乘到当前梯度上以后，所得的映射仍然连续。如果改在整个乘积空间上定义映射，再随意补充 $u = 0$ 处的步长，就失去了这一关系，映射也不再连续。

还需说明：从锥中每个状态出发的轨道都收敛。零状态是平凡的，以下设 $u \neq 0$。对任意 $(u, v) \in \Omega_{\mathrm{BB}}$，取满足 $v = (I - \eta H)u$ 的步长 η，令

$$x_0 = x_* + H^{-1}u, \qquad x_1 = x_* + H^{-1}v.$$

从 x_0 到 x_1 就是以 η 为步长的一次梯度迭代。记 $s = x_1 - x_0 = -\eta u$，$y = Hs = -\eta H u$，则下一个 BB1 步长恰为 $\|u\|^2 / (u^{\mathsf{T}} H u)$。这样，$T_I$ 从任一状态出发的轨道，都是 BB1 以 x_0 为初始点、以 η 为首步步长的实际迭代。由 Raydan 的全局收敛定理 [25]，对所有 $z \in \Omega_{\mathrm{BB}}$ 都有 $T_I^k z \to 0$。再由命题 5.2 和定理 5.1，就得到只依赖于 H 的常数 $C_H \ge 1$ 和 $\rho_H \in (0, 1)$，使式 (5.1) 对 T_I 成立。

5.3　加权 BB 方法与任意的首步步长

设 $S = W^{1/2}$，在乘积空间上定义 $\mathcal{S}(u, v) = (Su, Sv)$。由于 S 与 H 可交换，$\mathcal{S}$ 是 Ω_{BB} 到自身的双射，并且

$$\eta_I(Su) = \eta_W(u), \qquad \mathcal{S} T_W = T_I \mathcal{S}.$$

于是

$$\left\| T_W^k z \right\|_{\times} \le C_H \sqrt{\kappa_2(W)}\, \rho_H^k \, \|z\|_{\times}, \qquad z \in \Omega_{\mathrm{BB}}, \tag{5.7}$$

其中 $\kappa_2(W) = \|W\| \left\|W^{-1}\right\|$。一致的几何因子 ρ_H 与权 W 无关，权只通过范数等价影响常数。这里比较的是经过变换后相对应的初值，并不表示从同一个 g_0 出发时，BB1 和 BB2 每一步的表现相同。

线性共轭还保持上面的增长率 $r(T)$：共轭前后的状态范数至多相差固定的倍数 $\|S\|\,\|S^{-1}\|$，开 k 次方再令 $k\to\infty$，这个倍数的影响就消失了。因此，权固定的各种方法有相同的最优一致收敛因子，而它们在有限步内的实际表现仍可能相差很大。若每一步都改变 W，就不存在统一的坐标变换，上述论证不再适用。

按本报告的约定，先用任意正步长 η_0 由 g_0 得到 g_1。首步步长可能不在 $[L^{-1},m^{-1}]$ 中，(g_0,g_1) 也就不一定在锥中；但第一个 BB 步长 $\eta_1=\eta_W(g_0)$ 是允许的步长，所以 $(g_1,g_2)\in\Omega_{\mathrm{BB}}$。有限步终止后记梯度为零，则对 $k\ge 1$ 有

$$\|(g_k,g_{k+1})\|_\times \le C_H\sqrt{\kappa_2(W)}\,\rho_H^{k-1}\,\|(g_1,g_2)\|_\times\,. \tag{5.8}$$

误差满足 $\|x-x_*\|\le m^{-1}\|g\|$，函数值误差满足 $F(x)=\frac12 g^{\mathsf T}H^{-1}g\le\|g\|^2/(2m)$，故函数值误差的几何因子为 ρ_H^2。前面有限多步不影响渐近收敛因子；但首步步长可以任意大，所以以 $\|g_0\|$ 为基准时，常数不可能与首步步长无关。

这一分析建立在已有的收敛结论之上，并没有代替 Raydan 的全局收敛定理，也没有给出最优的 ρ_H。最坏因子究竟是多少，要由第 7 节对分量比例的遍历分析来回答。

6 LMSD 的全局收敛与 R-线性收敛

LMSD 利用一组历史梯度所含的谱信息，构造下一组的多个步长。如果直接分析单个 Ritz 值，难点不仅在于步长滞后，还在于归一化的历史矩阵可能越来越接近降秩。我们的工作 [35] 研究一组步长的乘子多项式：先利用它的行列式表示证明所有轨道都趋于零，再证明映射在历史矩阵降秩的状态处仍然连续，最后应用第 5 节的定理。这里的 LMSD 专指按精确秩求 Ritz 值、每个 Ritz 值的倒数在下一组中各用一次的版本，不考虑把步长重复使用多次的变体。

6.1 历史矩阵的秩与有限步终止

固定记忆长度 $p\ge 1$。第 j 组含 ℓ_j 步迭代，其历史矩阵由各步迭代之前的梯度组成：

$$G_j=[g_{j,0},\dots,g_{j,\ell_j-1}],\qquad g_{j,\ell_j}=g_{j+1,0}.$$

组末梯度 g_{j,ℓ_j} 就是下一组的组首梯度，不计入本组的历史矩阵。初始组以任意正步长迭代 p 步，若未提前终止，则 $\ell_0=p$。给定 G_j，取其列空间的标准正交基 Q_j，记 $d_j=\operatorname{rank}G_j$，由

$$Q_j^{\mathsf T}HQ_jz=\theta z$$

求出 d_j 个正的 Ritz 值。从组首梯度 $g_{j+1,0}$ 出发，依次以这些 Ritz 值的倒数为步长，构成下一组，于是 $\ell_{j+1}=d_j$。各步的因子 $I-\eta_hH$ 可以交换，所以组末梯度与步长的使用顺序无关，组内各步的梯度则可能不同。

任一由 ℓ 个正步长得到的历史矩阵满足

$$\operatorname{span}G=\mathcal K_\ell(H,u):=\operatorname{span}\{u,Hu,\dots,H^{\ell-1}u\}, \tag{6.1}$$

其中 u 为该组的组首梯度。事实上，第 r 列（$r=0,\dots,\ell-1$）是 $q_r(H)u$，其中多项式

$$q_r(t)=\prod_{h<r}(1-\eta_h t)$$

的次数为 r，首项系数不为零，所以 G 的各列与 $u, Hu, \dots, H^{\ell-1}u$ 之间只差一个可逆的上三角变换。

为了把重特征值的情形也包括进来，把 H 写成谱分解

$$H = \sum_{i=1}^{\nu} \lambda_i P_i, \qquad 0 < \lambda_1 < \cdots < \lambda_\nu,$$

其中 P_i 为相应特征子空间上的正交投影。记 $S(u) = \{i : P_i u \neq 0\}$，$s(u) = |S(u)|$。把非零向量 $P_i u$（$i \in S(u)$）单位化，作为它们所张成子空间的基，H 在这个子空间上的限制在该基下是以互异特征值为对角元的对角矩阵。由 Vandermonde 矩阵的秩可知

$$\operatorname{rank} K_p(u) = \min\{p, s(u)\}, \qquad K_p(u) = [u, Hu, \dots, H^{p-1}u]. \tag{6.2}$$

特别地，$s(u) = p$ 时矩阵仍然列满秩；这时迭代与 $s(u) < p$ 时一样会在有限步内终止，但历史矩阵并不秩亏。

命题 6.1（下一组有限步终止） 设某一组有 p 步，其组首梯度 $u \neq 0$，且 $s(u) = d \leq p$，则该组历史矩阵的列空间就是 H 的不变子空间 $\operatorname{span}\{P_i u : i \in S(u)\}$，求出的 *Ritz* 值恰为 $\{\lambda_i : i \in S(u)\}$；下一组从它的组首梯度 v（即本组的组末梯度）出发，以这些 *Ritz* 值的倒数为步长，至多 d 步就终止。

证明 由式 (6.1) 和式 (6.2)，历史矩阵列空间的维数为 d，且包含在同样是 d 维的 $\operatorname{span}\{P_i u : i \in S(u)\}$ 中，故二者相等。投影后的矩阵就是 H 在这个不变子空间上的限制。下一组的组首梯度是 $v = q(H)u$，其中 q 为本组各步步长对应的乘积多项式，从而 $S(v) \subset S(u)$，于是

$$\prod_{i \in S(u)} (I - \lambda_i^{-1} H) v = 0.$$

□

在不终止的迭代中，每组的历史矩阵都有 p 列，且 $s(g_{j,0}) > p$。否则，取第一个使 $d_j < p$ 的组：在它之前各组历史矩阵的秩都是 p，所以这一组仍有 p 步；由式 (6.2)，其组首梯度只在少于 p 个特征值上有分量，再由命题 6.1，下一组就会终止，矛盾。若历史矩阵满秩但 $s(g_{j,0}) = p$，同一命题也导出矛盾。因此不终止时每组都恰好有 p 步；若 $p \geq \nu$，则至多 $p + \nu$ 步就终止。

6.2 原证明中的一个极限问题

Fletcher 的收敛性证明 [14] 中有一步取极限，需要由“归一化的组首梯度在较小特征值上的分量趋于零”，推出取极限之前历史子空间中全部 Ritz 值都有相应的下界。仅凭这些信息，这个推断并不成立。取

$$H = \operatorname{diag}(1, 2, 3), \qquad u_\varepsilon = \frac{(\varepsilon, \varepsilon^2, 1)^{\mathsf{T}}}{(1 + \varepsilon^2 + \varepsilon^4)^{1/2}}, \qquad p = 2.$$

虽然 $u_\varepsilon \to e_3$，但 $(H - 3I)u_\varepsilon$ 归一化后趋于 $-e_1$，从而

$$\mathcal{K}_2(H, u_\varepsilon) \longrightarrow \operatorname{span}\{e_1, e_3\}, \qquad \{\theta_{\varepsilon,1}, \theta_{\varepsilon,2}\} \longrightarrow \{3, 1\}.$$

组首梯度在较小特征值上的分量趋于零，历史子空间却仍可以保留相应特征向量的方向。对照 Fletcher 附录中的式 (36)，$r = 3$ 时所需的下界是 $\theta > 2\lambda_3/3 = 2$，而此例中最小的 Ritz 值趋于

1。这说明在历史子空间不受控制时，这个下界没有依据。这只是一族向量取极限的例子，用来检验那一步推理，不是一条发散的 LMSD 迭代轨道，也不否定 Fletcher 的收敛结论。

Curtis 与 Guo [6] 的收敛速度分析明确假设历史矩阵列满秩，并且归一化后一致非退化：若 $G_j = Q_j R_j$，要求存在与 j 无关的常数 C，使 $\left\|R_j^{-1}\right\| \le C/\left\|g_{j,0}\right\|$。仅由每个 G_j 都列满秩，推不出这个一致的条件。Ferrandi 与 Hochstenbach [12] 着重讨论谱逼近、割线结构和算法实现，他们对数值秩的处理，与这里按精确秩计算的模型不能直接等同。我们的证明不假设上面的逆矩阵界，而是直接研究一组步长的乘子多项式，即全部因子 $1 - t/\theta_{j,h}$ 之积。

这样，我们的结果不再需要“归一化的历史矩阵一致非退化”这一额外条件；严格凸二次函数、每组用完全部 Ritz 值、按精确秩计算等设定则仍然保留。

6.3 乘子多项式的 Cauchy–Binet 表示

设迭代不终止，记 $u_j = g_{j,0}$。由 u_j 所在组的历史矩阵确定的乘子多项式为

$$\Psi_{u_j}(t) = \prod_{h=1}^{p} (1 - t/\theta_{j,h}), \qquad u_{j+2} = \Psi_{u_j}(H)\, u_{j+1}. \tag{6.3}$$

状态仍然要取相邻的两个组首梯度，而且多项式作用在下一组的组首梯度上。下面把依赖于 Ritz 值的乘子多项式 Ψ_u 写成固定多项式的凸组合，组合的权随状态变化且非负。

命题 6.2（乘子多项式的凸组合表示） 设 $s(u) \ge p$，则

$$\Psi_u(t) = \frac{\det\big(K_p(u)^{\mathsf T}(H - tI)K_p(u)\big)}{\det\big(K_p(u)^{\mathsf T} H K_p(u)\big)}. \tag{6.4}$$

设 $\mathcal{I}_p$ 为 $\{1, \dots, \nu\}$ 的所有 p 元子集组成的集合，$V = (\lambda_i^{h-1})$ 为 $\nu \times p$ 矩阵，V_I 为 V 中由 I 对应的行组成的 $p \times p$ 子矩阵，$\Delta_I = \det V_I$。定义

$$\beta_I(u) = \Delta_I^2 \prod_{i \in I} \lambda_i \left\|P_i u\right\|^2, \qquad \phi_I^{(r)} = \prod_{i \in I} (1 - \lambda_r/\lambda_i), \tag{6.5}$$

$$\pi_I(u) = \frac{\beta_I(u)}{\sum_{J \in \mathcal{I}_p} \beta_J(u)}. \tag{6.6}$$

则 $\pi_I(u) \ge 0$，$\sum_I \pi_I(u) = 1$，并且

$$\Psi_u(\lambda_r) = \sum_{I \in \mathcal{I}_p} \pi_I(u)\, \phi_I^{(r)}. \tag{6.7}$$

证明 记 $K = K_p(u)$。由 K 列满秩知 $K^{\mathsf T} H K \succ 0$。用 K 代替历史子空间的标准正交基，广义特征值不变，所以

$$\det(K^{\mathsf T} H K - t K^{\mathsf T} K) = \det(K^{\mathsf T} K) \prod_{h=1}^{p} (\theta_h - t).$$

把此式在 t 处和 0 处的值相除，即得式 (6.4)。

由谱分解，

$$K^{\mathsf T} H K = V^{\mathsf T} \operatorname{diag}\big(\lambda_i \left\|P_i u\right\|^2\big) V.$$

于是由 Cauchy–Binet 公式，分母可以写成

$$\det(K^{\mathsf T} H K) = \sum_{I \in \mathcal{I}_p} \Delta_I^2 \prod_{i \in I} \lambda_i \left\|P_i u\right\|^2 = \sum_{I} \beta_I(u) > 0.$$

同样，把分子展开，得

$$\det\big(K^{\mathsf T}(H-\lambda_r I)K\big)=\sum_{I}\Delta_I^2\prod_{i\in I}(\lambda_i-\lambda_r)\,\|P_i u\|^2=\sum_{I}\beta_I(u)\phi_I^{(r)}.$$

这里并不要求 $H-\lambda_r I$ 正定，正定性只用于分母。特征值互异保证 $\Delta_I\neq 0$，而 $s(u)\ge p$ 保证至少有一个 p 元子集 $I\subset S(u)$，使 β_I 严格为正。两式相除即得结论。 □

多项式的值 $\phi_I^{(r)}$ 只依赖于 H 和 p，对状态的依赖全部在权 $\pi_I(u)$ 中。但 $\phi_I^{(r)}$ 可以为负，绝对值也可以大于 1，所以凸组合本身并不意味着收缩。有用的是两点：一是不同项的权之比有显式的表达式；二是较小特征值上的分量趋于零时，含这些下标的项，其权也随之趋于零。

与直接估计 Ritz 值相比，这个表示保留了整个多项式中各因子之间的代数联系。分母是正的 Gram 行列式，即使某些分量很小也仍然可用；分子允许变号。因此分析时不必先把每个因子都控制在 $(-1,1)$ 内，也不必要求一组之内梯度范数单调下降；需要控制的，只是有限多个项的权之间的相对大小。

6.4 按特征值从小到大归纳

步长滞后带来了一个需要单独处理的问题：只有上一组组首梯度在 λ_r 上的分量足够大时，才能保证从当前组到下一组，这个分量是收缩的。下面的引理处理的就是这种下标错开的关系，其基本思路来自 Fletcher 的附录 [14]。

引理 6.3（数列引理） 设 $a_j\ge 0$，且对 $j\ge j_0$ 有 $a_{j+2}\le M a_{j+1}$，其中 $M\ge 1$。若存在 $c\in(0,1)$，使得对任意 $\varepsilon>0$，都存在 $j_\varepsilon\ge j_0$ 满足

$$j\ge j_\varepsilon,\quad a_j\ge\varepsilon\quad\Longrightarrow\quad a_{j+2}\le c\,a_{j+1},\tag{6.8}$$

则 $a_j\to 0$。

证明 固定 $\delta>0$。必有某个 $N\ge j_\delta+1$ 使 $a_N<\delta$；否则对所有 $j\ge j_\delta+1$ 都有 $a_j\ge\delta$，由条件 (6.8)，a_j 此后按几何级数下降，与这个下界矛盾。

下面证明 $a_j<M^2\delta$ 对所有 $j\ge N$ 成立。$j=N$ 时显然，且 $a_{N+1}\le Ma_N<M\delta$。设 $n\ge N+1$，且结论对 $N\le j\le n$ 成立。若 $a_{n-1}<\delta$，两次利用 $a_{j+2}\le Ma_{j+1}$，得 $a_{n+1}\le Ma_n\le M^2a_{n-1}<M^2\delta$；若 $a_{n-1}\ge\delta$，因 $n-1\ge j_\delta$，由条件 (6.8) 得 $a_{n+1}\le ca_n<M^2\delta$。归纳完成，所以 $\limsup_j a_j\le M^2\delta$；再令 $\delta\downarrow 0$ 即可。 □

定理 6.4（LMSD 的全局收敛） 设 f 为严格凸二次函数，*LMSD* 的记忆长度 p 固定，按精确秩求 *Ritz* 值，初始组采用任意正步长，则迭代或者在有限步内终止，或者所有组首梯度及组内各步的梯度都趋于零。

证明 只需讨论不终止的情形，这时每组恰有 p 步。与第 5 节一样，记 $B=\max\{1,L/m-1\}$。Ritz 值都在 $[m,L]$ 中，所以

$$\big|\Psi_{u_j}(\lambda_r)\big|\le B^p.\tag{6.9}$$

下面按互异特征值从小到大归纳。对 $r=1$，含下标 1 的项为零，其余各项的 $\phi_I^{(1)}$ 都非负且不超过 $\rho_1=(1-\lambda_1/\lambda_\nu)^p<1$。由式 (6.7) 得 $\big|\Psi_{u_j}(\lambda_1)\big|\le\rho_1$，再由式 (6.3) 得 $P_1u_j\to 0$。

设对 $i<r$ 已有 $P_iu_j\to 0$。固定 $\varepsilon>0$，只考虑满足 $\|P_ru_j\|\ge\varepsilon$ 的那些 j。把各项分成三类：含 r 的项为零；下标都大于 r 的项，绝对值不超过某个 $\rho_r<1$，可取 $\rho_r=(1-\lambda_r/\lambda_\nu)^p$（这一类为空时取 $\rho_r=0$）；其余的 I 不含 r，但含有某个 $i(I)<r$。对最后一类，令

$$J=(I\setminus\{i(I)\})\cup\{r\}.$$

若 $\beta_I(u_j)=0$，该项为零；否则由 $\|P_ru_j\|\ge\varepsilon$ 知 $\beta_J(u_j)>0$，并且

$$\pi_I(u_j)\le\frac{\beta_I(u_j)}{\beta_J(u_j)}=\frac{\Delta_I^2}{\Delta_J^2}\frac{\lambda_{i(I)}}{\lambda_r}\frac{\left\|P_{i(I)}u_j\right\|^2}{\|P_ru_j\|^2}\le\frac{C_I}{\varepsilon^2}\left\|P_{i(I)}u_j\right\|^2\longrightarrow 0. \tag{6.10}$$

可能造成放大的项只有有限多个，而且 $\phi_I^{(r)}$ 都是与 j 无关的常数，从而沿这些 j 有

$$\limsup_j\left|\Psi_{u_j}(\lambda_r)\right|\le\rho_r<1.$$

若这样的 j 只有有限个，条件 (6.8) 在其后自然成立。否则取 $c_r=(1+\rho_r)/2<1$，对每个 ε 都可选足够大的 j_ε，使乘子的绝对值不超过 c_r。

令 $a_j=\|P_ru_j\|$。精确的滞后递推是 $a_{j+2}=\left|\Psi_{u_j}(\lambda_r)\right|a_{j+1}$，而条件 (6.8) 的前提是关于 a_j 的。式 (6.9) 给出引理中的 $M=B^p$，上面的估计给出条件 (6.8)，于是 $a_j\to 0$。归纳完成，得 $u_j\to 0$。

每组至多有 p 个因子，每个因子的范数不超过 B，所以 $\max_{0\le r\le p}\|g_{j,r}\|\le B^p\|u_j\|\to 0$。再由 $g=H(x-x_*)$，全部迭代点都趋于 x_*。 □

定理 6.4 的证明没有假设历史矩阵的条件数一致有界。它的机制也比“Ritz 值逼近 Hesse 矩阵的特征值，所以应当加速”更具体：较小特征值上的分量趋于零以后，那些可能放大当前分量的项，其权也随之趋于零；数列引理再把这种有条件的收缩变成真正的收敛。

这个归纳并没有给出一个对所有 j 都成立的收缩因子：阈值 ε 越小，所需的 j_ε 可能越大。所以单凭这段证明，还得不到显式的几何收敛速度。数列引理先证明收敛，下面两小节再利用单位截面的紧性，得到对初值一致的收敛速度。

6.5 状态锥与映射的连续性

全局收敛还不是对所有状态一致的 R-线性收敛。为了应用定理 5.1，定义由 p 个允许的步长构成的多项式集合，以及状态锥

$$\begin{aligned}\mathcal{Q}_p&=\left\{q(t)=\prod_{h=1}^{p}(1-\eta_ht):\eta_h\in[L^{-1},m^{-1}]\right\},\\ \Omega_p&=\big\{(u,v):\ \text{存在 } q\in\mathcal{Q}_p,\ \text{使 } v=q(H)u\big\}.\end{aligned} \tag{6.11}$$

与 BB 的情形一样，由步长取值范围的紧性可得 Ω_p 是闭集；它显然是锥，并且其中的状态都满足 $S(v)\subset S(u)$。在 Ω_p 上定义

$$F_p(u,v)=\begin{cases}\Psi_u(H)v, & s(u)>p,\\ 0, & s(u)\le p,\end{cases}\qquad T_p(u,v)=(v,F_p(u,v)). \tag{6.12}$$

当 $s(u)=p$ 时，行列式之比的分母仍为正，Ψ_u 在 λ_i（$i\in S(u)$）处都等于零，又 $S(v)\subset S(u)$，因此 $\Psi_u(H)v=0$，与定义一致；当 $s(u)<p$ 时，按精确秩求 Ritz 值，下一组有限步终止，所以也定义为零。

T_p 把 Ω_p 映到自身，并且是一次正齐次的。不终止时 Ritz 值都在 $[m, L]$ 中，故 $\Psi_u \in \mathcal{Q}_p$，像仍在锥中。终止时 $s(v) \le s(u) \le p$：对每个 $i \in S(v)$ 取因子 $1 - t/\lambda_i$，不足 p 个时用任意允许的因子补足，所得多项式 $q \in \mathcal{Q}_p$ 满足 $q(H)v = 0$，即 $(v, 0) \in \Omega_p$。把 u 乘一个正数不改变 Ritz 值，零状态也满足齐次性。剩下要证明的是连续性。

命题 6.5 F_p 和 T_p 在整个 Ω_p 上连续；特别地，在 $s(u) = p$（历史矩阵满秩但迭代终止）的点和历史矩阵降秩的点处也连续。

证明 设 $(u_\ell, v_\ell) \to (u, v)$。把序列分成两个子列：满足 $s(u_\ell) > p$（不终止）的项和满足 $s(u_\ell) \le p$（终止）的项；后者的 F_p 恒为零。对前者总有 $\|\Psi_{u_\ell}(H)\| \le B^p$。下面分四种情形讨论。

若 $s(u) > p$，则 ℓ 充分大时 $s(u_\ell) > p$，且 $K_p(u)^{\mathsf{T}} H K_p(u)$ 正定。行列式之比在 u 附近连续，于是 $F_p(u_\ell, v_\ell) \to F_p(u, v)$。

若 $s(u) = p$，极限处的分母仍为正。对 $i \in S(u)$，$\Psi_u(\lambda_i) = 0$，由行列式之比的连续性，$F_p(u_\ell, v_\ell)$ 在这些特征值上的分量趋于零；对 $i \notin S(u)$，由锥的定义 $P_i v = 0$，再由乘子的一致上界得 $\|P_i F_p(u_\ell, v_\ell)\| \le B^p \|P_i v_\ell\| \to 0$。因此在两个子列上 F_p 都趋于零。

若 $1 \le s(u) < p$，不在 $S(u)$ 中的分量仍由上面的上界控制。固定 $r \in S(u)$。凡不含 r 的 p 元子集 I，因 $s(u) < p$，必含某个 $h \notin S(u)$。令 $J = (I \setminus \{h\}) \cup \{r\}$。当 ℓ 充分大时 $P_r u_\ell \neq 0$；当 $\beta_I(u_\ell) > 0$ 时 $\beta_J(u_\ell) > 0$，从而

$$0 \le \pi_I(u_\ell) \le \frac{\beta_I(u_\ell)}{\beta_J(u_\ell)} = \frac{\Delta_I^2}{\Delta_J^2} \frac{\lambda_h}{\lambda_r} \frac{\|P_h u_\ell\|^2}{\|P_r u_\ell\|^2} \longrightarrow 0.$$

权为零的项不影响估计，含 r 的项又恰为零。对有限多项求和，便得 $\Psi_{u_\ell}(\lambda_r) \to 0$；因 $P_r v_\ell$ 有界，F_p 在 λ_r 上的分量也趋于零。

最后，若 $u = 0$，由锥的定义 $v = 0$；由一致的上界得 $\|F_p(u_\ell, v_\ell)\| \le B^p \|v_\ell\| \to 0$。四种情形都得到了所需的极限，且两个子列趋于同一个值，所以整个序列收敛到 $F_p(u, v)$。T_p 的第一个分量是 v，故 T_p 也连续。 □

同样的权之比估计在这里又用了一次：前面用它控制归纳中可能造成放大的项，这里用它控制降秩极限中不含 r 的项。证明映射连续，并不需要单个 Ritz 值在维数变化的历史子空间上有一致的下界，而原证明恰恰在这一点上缺少依据。

满秩而终止的情形和真正秩亏的情形必须分开处理：前者的分母仍为正，可以直接对比值取极限；后者的分母可能趋于零，只看分子趋于零推不出比值趋于零，必须成对地比较权。原点处则不必估计任何比值，由锥的定义，当前的组首梯度也一起趋于零。边界情形需要这样分别论证，不能以一句“由 Ritz 值的连续性可知”带过。

6.6 R-线性收敛

还要验证：以 Ω_p 中任一状态为初始状态时，定理 6.4 都适用。给定 $(u_0, u_1) \in \Omega_p$，取满足 $u_1 = q(H)u_0$ 的步长 $\eta_1, \dots, \eta_p$，作为 LMSD 初始组的实际步长。若中途梯度为零，则 $u_1 = 0$，$F_p(u_0, u_1)$ 也为零；否则这一组的历史矩阵恰好张成 $\mathcal{K}_p(H, u_0)$。若 $s(u_0) > p$，求出的 Ritz 值正是 Ψ_{u_0} 的根；若 $s(u_0) \le p$，下一组终止，正好对应定义中的零。逐组重复，实际迭代得到的组首梯度与 T_p 的轨道重合（有限步终止后都记为零）。因此

$$T_p^j(u_0, u_1) \to (0, 0) \qquad \text{对所有} (u_0, u_1) \in \Omega_p.$$

定理 6.6（一致的 R-线性收敛） 设 $H \succ 0$，记忆长度 p 固定，则存在只依赖于 H 和 p 的常数 $C_* \geq 1$ 和 $\rho_* \in (0,1)$，使得

$$\left\| T_p^j Z \right\|_\times \leq C_* \rho_*^j \left\| Z \right\|_\times, \qquad Z \in \Omega_p, \quad j \geq 0, \tag{6.13}$$

其中 j 是组的序号；有限步终止的情形（终止后记为零）也包括在内。

证明 Ω_p 是闭锥，T_p 是它到自身的连续、一次正齐次映射，并且由定理前的论证，从每个状态出发的轨道都趋于零。由定理 5.1 即得结论。 □

权矩阵取为固定的 $W = \omega(H) \succ 0$（与 H 可交换）时，改用

$$Q_j^{\mathsf{T}} W H Q_j z = \theta Q_j^{\mathsf{T}} W Q_j z$$

求 Ritz 值。设矩阵 A 的各列构成历史子空间的一个基，则

$$(W^{1/2}A)^{\mathsf{T}} H (W^{1/2}A) = A^{\mathsf{T}} W H A, \qquad (W^{1/2}A)^{\mathsf{T}} (W^{1/2}A) = A^{\mathsf{T}} W A.$$

所以加权的 Ritz 值等于同一个 H 在变换后的历史矩阵 $W^{1/2}G_j$ 上的普通 Ritz 值。由于 $W^{1/2}$ 与 H 可交换，这一变换不改变历史矩阵的秩、下标集 $S(\cdot)$、每组的步数和终止的时刻。整条加权的轨道与不加权的轨道共轭，式 (6.13) 的右端只需再乘以 $\sqrt{\kappa_2(W)}$。$W = I$ 对应普通 Ritz 值，$W = H$ 对应调和 Ritz 值。

当 $p = 1$ 时，这个加权特征值问题只有一个特征值

$$\theta = \frac{u^{\mathsf{T}} W H u}{u^{\mathsf{T}} W u},$$

其倒数正是 $\eta_W(u)$；Ω_1 也就是第 5 节的 Ω_{BB}。可见 BB 与 LMSD 的联系不只是证明的形式相似：记忆长度为 1 时，两者的算法和状态模型完全相同。第 5 节借助已有的 BB 全局收敛定理，本节则从一组步长的代数结构出发，对所有 p 重新证明了全局收敛，两条证明路线互不依赖。

初始组的步长可以是任意正数，所以 $(g_{0,0}, g_{1,0})$ 不一定属于 Ω_p。若迭代不终止，应从第 1 组（即初始组之后、第一个以 Ritz 值倒数为步长的组）开始估计，即取

$$(u_0, u_1) = (g_{1,0}, g_{2,0});$$

只有初始组的各步长都在 $[L^{-1}, m^{-1}]$ 中时，才可以直接从初始组算起。这样，从第 1 组起有对初值一致的收敛速度，而对任意的初始步长，全局收敛的结论仍然成立。

为了把按组计的估计换算成按迭代步计的估计，从第 1 组开始重新编号：

$$\widehat{g}_{jp+r} = g_{j+1,r}, \qquad 0 \leq r \leq p-1.$$

不终止时每组恰有 p 步，于是

$$\left\| \widehat{g}_{jp+r} \right\| \leq B^p C_* \sqrt{\kappa_2(W)}\, \rho_*^j \left\| (u_0, u_1) \right\|_\times .$$

令 $\widehat{\rho}_* = \rho_*^{1/p}$。由于 $\rho_*^j = \widehat{\rho}_*^{jp+r-r} \leq \widehat{\rho}_*^{-(p-1)} \widehat{\rho}_*^{jp+r}$，可取

$$\widehat{C}_W = B^p C_* \sqrt{\kappa_2(W)}\, \widehat{\rho}_*^{-(p-1)},$$

使对所有 $\ell \ge 0$ 有

$$\|\widehat{g}_\ell\| \le \widehat{C}_W \widehat{\rho}_*^{\,\ell}\, \|(u_0,u_1)\|_\times , \tag{6.14}$$

$$\|\widehat{x}_\ell - x_*\| \le \frac{\widehat{C}_W}{m} \widehat{\rho}_*^{\,\ell}\, \|(u_0,u_1)\|_\times ,$$

$$F(\widehat{x}_\ell) \le \frac{\widehat{C}_W^2}{2m} (\widehat{\rho}_*^{\,2})^{\ell}\, \|(u_0,u_1)\|_\times^2 . \tag{6.15}$$

有限步终止的迭代当然也是 R-线性收敛的（常数随具体的迭代过程而定）；上面统一的换算用到的是“不终止时每组恰好 p 步”这一事实。

这里只证明了存在小于 1 的一致几何因子，并没有给出 $p>1$ 时最优因子的显式值，也不能把第 7 节 BB 方法的 c_H 直接搬到 LMSD 上。按数值阈值删列、迭代中改变权或记忆长度、只选用部分 Ritz 值、重复使用步长等变体，都需要分别检查状态的选取和每条轨道的收敛性。研究乘子多项式而不是单个 Ritz 值，保留了可以精确分析的多项式结构：全局收敛来自凸组合表示和按特征值的归纳；一致的收敛速度则来自映射的连续性、齐次性、单位截面的紧性，以及锥中每个状态都可由算法实际达到这一事实。

6.7 一个数值例子：满秩的历史矩阵也可能非常病态

定理 6.6 不需要 Curtis 与 Guo 的归一化条件。那么，这个条件中的常数在实际迭代中有多大？下面复述 LMSD 预印本 [35] 数值部分的一个例子，没有另做计算。取 60 阶对角矩阵 H，它的 20 个互异特征值是区间 $[1,2]$、$[10,15]$、$[50,60]$、$[200,250]$、$[800,1000]$ 中各 4 个等距点（含端点），每个重复 3 次；初始梯度的分量为

$$(g_0)_i = 1 + \sin\!\big(0.37(i+7)\big) + 0.3\cos\!\big(1.17(i+7)\big) + 0.5\sin\!\big(0.11(i+7)\big), \qquad i = 1,\dots,60,$$

它在每个特征子空间上的投影都不为零。取 $p=5$，第 0 组用精确线搜索步长，权取 $W = H^a$，$a = -1, -0.5, \dots, 2$，每组按 Ritz 值从大到小的顺序使用其倒数。计算在双精度下进行：历史矩阵的数值秩由奇异值按相对阈值 10^{-12} 判定，迭代到 $\|g\| / \|g_0\| \le 10^{-11}$ 为止。记

$$\xi_j = \frac{\|g_{j,0}\|}{\sigma_{\min}(G_j)}, \qquad A_j = \max_{0\le r\le \ell_j} \frac{\|g_{j,r}\|}{\|g_{j,0}\|}. \tag{6.16}$$

由 $G_j = Q_j R_j$ 有 $\sigma_{\min}(G_j) = \big\|R_j^{-1}\big\|^{-1}$，所以 ξ_j 就是 Curtis 与 Guo 的条件要求一致有界的量（预印本中记作 χ_k）；A_j 表示一组之内梯度范数最多被放大了多少倍。结果见表 2 和图 2。

表 2　不同权下 LMSD 的迭代次数以及 ξ_j、A_j 的最大值，数据取自文献 [35]。迭代次数包括第 0 组的 5 步。

权 W	迭代次数	$\max_j \xi_j$	$\max_j A_j$
H^{-1}	276	1.30×10^{8}	1.33×10^{3}
$H^{-1/2}$	231	2.35×10^{8}	1.55×10^{5}
I	269	2.51×10^{7}	2.26×10^{5}
$H^{1/2}$	265	1.59×10^{7}	6.93×10^{4}
H	280	5.59×10^{6}	3.30×10^{4}
$H^{3/2}$	270	1.99×10^{7}	2.21×10^{4}
H^{2}	264	1.48×10^{7}	6.99×10^{1}

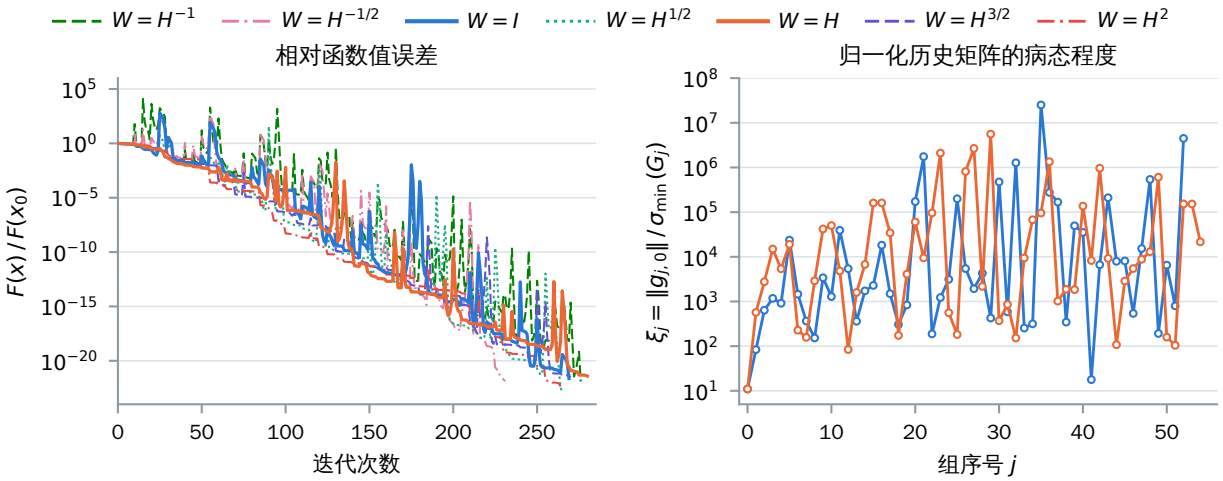


图 2　上述算例的计算结果，按文献 [35] 原图的数据重绘。左：七种权下相对函数值误差 $F(x)/F(x_0)$ 随迭代次数的变化（迭代次数包括第 0 组的 5 步）；右：$W = I$ 与 $W = H$ 时各组历史矩阵的 ξ_j，第 0 组是采用精确线搜索步长的初始组。

七种权下的迭代都在 231 到 280 步之内满足停止准则，检测到的历史矩阵数值秩始终是 5，数值阈值没有删去任何一列。然而 ξ_j 的最大值在 5.59×10^6 到 2.35×10^8 之间，$W = I$ 与 $W = H$ 时还在组与组之间剧烈起伏（图 2 右）。一组之内的放大也不小：A_j 最大为 2.26×10^5（$W = I$），不过远低于粗略的上界 $B^p = 999^5 \approx 10^{15}$。函数值误差也多次大幅反弹，$W = H^{-1}$ 时在第 15 步一度超过初值的 10^4 倍（图 2 左）；尽管如此，各条曲线总体上仍是几何下降的。

这些数据说明，每组的历史矩阵都满秩，并不意味着它们远离秩亏。在这个只有 60 维的问题上，对采用普通 Ritz 值（$W = I$）和调和 Ritz 值（$W = H$）的两种版本，要使 Curtis 与 Guo 的条件成立，常数分别至少要取 2.51×10^7 和 5.59×10^6；Curtis 与 Guo 的数值试验中，同一比值的最大值约为 2×10^{16} [6]，相当于 $\sigma_{\min}(G_j)/\|g_{j,0}\| \approx 5 \times 10^{-17}$，已经小于双精度的机器精度。有限步的试验不能判定 ξ_j 沿整条轨道是否有界，因此这并不构成反例；它只说明这个常数可能很大，而且无法事先估计。定理 6.6 及其加权形式不含这个常数：C_*、ρ_* 只依赖于 H 与 p，权只以因子 $\sqrt{\kappa_2(W)}$ 的形式出现在常数中；代价是这些常数由紧性得到，没有显式的值。

7　BB 方法的最坏渐近收敛因子

7.1　问题与结论

已有的全局收敛和 R-线性收敛理论，回答了 BB 迭代是否收敛到解、是否有几何收敛速度的上界，却没有确定最坏的轨道究竟有多慢。对非单调方法来说，这一区别尤其重要：某些步的梯度范数会明显上升，只用单步估计，往往得不到精确的渐近常数。Li 与 Sun 在首步步长也按 BB 公式计算的条件下，证明了 BB1 各分量有因子为 $1 - 1/\kappa$ 的上界，并给出了因子为 $(\kappa-1)/(\kappa+1)$ 的平衡轨道 [19]，上下界之间仍有差距。我们关于最坏收敛因子的预印本 [37] 证明了：所有轨道的渐近收敛因子都不超过 $(\kappa-1)/(\kappa+1)$，即没有比这条平衡轨道更慢的轨道。

逐个分量估计的方法能有效地控制各分量的增减，足以证明收敛并给出上界；但要得到精确的常数，就必须利用两个端点分量之间的联系。一个分量增大时，另一个分量未必同时增大；把它们分别放大再相加，就失去了这种制约。下面的状态和分解式保留了这种联系，并把长时间的增长问题化为紧空间上的平均增长问题。

以下仍考虑严格凸二次函数

$$f(x)=\frac{1}{2}x^{\mathsf T}Hx-b_0^{\mathsf T}x,\qquad H=H^{\mathsf T}\succ 0,\qquad g_k=H(x_k-x_*). \tag{7.1}$$

记 $m=\lambda_{\min}(H)$、$L=\lambda_{\max}(H)$，$\kappa=L/m$。用 P_λ 表示特征值 λ 的特征子空间上的正交投影，记使 $P_\lambda g_0\neq 0$ 的特征值 λ 组成的集合为

$$\Lambda(g_0)=\{\lambda : P_\lambda g_0\neq 0\},\qquad a=\min\Lambda(g_0),\quad b=\max\Lambda(g_0).$$

迭代不会产生新的分量，即 $\Lambda(g_k)\subset\Lambda(g_0)$。若 $a<b$，令 $c_{a,b}=(b-a)/(b+a)$；$\Lambda(g_0)$ 只含一个点时，约定这个常数为 0。对整个谱，记

$$c_H=\frac{L-m}{L+m}=\frac{\kappa-1}{\kappa+1}. \tag{7.2}$$

以下考虑全程使用 BB1 或全程使用 BB2 的迭代：首步步长 $\eta_0>0$ 可以任意选取，此后由前一个梯度计算步长；有限步终止后梯度记为零。

渐近收敛因子 $\rho_g(g_0,\eta_0)$ 的定义见式 (1.2)，它不受前面有限多步的影响；$\|g_k\|\le C\gamma^k\|g_0\|$ 只有在 C 与初值无关时，才是对初值一致的上界（见第 1 节）。

定理 7.1（有限维情形 BB 方法的最坏渐近收敛因子） 设 $g_0\neq 0$，首步步长 $\eta_0>0$，迭代全程使用 *BB1* 或全程使用 *BB2*，则

$$\rho_g(g_0,\eta_0)\le c_{a,b}\le c_H.$$

若 $m<L$，渐近收敛因子关于所有初始梯度和正的首步步长的上确界恰为 c_H，并且首步步长按 *BB* 公式计算的平衡轨道达到这个值。对每个 $\gamma\in(c_H,1)$，当首步步长按 *BB* 公式计算时，存在与初值无关的常数 $C_\gamma(H)$，使得

$$\|g_k\|\le C_\gamma(H)\gamma^k\|g_0\|\qquad(k\ge 0). \tag{7.3}$$

对 $\gamma<c_H$，不存在这样的一致上界。

这里“首步步长按 BB 公式计算”，对 BB1 是指 $\eta_0=\|g_0\|^2/(g_0^{\mathsf T}Hg_0)$，对 BB2 是指 $\eta_0=(g_0^{\mathsf T}Hg_0)/(g_0^{\mathsf T}H^2g_0)$。定理确定的是一致上界中因子的下确界，并没有证明 $\gamma=c_H$ 时也有一致的常数。对任意正的首步步长，渐近收敛因子的结论同样成立；但与第 5 节一样，第一步就可能把梯度任意放大，以 $\|g_0\|$ 为基准的常数不可能与首步步长无关。$\Lambda(g_0)$ 只含一个点时，首步之后至多再迭代一步就精确终止，需要单独处理。

下界由一条具体的轨道给出。对 BB1，取只在特征值 m 和 L 上有分量的初始梯度，且

$$\|P_mg_0\|^2=\|P_Lg_0\|^2,\qquad \eta_0=\frac{2}{m+L}.$$

较小特征值上的分量每步乘以 c_H，较大特征值上的分量每步乘以 $-c_H$，两个分量的范数始终相等，因此 $\|g_k\|=c_H^k\|g_0\|$。这就是 BB 原文二维分析中的平衡轨道（见第 3 节），Li 与 Sun 给出的例子也是这样的轨道 [5, 19]。新结果是与之相匹配、对所有轨道都成立的上界。虽然 BB 方法的最坏收敛因子与最优固定步长梯度法相同，但在许多其他轨道上 BB 方法可能快得多；最坏情形的比较不能代替典型表现的比较。

对 BB2，相应的平衡条件是 $m\|P_mg_0\|^2=L\|P_Lg_0\|^2$，首步步长同样取 $2/(m+L)$。若 $\kappa>1$，则 $c_H<1-1/\kappa$；当 $\kappa\to\infty$ 时，$c_H=1-2/\kappa+O(\kappa^{-2})$。这个改进是就梯度范数的最坏渐近收敛因子而言的，函数值误差相应地取平方。对具体问题达到给定精度所需的迭代次数，它并不能给出完整的预测，因为常数和开始阶段的过渡过程也会影响有限步的表现。

7.2 分量比例与端点不等式

以 BB1 为例，其迭代为

$$g_{k+1} = (I - \eta_k H) g_k, \qquad \eta_k = \frac{\|g_{k-1}\|^2}{g_{k-1}^{\mathsf{T}} H g_{k-1}} \quad (k \ge 1). \tag{7.4}$$

设梯度只在下标集 I 对应的特征值上有分量，I 至少含两个下标。定义各分量所占的比例

$$w_i^k = \frac{\|P_{\lambda_i} g_k\|^2}{\|g_k\|^2}, \qquad \mathcal{P}_I = \Big\{ w : \ w_i \ge 0,\ \textstyle\sum_{i\in I} w_i = 1 \Big\}.$$

当前的比例 w_k 给出各分量的相对大小，前一步的比例 w_{k-1} 决定这一步的步长，因此取 $\chi_k = (w_k, w_{k-1})$ 为状态，χ_{k+1} 就只依赖于 χ_k。对 $\chi = (w, z) \in X_I := \mathcal{P}_I \times \mathcal{P}_I$，令

$$\begin{aligned} u(z) &= \sum_{i\in I} \lambda_i z_i, & \phi_i(z) &= \left(1 - \frac{\lambda_i}{u(z)}\right)^2, \\ r(\chi)^2 &= \sum_{i\in I} \phi_i(z) w_i, & T(w,z) &= \left(\frac{\phi(z) \odot w}{r(\chi)^2}, w \right). \end{aligned} \tag{7.5}$$

当 $r(\chi_k) > 0$ 时，$\chi_{k+1} = T\chi_k$，且 $\|g_{k+1}\| = r(\chi_k)\,\|g_k\|$。于是范数之比是各步因子 $r(\chi_i)$ 的乘积，取对数后化为 $\log r$ 沿轨道的和。首步步长任取时，这一描述从 $\chi_1 = (w_1, w_0)$ 开始适用；首步步长按 BB 公式计算时，也可以等价地从 $\chi_0 = (w_0, w_0)$ 开始。分析在整个 X_I 上进行，其中也包括算法实际上达不到的状态。

这个离散模型精确地保留了 BB 方法的步长滞后，以及每一步各分量的乘子。Akaike 以及 Pronzato 等人对使用当前梯度的方法所做的概率分布分析，是重要的先例 [1, 24]；但只用当前的分布描述不了 BB 方法，状态中多出的那个分布，记录的就是上一步的信息。

先看决定 $c_{a,b}$ 的一个初等不等式。若 $0 < a < b$，$u \in [a, b]$，则

$$\left(\frac{u-a}{u}\right)^{\frac{b}{a+b}} \left(\frac{b-u}{u}\right)^{\frac{a}{a+b}} \le \frac{b-a}{b+a}, \tag{7.6}$$

等号当且仅当 $u = (a+b)/2$ 时成立。对 $u \in (a, b)$，令 $t = (u-a)/(b-u)$，则左边等于

$$\frac{(b-a)\, t^{b/(a+b)}}{a + bt},$$

由加权算术–几何平均不等式 $a + bt \ge (a+b) t^{b/(a+b)}$ 即得结论。以 $1/u$ 为步长时，左边的两个底数恰是两个端点分量的乘子的绝对值，所以这个不等式刻画了两个端点分量此消彼长的关系。

考虑这样的状态 (w, z)：存在下标子集 $J \subset I$，使相邻两步的比例都恰好以 J 为支撑，即 $i \in J$ 时 $w_i, z_i > 0$，$i \notin J$ 时 $w_i = z_i = 0$。记 J 对应的最小、最大特征值为 $a < b$，令

$$\begin{aligned} \beta &= \frac{b}{a+b}, \qquad 1 - \beta = \frac{a}{a+b}, \\ h(w,z) &= -\frac{\beta}{2} \log w_a - \frac{1-\beta}{2} \log w_b, \\ D(w,z) &= \log c_{a,b} - \beta \log \frac{u(z) - a}{u(z)} - (1-\beta) \log \frac{b - u(z)}{u(z)}. \end{aligned} \tag{7.7}$$

这时 $u(z) \in (a, b)$，由式 (7.6) 知 $D \ge 0$。由两个端点分量的精确更新式，直接验证可得

$$\log r = \log c_{a,b} - D + h \circ T - h. \tag{7.8}$$

式 (7.8) 把可正可负的一步对数增长 $\log r$ 分成三部分：临界值 $\log c_{a,b}$、一个非正的项 $-D$ 和一个差分项。沿周期轨道求和时差分项相互抵消，立即得到一个周期内的平均不超过 $\log c_{a,b}$。但一般的轨道不一定是周期的，而且端点分量的比例趋于零时 h 无界。不能只凭求和时中间各项相消，就断言 $[h(T^n\chi)-h(\chi)]/n\to 0$；也不能在不知道可积性的情况下直接对差分项积分。下面分两步解决这个问题。

7.3 从不变测度到每一条轨道

首先要处理有限步终止带来的奇异性。记

$$\Sigma_I=\{\chi\in X_I:\ r(\chi)=0\},$$

在 Σ_I 上梯度已经变为零，式 (7.5) 中归一化的分母为零。添加一个孤立点 † 表示迭代已经终止，令 Σ_I 和 † 都映到 †，就得到紧度量空间 $\widehat{X}_I=X_I\cup\{\dagger\}$ 上的 Borel 可测映射 $\widehat{T}$。定义

$$\varphi(\chi)=\log r(\chi)\quad(\chi\notin\Sigma_I\cup\{\dagger\}),\qquad \varphi=-\infty\quad\text{在 }\Sigma_I\cup\{\dagger\}\text{ 上.}$$

函数 φ 上半连续且有上界；映射 $\widehat{T}$ 则只在奇异集之外连续。

第一步是证明：对每个使 $\int\varphi\,d\nu$ 有限的 $\widehat{T}$-不变概率测度 ν，都有

$$\int\varphi\,d\nu\le\log c_I,\qquad c_I=\frac{\max_{i\in I}\lambda_i-\min_{i\in I}\lambda_i}{\max_{i\in I}\lambda_i+\min_{i\in I}\lambda_i}.\tag{7.9}$$

积分有限，说明孤立点 † 的测度为零；再由不变性和 $\widehat{T}^{-1}\{\dagger\}=\Sigma_I\cup\{\dagger\}$，奇异集的测度也为零。由遍历分解，只需考虑遍历测度。取最小的下标集 J，使测度集中在 X_J 上。一个分量一旦变为零，此后就始终为零，状态落在更小的下标集上；相应的正向不变集在遍历测度下的测度只能是 0 或 1，由 J 的最小性，不能是 1。因此对 $i\in J$，$w_i>0$ 几乎处处成立。又因 T 的第二个分量就是原来的第一个分量，不变测度的两个边缘分布相同，所以 z_i 也几乎处处为正。只含一个特征值时迭代立即终止，因此这里的 J 至少含两个下标。这样，每个遍历测度都对应一个固定的下标集，其上的分量几乎处处始终不为零，分解式 (7.8) 可以在其上使用。

这里“始终不为零”是对遍历测度几乎处处成立的性质，并不是说任意一条轨道上非零分量的集合都永不改变。还要构造一个正向不变的全测度集：取上述全测度集与它在 $\widehat{T}$ 下各次原像的可数交即可，这样求和时每一步都能使用同一个下标集上的分解式。于是，边界上的问题就化成了固定下标集内部的问题。

Birkhoff 遍历定理和 Poincaré 回复定理的结论都对几乎所有的点成立。取一点 χ，使两者的结论在 χ 处都成立，并且它的整条正向轨道都在上述全测度集中。由前者得

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi(\widehat{T}^k\chi)\longrightarrow\int\varphi\,d\nu;$$

由后者得一列回复时刻 n_ℓ，使 $\widehat{T}^{n_\ell}\chi\to\chi$。由于 h 在相应集合的相对内部连续且有限，$h(\widehat{T}^{n_\ell}\chi)\to h(\chi)$。沿这列时刻对式 (7.8) 求和，得

$$\frac{1}{n_\ell}\sum_{k=0}^{n_\ell-1}\varphi(\widehat{T}^k\chi)=\log c_J-\frac{1}{n_\ell}\sum_{k=0}^{n_\ell-1}D(\widehat{T}^k\chi)+\frac{h(\widehat{T}^{n_\ell}\chi)-h(\chi)}{n_\ell}.$$

最后一项趋于零，D 非负，所以平均值不超过 $\log c_J\le\log c_I$。这里 Birkhoff 定理给出时间平均，回复定理消去无界的差分项，并不要求 h 可积。由于 φ 有上界，经过截断，可以把这个结论从遍历测度推广到一般的、积分有限的不变测度。

这里也没有假设实际的 BB 轨道都回复：回复点是按不变测度选出来的，对所有初值的控制要在下一步得到。

然而，对不变测度几乎处处成立的结论，还不能推出对每个初值都成立。第二步需要一个带奇异集的半一致遍历定理。它与连续映射的半一致遍历定理 [30] 有联系，但需要单独处理 Borel 可测映射和取值为 $-\infty$ 的函数 [37]。

命题 7.2（带奇异集的半一致估计） 设 X 为紧度量空间，$T : X \to X$ 为 *Borel* 可测映射，*Borel* 集 S 包含 T 的所有不连续点。设 $\varphi : X \to [-\infty, \infty)$ 上半连续、有上界，且在 S 上等于 $-\infty$。若对每个满足 $\int \varphi \, d\mu > -\infty$ 的 T-不变概率测度 μ，都有 $\int \varphi \, d\mu \le \ell_0$，则对每个 $\varepsilon > 0$，存在 $B_\varepsilon < \infty$，使得对任意 $x \in X$ 和 $n \ge 1$，只要 $x, Tx, \dots, T^{n-1}x$ 都不在 S 中，就有

$$\sum_{k=0}^{n-1} \varphi(T^k x) \le n(\ell_0 + \varepsilon) + B_\varepsilon. \tag{7.10}$$

证明 反设结论不成立，则可以选出轨道段 (x_j, n_j)，使左边减去 $n_j(\ell_0+\varepsilon)$ 大于 j。由于 φ 有上界，$n_j \to \infty$。由紧性，经验测度 μ_j 有弱收敛的子列，不妨设

$$\mu_j = \frac{1}{n_j} \sum_{k=0}^{n_j-1} \delta_{T^k x_j} \implies \mu.$$

此时还不能断言 μ 是不变测度，因为 T 可能不连续。先令 $\varphi_N = \max\{\varphi, -N\}$。由 Portmanteau 定理得

$$\int \varphi_N \, d\mu \ge \limsup_j \int \varphi_N \, d\mu_j \ge \ell_0 + \varepsilon.$$

取 $U = \max\{0, \sup_X \varphi\}$，对 $U - \varphi_N$ 用单调收敛定理，令 $N \to \infty$，得 $\int \varphi \, d\mu \ge \ell_0 + \varepsilon > -\infty$。由于 φ 在 S 上为 $-\infty$，必有 $\mu(S) = 0$。

于是对任意连续函数 q，$q \circ T$ 有界且 μ 几乎处处连续，因此 $\int q \circ T \, d\mu_j \to \int q \circ T \, d\mu$。同时

$$\int (q \circ T - q) \, d\mu_j = \frac{q(T^{n_j} x_j) - q(x_j)}{n_j} \longrightarrow 0.$$

于是 $\int q \circ T \, d\mu = \int q \, d\mu$，即 μ 是不变测度；但 $\int \varphi \, d\mu \ge \ell_0 + \varepsilon > \ell_0$，矛盾。 $\square$

证明中的次序不能颠倒：先由积分有限且大于 ℓ_0 推出 $\mu(S) = 0$，再由此推出 μ 是不变测度。把这个命题用于 $\widehat{X}_I$，取 $S = \Sigma_I \cup \{\dagger\}$、$\ell_0 = \log c_I$，再取指数，就得到：对任意不终止的轨道段，乘子满足

$$\prod_{k=0}^{n-1} r(\widehat{T}^k \chi) \le e^{B_\varepsilon} (c_I e^{\varepsilon})^n.$$

对任意的首步步长，先把第一步的放大倍数并入常数，再令 $\varepsilon \downarrow 0$，即得渐近收敛因子不超过 $c_{a,b}$。首步步长按 BB 公式计算时，可以直接对紧状态空间中的所有初值使用这个上界，得到式 (7.3)。因此最后的结论对每一条 BB 轨道都成立，而不只是对某个不变测度几乎处处成立。

等号的情形还给出了更细的结构。对达到 $\log c_I$ 的遍历不变测度，要使上面各不等式都取等号，必须满足两点：始终不为零的分量包含 I 的两个端点，并且非负项 D 几乎处处为零。具体地说，选点时还要求 Birkhoff 定理的结论对可数多个函数 $\min\{D, N\}$（$N = 1, 2, \dots$）也成立；沿回复时刻，D 的平均只能趋于零，所以每个截断后的积分都是零，令 $N \to \infty$ 即得 $D = 0$ 几乎处处成立。于是 $u(z)$ 几乎处处等于两个端点的中点；由于两个边缘分布相同，$u(w)$ 也是如此。这时两

个端点上的 ϕ_i（即乘子的平方）都等于 c_1^2，每个中间特征值上的 ϕ_i 却严格更小。平均增长已经达到最大值，这就迫使中间分量的比例为零，两个端点的比例各为 $1/2$。

再由遍历分解可知，使平均增长达到最大的不变测度是唯一的，即平衡状态（两个端点各占一半）上的 Dirac 测度 [37]。这从不变测度的角度说明了平衡轨道的特殊地位。不过，这样的测度唯一，并不等于对所有达到最坏因子的轨道作了完整的分类；也不能由此推出，达到最坏因子的轨道一开始就处在这个平衡状态。

BB2 和其他权函数固定且为正的步长规则，可以用变量替换来处理：若

$$\eta_k^{\omega} = \frac{g_{k-1}^{\mathsf{T}}\omega(H)g_{k-1}}{g_{k-1}^{\mathsf{T}}\omega(H)Hg_{k-1}}, \qquad \omega(\lambda) > 0,$$

令 $\widetilde{g}_k = \omega(H)^{1/2}g_k$，就得到同一个 H 上的 BB1。范数等价常数开 k 次方后趋于 1，因此渐近收敛因子的结论不变；BB2 对应 $\omega(\lambda) = \lambda$。

8 BB 方法的 Lyapunov 函数

8.1 构造

式 (7.8) 中的 h 不只是证明收敛速度时的辅助函数。把归一化的比例与梯度的范数重新合在一起，有

$$\|g\|\, e^{-h} = \|g\|\, w_a^{\beta/2} w_b^{(1-\beta)/2} = \|P_a g\|^{\beta}\, \|P_b g\|^{1-\beta}\,. \tag{8.1}$$

于是差分项可以并入一个只依赖于当前点的显式函数。我们随后关于 Lyapunov 函数的预印本 [36] 单独展开了这一想法，并用它对谱为有限集的情形给出了 R-线性收敛的另一个证明。

本节采用与前文一致的编号：g_0 为初始梯度，g_1 为首步之后的梯度，从 $k = 1$ 起每一步的步长都由 g_{k-1} 算出。记 $\Lambda = \Lambda(g_0)$，设它至少含两个点，且 g_1 在端点 a、b 上的分量仍不为零。对 BB1、BB2 和其他权函数固定且为正的步长规则，步长的倒数 $\tau_k = 1/\eta_k$ 都是 $\Lambda(g_{k-1})$ 中各特征值的加权平均，权都为正。由于两个端点上的分量都不为零，$a < \tau_k < b$，从而

$$P_a g_{k+1} = \left(1 - \frac{a}{\tau_k}\right) P_a g_k \neq 0, \qquad P_b g_{k+1} = \left(1 - \frac{b}{\tau_k}\right) P_b g_k \neq 0.$$

由归纳法，首步之后两个端点上的分量一直不为零；中间的分量仍可能被消去。所以端点及相应的投影可以固定下来，不必每一步重新定义函数。

定义

$$V(x) = \|P_a \nabla f(x)\|^{\frac{2b}{a+b}}\, \|P_b \nabla f(x)\|^{\frac{2a}{a+b}}\,. \tag{8.2}$$

它是误差 $x - x_*$ 的二次正齐次函数。记

$$D_{a,b}(u) = \log c_{a,b} - \beta \log \frac{u-a}{u} - (1-\beta)\log \frac{b-u}{u} \geq 0.$$

把两个端点分量的更新式直接代入，得到精确的下降关系

$$\begin{aligned} V(x_{k+1}) &= \left(\frac{\tau_k - a}{\tau_k}\right)^{2\beta} \left(\frac{b-\tau_k}{\tau_k}\right)^{2(1-\beta)} V(x_k) \\ &= c_{a,b}^2 e^{-2D_{a,b}(\tau_k)} V(x_k) \leq c_{a,b}^2 V(x_k), \qquad k \geq 1. \end{aligned} \tag{8.3}$$

由于首步之后两个端点分量都不为零，$V(x_k)>0$，而且每一步都严格下降；函数值和梯度范数的反弹并不影响这种单调性。逐步相乘还得到

$$V(x_k)=c_{a,b}^{2(k-1)}\exp\left(-2\sum_{i=1}^{k-1}D_{a,b}(\tau_i)\right)V(x_1).$$

超出临界因子 $c_{a,b}$ 的那部分收敛速度，体现在 $D_{a,b}$ 的累积和中。

同时也要说明它的局限。V 依赖于 H 的特征投影和 g_0 的谱端点，不能在不知道谱信息时直接用作停止准则；若某个端点上的分量为零，V 就可能为零而梯度不为零，所以 V 不是正定函数，与 $\|g\|^2$ 不等价。即使沿这条轨道 V 始终为正，也不能由 $V\to 0$ 直接推出整个梯度按 $c_{a,b}$ 衰减。下降关系本身只用到步长的倒数在 (a,b) 中；BB 方法步长滞后的特点，还体现在这些步长怎样由状态决定，以及整个归一化系统怎样演化。

8.2 指数的选取

这两个指数由下面的极小化问题确定。考虑函数族

$$V_\beta(x)=\|P_a g(x)\|^{2\beta}\,\|P_b g(x)\|^{2(1-\beta)},\qquad 0<\beta<1.$$

若某一步步长的倒数为 $u\in[a,b]$，则经过这一步，$\sqrt{V_\beta}$ 变为原来的 $R_\beta(u)$ 倍，其中

$$R_\beta(u)=\left(\frac{u-a}{u}\right)^{\beta}\left(\frac{b-u}{u}\right)^{1-\beta}.$$

对 u 求导可知最大值点为

$$u_\beta^*=\frac{ab}{b(1-\beta)+a\beta},\qquad \Gamma(\beta):=\max_{u\in[a,b]}R_\beta(u)=\frac{(b-a)\beta^\beta(1-\beta)^{1-\beta}}{b^\beta a^{1-\beta}}.$$

再关于 β 求最小值：

$$\frac{d}{d\beta}\log\Gamma(\beta)=\log\frac{a\beta}{b(1-\beta)},\qquad \frac{d^2}{d\beta^2}\log\Gamma(\beta)=\frac{1}{\beta}+\frac{1}{1-\beta}>0.$$

因此使最坏的单步因子最小的指数是唯一的，即 $\beta_*=b/(a+b)$，相应的最小值为 $\Gamma(\beta_*)=c_{a,b}$。这既解释了式 (8.2) 中不对称的指数，也说明最优的单步因子恰为同一个常数 $c_{a,b}$。唯一性只限于上面这个函数族，不能推广为在所有 Lyapunov 函数中唯一最优。

如果简单地取对称的指数 $\beta=1/2$，由上面的公式得 $\Gamma(1/2)=(b-a)/(2\sqrt{ab})$，它比 $c_{a,b}$ 大，而且可能大于 1，因此不能保证每一步都下降。两个端点在乘子中的地位并不对称，选取不对称的指数，就是为了补偿这种差别。上面是对指数 β 极小化“步长倒数取遍 $[a,b]$ 时单步因子的最大值”；这并不意味着每条轨道都会反复取到最坏的步长；实际取到哪些步长，由 BB 方法的状态决定。

8.3 端点分量为零时的处理

由于 V 不是正定的，需要把使 V 为零的边界也纳入分析。在有限的下标集 I 上沿用式 (7.5)，定义有限步的乘子（终止后取零）

$$\mathcal{R}_0(\chi)=1,\qquad \mathcal{R}_{n+1}(\chi)=\begin{cases}r(\chi)\mathcal{R}_n(T\chi), & r(\chi)>0,\\ 0, & r(\chi)=0.\end{cases}$$

尽管 T 在迭代终止的状态处是奇异的，$\mathcal{R}_n$ 却在整个紧空间 X_I 上连续。这可以用归纳法证明：记 $M_I = \max\{1, \max_{X_I} r\}$，在迭代终止的状态附近

$$0 \le \mathcal{R}_{n+1}(\chi) \le M_I^n r(\chi) \longrightarrow 0;$$

在不终止的点处，则由 T 的连续性和归纳假设中 $\mathcal{R}_n$ 的连续性即得。考虑有限步乘子而不是归一化映射本身，就避开了归一化映射在迭代终止的状态处不连续的困难。

令 $\psi_I(w) = w_a^{\beta_*} w_b^{1-\beta_*}$，则 $V(x) = \|g(x)\|^2 \psi_I(w)$。对不终止的轨道段，端点不等式在边界上仍然可用，得

$$\mathcal{R}_n(\chi)^2 \psi_I(w_n) \le c_I^{2n} \psi_I(w_0). \tag{8.4}$$

对下标集归纳，要证明的是：对每个非空的有限集 I，存在 $L_I \ge 1$ 和 $q_I \in (0,1)$，使

$$\min_{1 \le \ell \le L_I} \mathcal{R}_\ell(\chi) \le q_I \qquad (\chi \in X_I). \tag{8.5}$$

也就是说，从每个状态出发，都能在某个不超过 L_I 的步数内收缩，但不要求同一个步数对所有状态都适用。

I 只含一个下标时，迭代一步就终止，归纳的起点成立。假设 I 的所有真子集都已有上述性质。考虑端点分量为零的边界 $B_I = \{(w,z) : w_a w_b = 0\}$：若当前点不终止，记 $J = \operatorname{supp} w$，则 $J \subsetneq I$，且下一个状态的两个分布都集中在 J 上。先至多放大 M_I 倍，再反复使用 J 上的有限步收缩，就能使乘子小于 $1/2$。真子集只有有限个，所以可以取一个一致的步数上界 L_B。由有限步乘子的连续性，

$$U_B = \left\{ \chi : \min_{1 \le \ell \le L_B} \mathcal{R}_\ell(\chi) < \frac{1}{2} \right\}$$

是整个 B_I 的一个开邻域。这样，端点分量为零的边界由低维情形的结论控制。

若某个初始状态 χ 使对所有 n 都有 $\mathcal{R}_n(\chi) \ge 1$，则由式 (8.4)，$\psi_I(w_n) \to 0$。由于 $X_I \setminus U_B$ 是紧的，且与 ψ_I 的零点集不相交，从某一步起轨道必定一直留在 U_B 中。此后可依次取长度不超过 L_B、乘子小于 $1/2$ 的轨道段，于是

$$\mathcal{R}_{t_j}(\chi) \le \mathcal{R}_{t_0}(\chi) 2^{-j} \longrightarrow 0,$$

矛盾。所以从每个初始状态出发，都会在某个步数内收缩。开集 $\{\mathcal{R}_n < 1\}$（$n \ge 1$）覆盖了紧空间 X_I；取有限子覆盖，令 L_I 为其中最大的 n，再由连续性和紧性，$\min_{1 \le \ell \le L_I} \mathcal{R}_\ell$ 在 X_I 上的最大值 q_I 小于 1，这就得到了式 (8.5)。

仅仅知道从每个状态出发都会在某一步收缩，还不足以得到线性收敛：如果所需步数随状态无限增大，或者收缩因子可以任意接近 1，就无法得到关于迭代次数的几何收敛速度。有限步乘子的连续性和状态空间的紧性排除了这两种情况。之后还要补上每段之内放大倍数的上界，结论才对整个序列成立，而不只是对每段的端点成立。

把这些长度不同的段接起来，到第 j 段末端，范数至多变为原来的 q_I^j 倍，每段之内至多放大 $M_I^{L_I}$ 倍，于是

$$\|g_{1+n}\| \le C_I \rho_I^n \|g_1\|, \qquad \rho_I = q_I^{1/L_I} < 1, \qquad C_I = M_I^{L_I} q_I^{-1}.$$

权函数固定且为正的情形，可以用前面的变量替换化为这一情形。这给出了首步之后整个序列的 R-线性收敛；这里的 ρ_I 由紧性得到，不能把它当成第 7 节的精确常数 c_I。两种论证的目的不同：对下标集的归纳说明，单调量退化时迭代为什么仍然收敛；遍历论证则进一步确定了所有轨道的最坏渐近收敛因子。

8.4　一个例子

取预印本中已有的二维例子 [36]：

$$H=\operatorname{diag}(1,10),\qquad x_*=0,\qquad g_0=(1,2)^{\mathsf T},\qquad \eta_0=\frac{5}{41},\qquad g_1=\Big(\frac{36}{41},-\frac{18}{41}\Big)^{\mathsf T}.$$

从 $k=1$ 起，按式 (7.4) 作 BB1 迭代。两个端点是 $a=1$、$b=10$，于是

$$V_k=|(g_k)_1|^{20/11}\,|(g_k)_2|^{2/11},\qquad \frac{V_{k+1}}{V_k}\le\frac{81}{121}.$$

图 3 画的就是预印本数表中的这条轨道。记 $F_k=f(x_k)-f(x_*)$，按本报告的编号，$F_4/F_3\approx1.49217$、$F_5/F_4\approx6.55229$，而 $F_{10}/F_9\approx80.9996$；这几步中梯度范数也上升了。与此同时，V_k 每一步都下降，这与式 (8.3) 一致。

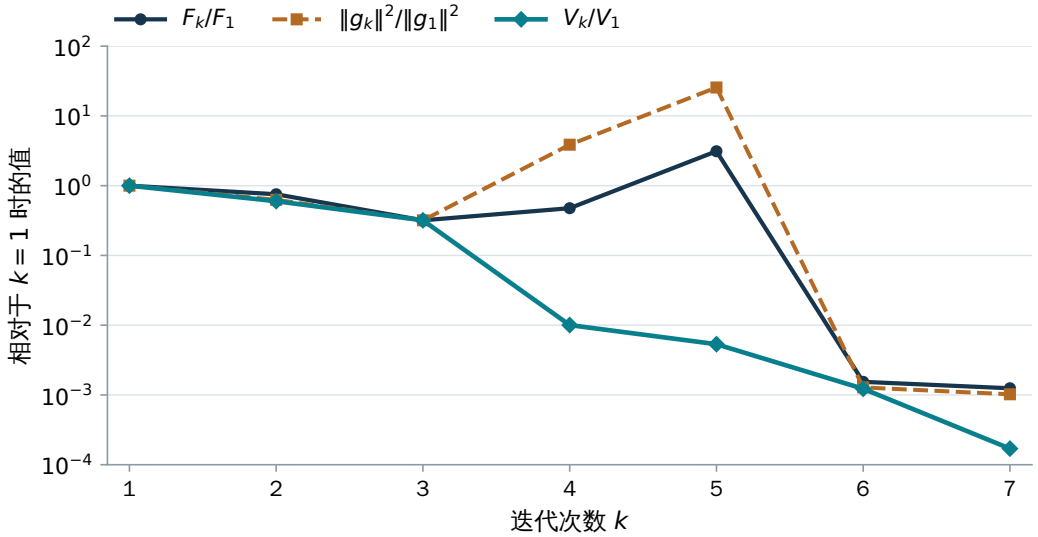


图 3　预印本中二维 BB1 例子的前七个迭代点。初值、首步步长和此后的迭代如正文所述；本报告的编号 $k=1,\dots,7$ 对应原数表的第 $0,\dots,6$ 行。三条曲线分别为 F_k/F_1、$\|g_k\|^2/\|g_1\|^2$ 和 V_k/V_1，纵轴为对数坐标。函数值误差和梯度范数都有反弹，Lyapunov 函数 V_k 逐步下降。数据按原递推式以高精度重新计算，只画出这一条轨道；全部十一行见表 3。

表 3　同一例子的全部十一行数值。行号比原表整体加一；$F_k=\frac12 g_k^{\mathsf T}H^{-1}g_k$，数值取六位有效数字。

k	F_k	$\|g_k\|$	V_k
1	0.395122	0.981688	0.679680
2	0.297661	0.776970	0.407264
3	0.125892	0.554124	0.216556
4	0.187852	1.92985	0.00682321
5	1.23086	4.96021	0.00362813
6	6.08979×10^{-4}	0.0351434	8.33213×10^{-4}
7	4.92497×10^{-4}	0.0313846	1.15605×10^{-4}
8	7.28493×10^{-6}	0.00381705	3.64247×10^{-6}
9	1.40686×10^{-11}	1.67741×10^{-5}	7.25694×10^{-23}
10	1.13955×10^{-9}	1.50967×10^{-4}	5.34610×10^{-33}
11	7.64263×10^{-36}	4.31747×10^{-18}	1.31467×10^{-35}

这个例子说明，在函数值和梯度范数反弹时，确实有一个逐步下降的量。分解式 (7.8) 中的差分项去掉归一化以后，就成为当前点的函数 V；V 的零点集又指出了证明中必须处理的边界情形。对下标集的归纳不能直接用于连续谱，这里也没有给出一般非二次函数的全局单调量；这些问题和已有推广的条件，留到下一节讨论。

9 比较、推广与展望

9.1 几种分析方法的比较

状态锥的方法和分量比例的方法，保留的是同一个算法的不同信息。前者把梯度的大小留在状态里，用连续的正齐次映射研究原点的稳定性；后者把范数分离出来，写成各步因子的乘积，在紧的归一化状态空间上研究长期平均增长率的最大值。前者由所有状态都趋于零得到某个一致的几何上界，后者进一步确定这个上界中最优的因子。第 8 节的 Lyapunov 函数则把归一化的分解式重新写成原变量上的下降关系。它们之间的关系可以概括为表 4。

表 4　本报告中的几种分析对象及其作用。

对象	主要工具	得到的结论
连续轨道与能量	时间缩放、微分方程与耗散恒等式；另需控制离散误差	加速机制、连续时间的收敛速度，以及构造离散能量的线索
状态锥	有限维带来的紧性、连续性、正齐次性；从每个状态出发都收敛	有限步内的一致收缩，R-线性收敛（常数只证明了存在）
相邻两步的分量比例	精确的迭代关系、端点不等式给出的分解式、回复定理、带奇异集的半一致遍历定理	所有 BB 轨道的最坏渐近收敛因子
Lyapunov 函数（第 8 节）	端点不等式（要求端点分量始终不为零、步长倒数在 (a,b) 中）	显式的一步下降关系；整个序列的收敛还需处理端点分量为零的情形

连续模型和离散模型都可以研究谱的作用，也都可以使用 Lyapunov 函数；两者的差别在于针对当前问题保留哪些信息、对哪种极限作估计。低分辨率微分方程研究小步长下有限时间内的轨道；高分辨率微分方程保留更高阶的修正；离散迭代本身则直接描述给定的算法及其长期行为。Wibisono、Wilson 与 Jordan 以及 Wilson、Recht 与 Jordan 的研究也表明，从连续时间的机制出发设计离散算法时，离散能量的性质仍需单独证明 [32, 33]。

9.2 Hilbert 空间中的推广

Azmi 与 Kunisch[3, 4] 已经把 BB 方法的收敛分析推广到 Hilbert 空间和偏微分方程约束的优化问题。我们关于最坏收敛因子的工作 [37] 进一步考虑由有界、自伴、一致正定的算子 A（$0 < mI \le A \le LI$）定义的二次问题。记 E_A 为 A 的谱分解（投影值测度），对非零的 g，定义谱测度

$$\nu_g(B) = \frac{\|E_A(B)g\|^2}{\|g\|^2}.$$

取 $K = \mathrm{spec}(A)$，记 $\mathcal{P}(K)$ 为 K 上的概率测度全体。由相邻两步的谱测度组成的状态仍满足一个自治的递推，状态空间 $\mathcal{P}(K)\times\mathcal{P}(K)$ 在弱拓扑下是紧的。这里不能照搬有限维中单位截面在范数拓扑下的紧性（参见第 5.1 小节末的例子）。

记 $[a,b]$ 为包含 $\operatorname{supp}\nu_{g_0}$ 的最小闭区间。连续谱的端点上不一定有原子，所以 $\log\nu_g(\{a\})$ 一般没有意义。证明中用左右端点附近一小段谱的测度代替端点上的分量，对固定的长度用回复定理，再令长度趋于零。得到的渐近收敛因子的上界仍是 $(b-a)/(b+a)$；首步步长按 BB 公式计算时，一致上界中最优的因子由算子整个谱的端点确定。用固定的权作变量替换时，还要求权函数在谱上有正的上界和下界，以保证变换有界可逆。

“一致上界中的因子不能更小”与“每个算子都有达到这个因子的轨道”是两个不同的结论。取谱测度集中在两个端点附近的初值，可以使轨道在任意给定的有限步内任意接近平衡状态，这足以说明不存在更小的一致因子；但这个论证本身并不保证存在渐近收敛因子恰好等于这个值的轨道。预印本中有一个纯连续谱的例子：在 $L^2([a,b])$ 上令 A 为乘法算子 $(Ag)(\lambda)=\lambda g(\lambda)$，BB1 迭代取 $g_0\equiv 1$，首步步长按 BB 公式计算，则

$$\frac{\|g_k\|}{\|g_0\|}=\frac{c^k}{\sqrt{2k+1}},\qquad c=\frac{b-a}{b+a}.$$

它的渐近收敛因子为 c，但并不是精确的几何下降。这个例子也说明，像 $1/\sqrt{2k+1}$ 这样的多项式型因子不影响渐近收敛因子，却可能影响有限步的表现。

9.3 非线性问题的局部结果

对非二次的目标函数，关于最坏收敛因子的预印本证明的是局部结果。记 $G=\nabla f$，并设 $G(x_*)=0$。所需的主要条件是 G 在 x_* 处严格 Fréchet 可微，即存在有界、自伴、一致正定的 H_*，使

$$\delta(r):=\sup_{\substack{x\neq y\\ x,y\in B_r(x_*)}}\frac{\|G(x)-G(y)-H_*(x-y)\|}{\|x-y\|}\longrightarrow 0. \tag{9.1}$$

这个条件控制了邻域内所有的两点割线，比只要求在 x_* 处可微要强；$f\in C^2$ 且 Hesse 矩阵在 x_* 处按算子范数连续，就足以保证它成立，不需要 Hesse 矩阵局部 Lipschitz 连续。记

$$m_*=\min\operatorname{spec}(H_*),\qquad L_*=\max\operatorname{spec}(H_*),\qquad c_*=\frac{L_*-m_*}{L_*+m_*}.$$

全程使用 BB1 或全程使用 BB2 时，对每个 $\gamma\in(c_*,1)$，都有相应的半径 $r_\gamma>0$ 和一致的常数 C_γ：从球 $B_{r_\gamma}(x_*)$ 内任意两个不同的初始点出发，迭代都有定义并停留在 x_* 的一个邻域内，误差有形如 $C_\gamma\gamma^k$ 的局部几何上界（以两个初始点的误差为基准）。对有定义且收敛到 x_* 的轨道，也有相应的渐近收敛因子的上界 [37]。

证明并不是把二次函数的结果形式地套到 H_* 上。需要先对以 H_* 为 Hesse 矩阵的二次模型，证明首步步长属于某个紧区间时的一致上界；再在梯度第一次小于某个阈值之前控制割线的分母，在固定的有限步内跟踪二次模型的轨道；最后在小邻域内分段重新启动。这样，二次模型上的结论才能转移到原来的非线性迭代。这个结论不包括初始点离 x_* 很远时的全局收敛；当 $c_*=0$ 时，渐近收敛因子为零，表示比任意固定的几何速度都快，但不能由此断言迭代是 Q-超线性收敛的。

9.4 进一步的问题

这次阅读的一个体会是：动力系统的工具是否有效，取决于所选的状态是否与算法的结构相匹配。以 LMSD 为例，单个 Ritz 值在降秩的极限中难以控制，一组步长的乘子多项式却可以写成行列式之比，进而写成凸组合；以 BB 方法为例，某一步范数变大，使直接的收缩估计失效，端点不等式给出的分解式却能控制最坏的长期平均。

首先，按精确秩计算与实际中按数值秩阈值截断之间还有距离。现有的 LMSD 定理针对的是保留历史矩阵全部列空间的算法；按数值阈值删列会改变这个空间和乘子，需要把截断误差放进状态，或者建立误差逐组累积时仍可控制的扰动估计。

其次，用固定的权作变量替换，可以解释 BB1、BB2 和其他固定正权的步长规则共同的渐近性质，但不包括权随迭代变化或任意切换的情形。若要推广，需要研究各规则共同的不变量、把切换信息放进状态，或者研究联合谱半径一类的长期增长率。对 LMSD 来说，现有的 R-线性收敛框架只证明了一致因子的存在，$p > 1$ 时最优因子的显式值尚未确定。

最后，第 8 节的 Lyapunov 函数表明，函数值不单调时，仍然存在可以利用的单调下降量；但它依赖于特征投影，在端点分量为零处退化。按其他特征值子集构造、更容易计算的类似组合，对特征值估计误差不敏感的替代函数，以及非二次问题中的局部 Lyapunov 函数，都值得进一步研究。推广时应当保留两个要求：函数的下降能够由实际的迭代公式验证；它为零的边界也要单独控制。

参考文献

[1] H. Akaike. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Annals of the Institute of Statistical Mathematics*, 11(1):1–16, 1959. doi: 10.1007/BF01831719.

[2] Uri M. Ascher, Kees van den Doel, Hui Huang, and Benar F. Svaiter. Gradient descent and fast artificial time integration. *ESAIM: Mathematical Modelling and Numerical Analysis*, 43(4):689–708, 2009. doi: 10.1051/m2an/2009025.

[3] Behzad Azmi and Karl Kunisch. Analysis of the Barzilai–Borwein step-sizes for problems in Hilbert spaces. *Journal of Optimization Theory and Applications*, 185(3):819–844, 2020. doi: 10.1007/s10957-020-01677-y.

[4] Behzad Azmi and Karl Kunisch. On the convergence and mesh-independent property of the Barzilai–Borwein method for PDE-constrained optimization. *IMA Journal of Numerical Analysis*, 42(4):2984–3021, 2022. doi: 10.1093/imanum/drab056.

[5] Jonathan Barzilai and Jonathan M. Borwein. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148, 1988. doi: 10.1093/imanum/8.1.141.

[6] Frank E. Curtis and Wei Guo. R-linear convergence of limited memory steepest descent. *IMA Journal of Numerical Analysis*, 38(2):720–742, 2018. doi: 10.1093/imanum/drx016.

[7] Yu-Hong Dai. Alternate step gradient method. *Optimization*, 52(4–5):395–415, 2003. doi: 10.1080/02331930310001611547.

[8] Yu-Hong Dai. A new analysis on the Barzilai–Borwein gradient method. *Journal of the Operations Research Society of China*, 1(2):187–198, 2013. doi: 10.1007/s40305-013-0007-x.

[9] Yu-Hong Dai and Roger Fletcher. On the asymptotic behaviour of some new gradient methods. *Mathematical Programming*, 103(3):541–559, 2005. doi: 10.1007/s10107-004-0516-9.

[10] Yu-Hong Dai and Li-Zhi Liao. R-linear convergence of the Barzilai and Borwein gradient method. *IMA Journal of Numerical Analysis*, 22(1):1–10, 2002. doi: 10.1093/imanum/22.1.1.

[11] Yu-Hong Dai, William W. Hager, Klaus Schittkowski, and Hongchao Zhang. The cyclic Barzilai–Borwein method for unconstrained optimization. *IMA Journal of Numerical Analysis*, 26(3):604–627, 2006. doi: 10.1093/imanum/drl006.

[12] Giulia Ferrandi and Michiel E. Hochstenbach. Limited memory gradient methods for unconstrained optimization. *Numerical Algorithms*, 99:735–764, 2025. doi: 10.1007/s11075-024-01895-9.

[13] Roger Fletcher. On the Barzilai–Borwein method. In L. Qi, K. L. Teo, and X. Yang, editors, *Optimization and Control with Applications*, volume 96 of *Applied Optimization*, pages 235–256. Springer, Boston, 2005. doi: 10.1007/0-387-24255-4_10.

[14] Roger Fletcher. A limited memory steepest descent method. *Mathematical Programming*, 135:413–436, 2012. doi: 10.1007/s10107-011-0479-6.

[15] George E. Forsythe. On the asymptotic directions of the s-dimensional optimum gradient method. *Numerische Mathematik*, 11(1):57–76, 1968. doi: 10.1007/BF02165472.

[16] A. Friedlander, J. M. Martínez, B. Molina, and Marcos Raydan. Gradient method with retards and generalizations. *SIAM Journal on Numerical Analysis*, 36(1):275–289, 1999. doi: 10.1137/S003614299427315X.

[17] Gene H. Golub and Gérard Meurant. *Matrices, Moments and Quadrature with Applications*. Princeton Series in Applied Mathematics. Princeton University Press, 2010.

[18] Ran Gu and Qiang Du. A modified limited memory steepest descent method motivated by an inexact super-linear convergence rate analysis. *IMA Journal of Numerical Analysis*, 41(1):247–270, 2021. doi: 10.1093/imanum/drz059.

[19] Dawei Li and Ruoyu Sun. On a faster R-linear convergence rate of the Barzilai–Borwein method, 2021. URL https://arxiv.org/abs/2101.00205v2. Preprint, arXiv:2101.00205v2.

[20] Dawei Li, Xiaotian Jiang, and Mingyi Hong. Barzilai–Borwein fails superlinear convergence on an open set of quadratics for every dimension $n \geq 4$, 2026. URL https://arxiv.org/abs/2607.21579v1. Preprint, arXiv:2607.21579v1.

[21] Xin-Rui Li and Ya-Kui Huang. A note on R-linear convergence of nonmonotone gradient methods. *Journal of the Operations Research Society of China*, 13(1):313–325, 2025. doi: 10.1007/s40305-023-00468-2.

[22] Chris C. Paige, Beresford N. Parlett, and Henk A. van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2(2):115–133, 1995.

[23] Luc Pronzato and Anatoly Zhigljavsky. Gradient algorithms for quadratic optimization with fast convergence rates. *Computational Optimization and Applications*, 50:597–617, 2011. doi: 10.1007/s10589-010-9319-5.

[24] Luc Pronzato, Henry P. Wynn, and Anatoly A. Zhigljavsky. Asymptotic behaviour of a family of gradient algorithms in $\mathbb{R}^d$ and Hilbert spaces. *Mathematical Programming*, 107(3):409–438, 2006. doi: 10.1007/s10107-005-0602-7.

[25] Marcos Raydan. On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13(3):321–326, 1993. doi: 10.1093/imanum/13.3.321.

[26] Marcos Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, 7(1):26–33, 1997. doi: 10.1137/S1052623494266365.

[27] Jinglai Shen and Jianghai Hu. Stability of discrete-time switched homogeneous systems on cones and conewise homogeneous inclusions. *SIAM Journal on Control and Optimization*, 50(4):2216–2253, 2012. doi: 10.1137/110845215.

[28] Bin Shi, Simon S. Du, Weijie J. Su, and Michael I. Jordan. Acceleration via symplectic discretization of high-resolution differential equations. In *Advances in Neural Information Processing Systems*, volume 32, pages 5745–5753, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/a9986cb066812f440bc2bb6e3c13696c-Abstract.html.

[29] Bin Shi, Simon S. Du, Michael I. Jordan, and Weijie J. Su. Understanding the acceleration phenomenon via high-resolution differential equations. *Mathematical Programming*, 195:79–148, 2022. doi: 10.1007/s10107-021-01681-8.

[30] Rob Sturman and Jaroslav Stark. Semi-uniform ergodic theorems and applications to forced systems. *Nonlinearity*, 13(1):113–143, 2000. doi: 10.1088/0951-7715/13/1/306.

[31] Weijie Su, Stephen Boyd, and Emmanuel J. Candès. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, 17(153): 1–43, 2016. URL https://jmlr.org/papers/v17/15-084.html.

[32] Andre Wibisono, Ashia C. Wilson, and Michael I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences of the United States of America*, 113(47):E7351–E7358, 2016. doi: 10.1073/pnas.1614734113. URL https://arxiv.org/abs/1603.04245v1.

[33] Ashia C. Wilson, Ben Recht, and Michael I. Jordan. A Lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, 22(113):1–34, 2021. URL https://jmlr.org/papers/v22/20-195.html.

[34] Shutai Yang and Ya-xiang Yuan. R-linear convergence of Barzilai–Borwein methods on strictly convex quadratics via closed-cone homogeneous dynamics, 2026. URL https://arxiv.org/abs/2609.12100v1. Preprint, arXiv:2609.12100v1, 10 September 2026.

[35] Shutai Yang and Ya-xiang Yuan. Global and unconditional R-linear convergence of rank-compressed weighted LMSD sweeps for strictly convex quadratics, 2026. URL https://arxiv.org/abs/2609.12114v1. Preprint, arXiv:2609.12114v1, 10 September 2026.

[36] Shutai Yang and Ya-xiang Yuan. Lyapunov functions and R-linear convergence for quadratic Barzilai–Borwein dynamics, 2026. URL https://arxiv.org/abs/2609.12084v1. Preprint, arXiv:2609.12084v1, 10 September 2026.

[37] Shutai Yang and Ya-Xiang Yuan. The sharp worst-case asymptotic rate of the Barzilai–Borwein method in $\mathbb{R}^d$ and Hilbert spaces, 2026. URL https://arxiv.org/abs/2608.07839v1. Preprint, arXiv:2608.07839v1, 8 August 2026.

[38] Qinmeng Zou and Frédéric Magoulès. On extensions of limited memory steepest descent method. In *2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, pages 172–175. IEEE, 2019. doi: 10.1109/DCABES48411.2019.00050. URL https://arxiv.org/abs/1912.01222v1. Author preprint: arXiv:1912.01222v1.

[39] Qinmeng Zou and Frédéric Magoulès. Delayed gradient methods for symmetric and positive definite linear systems. *SIAM Review*, 64(3):517–553, 2022. doi: 10.1137/20M1321140.

[40] 袁亚湘. 梯度法步长基于子空间逼近的选取. 计算数学, 48(3):395–404, 2026. doi: 10.12286/jssx.j2026-1387.