中国科学技术大学

本科毕业论文



块映射视角下固定记忆谱梯度法的统一收敛理论及应用

作者姓名:	杨书泰
学　　号:	公开版已隐去
专　　业:	数学与应用数学
导　　师:	陈士祥特任研究员 袁亚湘研究员
完成时间:	2026 年 5 月 10 日

中国科学技术大学本科毕业论文学术诚信承诺书

本人郑重承诺：所呈交的本科毕业论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。如出现侵犯他人知识产权的行为，由本人承担相应法律责任。本人承诺不存在抄袭、伪造、篡改、代写、买卖毕业论文（设计）等违纪行为。

作者签名：_____

签字日期：_____

摘要

本文研究固定记忆谱梯度法在连续正齐次块映射视角下的统一收敛理论，主要对象包括 BB1/BB2、LMSD、harmonic-LMSD 以及 ω 加权 Ritz-sweep 方法族。核心做法是把最近若干步误差或梯度端点组成固定维数块状态，并利用谱步长对整体缩放的零次齐次性，将一个完整块的迭代写成闭锥上的连续一次正齐次映射。本文证明：若严格凸二次模型的块映射在闭锥上逐点收敛到零，则自动存在统一有限步收缩，从而得到全局 R-线性收敛；对于只在正则区域自然定义的非线性算法，也给出紧子集上的局部化版本，为后续 GLL 全局化分析提供工具。

在严格凸二次问题上，本文从矩阵最小二乘和子空间投影视角出发，统一刻画 pD(S)-BB 与 pD(Y)-BB 步长：前者对应 Hessian 的 Ritz 值，后者对应逆 Hessian 的 inverse harmonic Ritz 值。借助块乘子的行列式比表示、Cauchy–Binet 展开和乘子权重引理，本文证明精确秩压缩 pD-BB 方法在 $p \geqslant n$ 时至多 $p+n$ 步有限终止，并建立 LMSD、harmonic-LMSD 及其 ω 加权推广的逐点收敛性质。进一步，针对 LMSD 与 harmonic-LMSD 中历史束可能降秩的问题，本文把状态从历史矩阵改为相容块端点梯度对，构造闭锥上的连续正齐次映射，证明其在严格凸二次精确秩压缩模型中无须附加 Curtis–Guo 式历史束非退化假设即可 R-线性收敛。同时，本文指出并修正 Fletcher 关于 LMSD 原始收敛证明中 Ritz 值极限连续性的漏洞。

对于一般非线性问题，本文引入对称化小规模广义特征值模型，并结合秩压缩、非正特征值替代、步长截断和 GLL 非单调线搜索，构造 pD(S)-BB 的全局化实现。在 $C^{2,1}$ 局部光滑、尾轨道进入正则区域且 GLL 线搜索最终全接受的条件下，证明该全局化算法与局部二次模型具有相同尾部块收缩，从而得到局部 R-线性收敛。数值实验验证了 $p \geqslant n$ 时有限终止上界、LMSD/harmonic-LMSD 相比 BB1/BB2 与若干自适应 BB 方法的迭代优势、ω 权重对谱分布的敏感性、以及非线性测试问题中 GLL 版本的局部 R-线性包络。最后，作为块动力系统视角的独立延伸，本文证明经典 BB 方法在严格凸二次问题上的紧渐近速率常数为 $(\kappa(H)-1)/(\kappa(H)+1)$，改进了已有 $1-1/\kappa$ 型上界。

关键词： 连续正齐次块映射；固定记忆谱梯度法；Barzilai–Borwein 步长；LMSD；矩阵最小二乘；GLL 非单调线搜索；局部 R-线性收敛

ABSTRACT

This thesis studies a unified convergence theory for fixed-memory spectral gradient methods from the viewpoint of continuous positively homogeneous block mappings. The main methods under consideration include BB1/BB2, LMSD, harmonic-LMSD, and a family of ω-weighted Ritz-sweep methods. The central idea is to collect several recent error vectors or gradient endpoint vectors into a fixed-dimensional block state and, using the degree-zero homogeneity of spectral steplength rules under global scaling, to represent one complete block of iterations as a continuous degree-one positively homogeneous mapping on a closed cone. It is proved that if the block mapping associated with a strictly convex quadratic model converges pointwise to zero on the closed cone, then a uniform finite-step contraction automatically follows, which yields global R-linear convergence. A localized compact-set version is also established for nonlinear algorithms that are naturally defined only on regular regions, providing a tool for the subsequent GLL-globalized analysis.

For strictly convex quadratic problems, this thesis provides a unified characterization of the pD(S)-BB and pD(Y)-BB steplengths from matrix least-squares and subspace-projection viewpoints: the former corresponds to Ritz values of the Hessian, while the latter corresponds to inverse harmonic Ritz values associated with the inverse Hessian. Based on a determinant-ratio representation of block multipliers, the Cauchy–Binet expansion, and a multiplier-weight lemma, this thesis proves that the exact rank-compressed pD-BB method terminates in at most $p+n$ steps when $p \geqslant n$, and establishes pointwise convergence for LMSD, harmonic-LMSD, and their ω-weighted extensions. Furthermore, to address possible rank deficiency of the history bundle in LMSD and harmonic-LMSD, the state is changed from history matrices to compatible block-end gradient pairs. This construction yields a continuous positively homogeneous mapping on a closed cone and proves R-linear convergence for the exact rank-compressed strictly convex quadratic models without imposing any Curtis–Guo type nondegeneracy assumption on the history bundle. This thesis also identifies and corrects a gap in Fletcher's original convergence proof for LMSD concerning the limiting continuity of Ritz values.

For general nonlinear problems, this thesis introduces a symmetrized small-scale generalized eigenvalue model and combines it with rank compression, replacement of

nonpositive eigenvalues, steplength truncation, and the GLL nonmonotone line search to obtain a globalized implementation of pD(S)-BB. Under local $C^{2,1}$ smoothness, eventual entrance of the tail orbit into a regular region, and eventual full acceptance of the GLL line search, the globalized algorithm is shown to have the same tail block contraction as the local quadratic model, which gives local R-linear convergence. Numerical experiments verify the finite-termination bound for $p \geqslant n$, the iteration advantage of LMSD/harmonic-LMSD over BB1/BB2 and several adaptive BB methods, the sensitivity of ω-weights to the spectral distribution, and the local R-linear envelope of the GLL version on nonlinear test problems. Finally, as an independent extension of the block dynamical-system viewpoint, this thesis proves that the sharp asymptotic rate constant of the classical BB method on strictly convex quadratic problems is $(\kappa(H)-1)/(\kappa(H)+1)$, improving the existing $1-1/\kappa$ type upper bound.

Key Words: Continuous positively homogeneous block mapping; Fixed-memory spectral gradient method; Barzilai–Borwein steplength; LMSD; Matrix least squares; GLL nonmonotone line search; Local R-linear convergence

目录

第一章　引言 …… 4

第二章　连续正齐次映射的收敛与 R-线性收敛 …… 8
第一节　基本定义 …… 8
第二节　统一有限步收缩原理 …… 8
第三节　闭锥状态下一般固定记忆模型的 R-线性收敛 …… 11
第四节　紧子集上的局部收敛 …… 13
第五节　一些例子 …… 15

第三章　LMSD 的高维最小二乘解释与子空间结构 …… 19
第一节　BB1/BB2 的最小二乘解释 …… 19
第二节　p 阶矩阵的最小二乘推广 …… 19
第三节　严格凸二次情形下与原始公式的等价性 …… 20
第四节　二次模型中的块步长定义及其与 LMSD 的对应关系 …… 21

第四章　二次模型的有限终止、逐点收敛与 R-线性收敛 …… 24
第一节　pD-BB 梯度法的有限终止定理 …… 24
第二节　二次模型上的逐点收敛 …… 26
一、几个引理 …… 27
二、S 型方法的逐点收敛 …… 31
三、Y 型方法的逐点收敛 …… 32
第三节　二次模型上的 R-线性收敛 …… 33
一、相容块端点梯度对闭锥 …… 33
二、R-线性收敛定理 …… 37
第四节　ω 加权 Ritz-sweep 方法族 …… 39
第五节　λ_1-分量的显式收缩界 …… 42

第五章　一般问题下的对称化算法 …… 43
第一节　一般非线性问题下的对称化改进步长 …… 43
第二节　二维显式公式 …… 44
第三节　一些约定 …… 46
第四节　纯算法与 GLL 全局化算法 …… 47

第六章　GLL 全局化下 *p*D(S)-BB 算法的收敛与 *R*-线性收敛 50
第一节　GLL 检测下 *p*D(S)-BB 算法的全局收敛性 50
第二节　关于无 GLL 检测时算法全局收敛性的反例 50
第三节　局部假设与块映射 51
第四节　局部二次逼近与步长扰动估计 53
第五节　GLL 全局化方法的 *R*-线性收敛 59

第七章　数值实验 64
第一部分：二次模型实验 64
第一节　实验设置 64
第二节　有限终止现象 64
第三节　与 BB1、BB2、共轭梯度法的比较 66
第四节　历史束最小奇异值的数值衰减 67
第五节　不同谱下的 ω-加权表现 68
第六节　处理大规模问题的性能 69
第七节　经验 *R*-线性包络 69
第八节　与自适应 BB 变体的对比 71
第二部分：非二次模型实验 72
第九节　非二次实验设置与对比基线 72
第十节　精确 Hessian 公式与改进对称化公式的比较 73
第十一节　Trigonometric 测试问题 74
第十二节　Chained Rosenbrock 测试问题 75
第十三节　Convex 2 测试问题 75
第十四节　局部 *R*-线性渐近行为的经验拟合 76
第十五节　讨论 76

第八章　总结与展望 78
第一节　本文证明范式的意义 78
第二节　对其他方法的适用性 79
第三节　动力系统视角的拓展：经典 BB 方法的紧渐近速率 79
一、问题陈述与已有结果 79
二、谱压缩与归一化两步动力系统 81
三、二维情形与速率下界 82
四、终止态紧化 83
五、端点恒等式与最优常数 84

六、从测度回到轨道：定理 8.2 与 8.3 的证明 …… 90
七、精确周期轨道的初等证明 …… 93
八、与本文整体框架的关系 …… 94
第四节　问题与展望 …… 94

参考文献 …… 95
附录 A　从最小二乘到正规方程：标量与矩阵情形 …… 98
第一节　标量情形：BB1 与 BB2 …… 98
第二节　矩阵情形：原始 OLS 模型 …… 98
附录 B　二维对称化公式的代数推导 …… 100
附录 C　三维与四维显式公式：原始 OLS 与对称化模型的统一写法 …… 101
第一节　系数的来源：迹、主子式与 Newton 恒等式 …… 101
第二节　三维情形 …… 101
第三节　四维情形 …… 102
附录 D　核心代码片段 …… 104
第一节　对称化 S 型步长生成器 …… 104
第二节　精确 Hessian 投影基线的实现 …… 104
第三节　ω-加权 Ritz-sweep 步长 …… 105
第四节　块队列执行器 …… 106
第五节　GLL 非单调线搜索内核 …… 107
第六节　包络拟合与稳定 $\log\cosh$ …… 107
致谢 …… 109

第一章　引言

考虑无约束优化问题

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

其梯度法基本迭代格式为

$$x_{k+1} = x_k - \alpha_k g_k, \qquad g_k = \nabla f(x_k). \tag{2}$$

Barzilai 与 Borwein 提出的 BB1/BB2 步长[2]

$$\alpha_k^{\mathrm{BB1}} = \frac{s_k^T s_k}{s_k^T y_k}, \qquad \alpha_k^{\mathrm{BB2}} = \frac{s_k^T y_k}{y_k^T y_k}, \tag{3}$$

其中

$$s_k = x_{k+1} - x_k, \qquad y_k = g_{k+1} - g_k$$

是谱梯度法最重要的出发点之一。对严格凸二次问题，Raydan 证明了 BB 方法的全局收敛[3]，并把 BB 与 Grippo–Lampariello–Lucidi 提出的非单调线搜索结合起来，得到一般非线性问题上的全局化算法[1,5]。Dai–Liao 证明了 BB 方法的 R-线性收敛[7]，Dai 进一步把这一证明思路抽象为 Property A，形成了一条影响很大的谱分量递推路线[9]。

由 BB 方法衍生出众多固定记忆谱梯度方法，其中 Fletcher 的 limited memory steepest descent (LMSD) 是较为重要的一个分支，也是本文的主要讨论对象。从严格凸二次问题的角度看，BB、Yuan 的 2D/3D 步长、Fletcher 的 LMSD 以及 harmonic-LMSD，本质上都属于同一类固定记忆谱梯度法：它们都利用最近若干步的位移、梯度或梯度差，在一个低维子空间上提取谱信息，再把这些信息转化为一整块步长，在算法中依次使用。Fletcher 在二次情形下把一个 BB 步长推广成一整个 Ritz-sweep[11]，Curtis–Guo 证明了 fixed-cycle LMSD 在强凸二次问题上的 R-线性收敛，但其证明依赖一个非退化的附加条件，即在 Curtis–Guo 的块记号下 QR 因子满足 $\|R_k^{-1}\| \leqslant \rho\|g_{k,1}\|^{-1}$[13]。这个条件在 $p = 1$ 时自动成立，而在 $p > 1$ 时，由于梯度向量可能随着迭代趋于平行，该条件并非无条件事实：Curtis–Guo 在数值实验中记录了运行过程中观察到当 $m = 5$ 时，在某些测试问题上这一最大值可达到约 2×10^{16}[13] 中 Table 1。Curtis–Guo 还讨论了非正曲率下的 LMSD 变体及其非单调线搜索全局化[12]。Ferrandi–Hochstenbach 则从 Hessian 矩阵的子空间投影、割线最小二乘与对称化的角度，对 LMSD 及其在解决一般非线性问题时的若干拓展方法给出了统一解释[19]。

本文对上述固定记忆谱梯度法的严格凸二次问题收敛性质给出一个统一的视角和新的证明方法，把研究对象分为两大层。第一层是严格凸二次问题上的精

确理论模型：在该模型中，历史块的秩等价于活跃不同 Ritz 值个数，历史矩阵与 Krylov 矩阵具有相同秩，精确秩下降来自活跃不同特征值个数已经小于 p，此时带秩压缩的算法在实际活跃子空间上执行步数相应减少的算法并有限终止。第二层是一般非线性与浮点计算中的实际操作模型：数值秩压缩、删列、非正特征值替代、步长截断以及 GLL 检测都是实现层面的保护机制。第 4 章的主要新结论属于第一层，第 5–6 章则处理第二层中的算法定义、扰动估计和全局化问题。

首先考虑历史窗口固定（设为 p）的情形。固定一个块长 $p \in \mathbb{N}$，则对固定记忆谱方法，本文统一把第 k 个完整块中第 r 个位置的点、梯度与位移分别记为

$$x_{k,r}, \qquad g_{k,r} = \nabla f(x_{k,r}), \qquad s_{k,r} = x_{k,r+1} - x_{k,r}, \qquad r = 0, \cdots, p-1.$$

相应地，

$$E_k = [x_{k,0}, x_{k,1}, \cdots, x_{k,p}],$$
$$S_k = [s_{k,0}, s_{k,1}, \cdots, s_{k,p-1}],$$
$$G_k = [g_{k,0}, g_{k,1}, \cdots, g_{k,p-1}]$$

分别表示第 k 个完整块的状态、位移束与梯度束。则在固定块长模型中有

$$x_{kp+r} = x_{k,r}, \qquad g_{kp+r} = g_{k,r}, \qquad s_{kp+r} = s_{k,r}, \qquad x_{k,p} = x_{k+1,0}.$$

不失一般性，本文对于二次问题做出如下标准形约定。

假设 1.1 凡涉及严格凸二次函数

$$q(x) = q(x^*) + \frac{1}{2}(x - x^*)^T H(x - x^*), \qquad H \succ 0,$$

的理论论证，我们总是先作平移，把极小点移到原点，再作正交变换把 H 对角化。若某个特征值有重数，则先把问题压缩到由初始梯度生成的活跃不变子空间，并在每个活跃特征子空间内选择正交基，使初始梯度至多在其中一个方向上有非零坐标。由于一旦某个坐标为零，二次梯度递推会永远保持它为零，因此二次情形下的收敛与终止性质只取决于活跃的不同特征值个数，与代数重数无关。于是，对后续所有严格凸二次证明，都可以不失一般性地假设

$$q(x) = \frac{1}{2}x^T H x, \qquad H = \mathrm{diag}(\lambda_1, \cdots, \lambda_n), \qquad 0 < \lambda_1 < \cdots < \lambda_n,$$

其中 $\lambda_1, \cdots, \lambda_n$ 是活跃的不同特征值。

在这个约定下，二次理论里 x 同时也是迭代点相对最优点的误差变量。

本文的核心命题有两层。第一层是全局闭锥版本：若某一固定记忆方法能够在合适闭锥上写成连续正齐次自映射，则二次模型的逐点收敛会自动升级为全局 R-线性收敛。第二层是局部正则版本：对一般非线性问题或带保护机制的实际实现，若归一化尾轨道最终停留在与退化边界保持正距离的紧子集中，则仍可在该紧集诱导的锥上得到局部化的块收缩。第 4 章表明，严格凸二次精确模型中

的历史束秩退化可以通过重新选择状态空间来纳入第一层，而不必再引入额外非退化假设（如 Curtis Guo[13]）排除。

本文的主要工作可以概括为以下五点。

(i) 在 BB1/BB2 的一维最小二乘解释以及 Ferrandi–Hochstenbach 对 LMSD 的高维最小二乘视角基础上，采用统一块记号重写最近 p 个割线对上的小矩阵模型，从而把 Yuan 的 2D/3D 公式、LMSD 以及 harmonic-LMSD 纳入同一代数框架。

(ii) 在统一块记号和二次标准形下，把有限终止机制重写为该记号下的一个一般命题：当历史窗口大小 p 不小于初始梯度激活的不同特征值个数 n 时，纯 pD-BB 方法在严格凸二次问题上至多经过 $p+n \leqslant 2p$ 步有限终止。

(iii) 把完整 Ritz-sweep 在谱分量上的乘子写成行列式之比，再由 Cauchy–Binet 展开得到非负加权平均，从而避开 Krylov 子空间维数下降时单个 Ritz 值极限不连续的问题。基于这一框架，指出 Fletcher 原始文章中关于 LMSD 收敛性证明的一个错误，给出 LMSD 与 harmonic-LMSD 的全局收敛证明，并进一步推广到统一的 ω-加权 Ritz 结构。

(iv) 给出严格凸二次精确模型中 LMSD 的相容块端点梯度对闭锥证明。若 Krylov 维数下降，算法作用退化为有限终止分支。由第 2 章的闭锥齐次映射原理，可得到无历史束非退化假设的 R-线性收敛。

(v) 在一般非线性情形下，引入对称化广义特征值模型，给出 pD(S)-BB 的带压缩纯实现与 GLL 全局化版本，并证明 GLL 全局化实现的全局收敛及局部 R-线性收敛率。

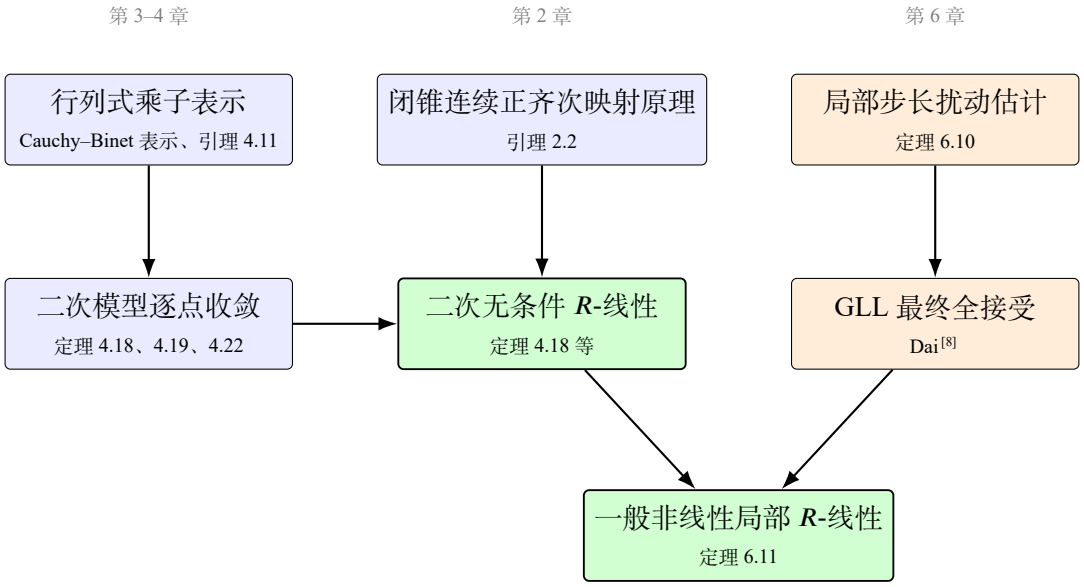


图 1.1　本文技术路线。

文献	算法	历史束假设	范数 $\|g_k\|$ 速率
Raydan[3]	BB（p=1）	无	无
Dai–Liao[7]	BB（p=1）	无	R-线性（无条件）
Fletcher[11]	LMSD	无（声称）*	无
Curtis–Guo[13]	LMSD	$\|R_k^{-1}\| \leqslant \rho \|g_{k,1}\|^{-1}$	R-线性（条件性）
本文	ω 加权 Ritz 方法族	无	R-线性（无条件）

表 1.1　本文与既有 BB / LMSD 收敛结果的对比。*Fletcher[11] 附录中证明的 LMSD 收敛存在 Ritz 值极限不连续的问题，本文予以了修正。

本文结构如下：第 2 章首先在闭锥状态空间上给出抽象齐次块映射工具，并把固定步长梯度法以及 BB 方法的梯度对状态作为一个例子，给出一个不同于既有方法[7]的 R-线性收敛证明，同时给出紧集上的局部化版本。第 3–4 章讨论严格凸二次问题：第 3 章从 Ferrandi–Hochstenbach 对 LMSD 的最小二乘解释和历史割线投影，以及 Krylov 子空间结构出发，统一固定记忆谱步长的一般框架；第 4 章利用相容块端点梯度对闭锥，在这一框架下讨论有限终止和收敛性质，从而给出精确秩压缩 LMSD / harmonic-LMSD 的无历史束非退化假设 R-线性收敛证明，进一步把这一证明推广到任意正权 ω 加权的 Ritz-sweep 方法族，并把 BB1、BB2 作为 $p = 1$ 实例放回同一闭锥框架，作为方法特例再次证明了 Dai–Liao 的经典 R-线性结论。第 5–6 章讨论一般非线性问题：第 5 章引入对称化步长与相应带压缩纯方法和 GLL 全局化实现，第 6 章对 pD-BB 处理一般非线性问题时的收敛性质进行系统讨论。第 7 章合并给出二次与非二次模型的数值实验：前半部分验证有限终止，R-线性包络等定性现象，后半部分把 GLL 全局化版本应用到传统非线性测试问题上，并对尾轨道做经验 R-线性拟合。第 8 章在讨论本方法的使用范围、与现有路线的关系、以及它对其他固定记忆谱梯度方法的适用性的基础上，进一步把块动力系统视角应用于经典 BB 方法本身，给出严格凸二次问题上 BB 紧渐近速率 $(\kappa-1)/(\kappa+1)$ 的证明，作为方法的一个独立延伸。附录收录代数推导与实现细节。

第二章　连续正齐次映射的收敛与 R-线性收敛

本节给出全文的主要工具。核心思想是：在有限维赋范空间中，连续正齐次复合映射只要对每个初值逐点收敛到零，就自动具有统一有限步收缩，因此必然全局 R-线性收敛。然而应用到具体算法收敛性的证明中时，全局逐点收敛的条件过强，且很多初始状态并不是算法所能达到的自然状态。因此我们首先在“闭锥状态空间”上陈述并证明相关抽象结论，再以固定步长梯度法及 BB 方法作为例子说明如何具体应用。

第一节　基本定义

定义 2.1 (闭锥状态空间)　设 X 是有限维赋范空间。若非空集合 $\Omega \subset X$ 满足

$$0 \in \Omega, \qquad x \in \Omega,\ t \geqslant 0 \Longrightarrow tx \in \Omega,$$

并且 Ω 在 X 中闭，则称 Ω 为 X 中的一个闭锥。

定义 2.2 (闭锥上的连续正齐次映射)　设 $\Omega \subset X$ 是闭锥。若映射 $T : \Omega \to \Omega$ 满足

$$T(tx) = tT(x), \qquad \forall t \geqslant 0,\ \forall x \in \Omega,$$

则称 T 为一次正齐次映射。若进一步 T 在 Ω 上连续，则称 T 为闭锥 Ω 上的连续正齐次映射。

定义 2.3 (闭锥上的全局 R-线性收敛)　设 $T : \Omega \to \Omega$ 为闭锥上的映射。若存在常数 $C > 0$ 与 $\rho \in (0,1)$，使得对任意 $x \in \Omega$ 与 $k \geqslant 0$ 都有

$$\left\|T^k(x)\right\| \leqslant C\rho^k \|x\|,$$

则称 T 在 Ω 上全局 R-线性收敛到零。

第二节　统一有限步收缩原理

引理 2.1　设 $\Omega \subset X$ 是有限维赋范空间上的闭锥，$T : \Omega \to \Omega$ 为连续正齐次映射。则下列两条等价：

(i) 存在 $C > 0$ 与 $\rho \in (0,1)$ 使

$$\left\|T^k(x)\right\| \leqslant C\rho^k \|x\|, \qquad \forall x \in \Omega,\ \forall k \geqslant 0;$$

(ii) 存在整数 $N \geqslant 1$ 与常数 $\theta \in (0,1)$ 使

$$\left\|T^N(x)\right\| \leqslant \theta \|x\|, \qquad \forall x \in \Omega.$$

我们称此 T 在 Ω 上具有统一 N 步收缩。

证明 (i)⇒(ii)：只需取 N 使 $C\rho^N < 1$ 即可。

(ii)⇒(i)：对任意 $k \geqslant 0$，写作 $k = mN + r$ 且 $0 \leqslant r \leqslant N-1$。令

$$S_\Omega := \Omega \cap \{x \in X : \|x\| = 1\}.$$

由于 Ω 是闭锥且单位球面紧，S_Ω 亦为紧集。由连续性可知

$$K := \max\left\{1, \sup_{z\in S_\Omega} \|T(z)\|, \cdots, \sup_{z\in S_\Omega} \left\|T^{N-1}(z)\right\|\right\} < \infty.$$

于是对任意 $x \in \Omega$ 都有

$$\left\|T^k(x)\right\| = \left\|T^r(T^{mN}(x))\right\| \leqslant K\left\|T^{mN}(x)\right\| \leqslant K\theta^m \|x\|.$$

由于 $m = (k-r)/N \geqslant k/N - 1$，故

$$K\theta^m \leqslant \frac{K}{\theta}\left(\theta^{1/N}\right)^k.$$

再令 $\rho = \theta^{1/N} \in (0,1)$ 与 $C = K/\theta$ 即得结论。 ∎

引理 2.2 设 $\Omega \subset X$ 是有限维赋范空间上的闭锥，$T : \Omega \to \Omega$ 是连续正齐次映射，且对每个 $x \in \Omega$ 都有

$$T^k(x) \to 0.$$

则 T 在 Ω 上全局 R-线性收敛到零。从而，存在整数 $N \geqslant 1$ 与 $\theta \in (0,1)$ 使得

$$\left\|T^N(x)\right\| \leqslant \theta\|x\|, \qquad \forall x \in \Omega.$$

证明 令

$$S_\Omega := \Omega \cap \{x \in X : \|x\| = 1\}.$$

由 Ω 闭且单位球面紧，S_Ω 是紧集。对任意 $x \in S_\Omega$，由 $T^k(x) \to 0$ 可取整数 $n_x \geqslant 1$ 使

$$\|T^{n_x}(x)\| < \frac{1}{2}.$$

由于 T^{n_x} 连续，存在 x 在 S_Ω 中的开邻域 U_x，使得对所有 $y \in U_x$ 都有

$$\|T^{n_x}(y)\| < \frac{1}{2}.$$

于是 $\{U_x\}_{x\in S_\Omega}$ 构成 S_Ω 的开覆盖。由紧性可取有限子覆盖 $U_{x_1}, \cdots, U_{x_r}$，并令

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

于是对任意 $y \in S_\Omega$，总存在某个 i 使 $y \in U_{x_i}$，从而可取 $m(y) := n_{x_i} \in \{1, \cdots, N_0\}$ 使

$$\left\|T^{m(y)}(y)\right\| < \frac{1}{2}.$$

接下来把这一“至多 N_0 步减半”性质推广到任意向量。由于 T 连续且正齐次，存在

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

并且对任意 $z \in \Omega$ 与 $0 \leqslant j \leqslant N_0$ 都有

$$\left\|T^j(z)\right\| \leqslant K \|z\|.$$

现在固定任意 $x \in \Omega$。若 $x = 0$ 则结论显然成立，以下设 $x \neq 0$。递归定义

$$z_0 := x, \qquad z_q := T^{m_q}(z_{q-1}) \quad (q \geqslant 1),$$

其中每个 $m_q \in \{1, \cdots, N_0\}$ 取为归一化向量 $z_{q-1}/\left\|z_{q-1}\right\| \in S_\Omega$ 所对应的减半步数。若某一步得到 $z_q = 0$，则此后 $T^k(x) = 0$ 对所有足够大的 k 成立，结论显然成立；因此以下只需讨论所有 $z_q \neq 0$ 的情形。由正齐次性，

$$\left\|z_q\right\| = \left\|T^{m_q}(z_{q-1})\right\| \leqslant \frac{1}{2}\left\|z_{q-1}\right\|,$$

故归纳可得

$$\left\|z_q\right\| \leqslant 2^{-q}\|x\|, \qquad q \geqslant 0.$$

记 $s_q := \sum_{j=1}^{q} m_j$。给定任意 $k \geqslant 1$，取唯一整数 $p \geqslant 0$ 使

$$s_p \leqslant k < s_{p+1}.$$

由 $k - s_p < m_{p+1} \leqslant N_0$，得到

$$\left\|T^k(x)\right\| = \left\|T^{k-s_p}(T^{s_p}(x))\right\| \leqslant K\left\|T^{s_p}(x)\right\| = K\left\|z_p\right\| \leqslant K2^{-p}\|x\|.$$

另一方面，由 $m_j \leqslant N_0$ 可知

$$k < s_{p+1} \leqslant (p+1)N_0,$$

故 $p > k/N_0 - 1$。代回上式可得

$$\left\|T^k(x)\right\| \leqslant K2^{-k/N_0+1}\|x\| = 2K\left(2^{-1/N_0}\right)^k\|x\|.$$

因此，令

$$C := 2K, \qquad \rho := 2^{-1/N_0} \in (0,1),$$

便有

$$\left\|T^k(x)\right\| \leqslant C\rho^k\|x\|, \qquad \forall x \in \Omega,\ \forall k \geqslant 0.$$

即 T 在 Ω 上全局 R-线性收敛到零。最后由引理 2.1，存在整数 $N \geqslant 1$ 与 $\theta \in (0,1)$ 使

$$\left\|T^N(x)\right\| \leqslant \theta\|x\|, \qquad \forall x \in \Omega.$$

∎

注 引理 2.2 的关键在于“齐次性把逐点收敛升级成统一有限步收缩”。这一结论的证明完全基于有限覆盖、连续性和正齐次性。应用到具体算法时，真正需要认真选择的是状态空间本身：它必须是一个在算法递推下封闭的闭锥，并且要能容纳算法在边界处可能出现的退化行为。

第三节 闭锥状态下一般固定记忆模型的 R-线性收敛

在假设 1.1 的约定下，严格凸二次模型可写为

$$q(x) = \frac{1}{2}x^T Hx, \qquad H \succ 0. \tag{4}$$

固定一个块长 $p \in \mathbb{N}$。采用记号

$$x_{k,r} = x_{kp+r}, \qquad g_{k,r} = Hx_{k,r}, \qquad s_{k,r} = x_{k,r+1} - x_{k,r}, \qquad r = 0, 1, \cdots, p-1,$$

我们假设研究的算法以一个在递推下闭合，并且对整体缩放保持不变的闭锥 $\Omega \subset X$ 作为状态空间，并把第 k 个完整块的状态写为

$$E_k = (x_{k,0}, x_{k,1}, \cdots, x_{k,p}) \in X := (\mathbb{R}^n)^{p+1}.$$

在固定记忆谱梯度方法中，下一轮完整迭代的一整块步长由当前块的状态决定。为突出这一结构，我们把块内步长写成状态块

$$E = (x_0, x_1, \cdots, x_p)$$

的函数

$$\alpha_0(E), \alpha_1(E), \cdots, \alpha_{p-1}(E).$$

假设 2.1 设 $\Omega \subset X$ 为闭锥。对每个 $r = 0, \cdots, p-1$，步长函数 $\alpha_r : \Omega \setminus \{0\} \to \mathbb{R}_+$ 连续，并满足

$$\alpha_r(tE) = \alpha_r(E), \qquad \forall t > 0,\ \forall E \in \Omega \setminus \{0\}.$$

同时存在常数 $0 < \underline{\alpha} \leqslant \overline{\alpha} < \infty$，使得

$$\underline{\alpha} \leqslant \alpha_r(E) \leqslant \overline{\alpha}, \qquad \forall E \in \Omega \setminus \{0\},\ \forall r = 0, \cdots, p-1.$$

再由块内递推

$$\eta_0 = x_p, \qquad \eta_{r+1} = (I - \alpha_r(E)H)\eta_r, \qquad r = 0, \cdots, p-1,$$

定义块映射

$$\mathcal{T}_q(E) = (\eta_0, \eta_1, \cdots, \eta_p), \tag{5}$$

并假设它满足 $\mathcal{T}_q(\Omega) \subset \Omega$ 及 $\mathcal{T}_q(0) = 0$。

命题 2.3 在假设 2.1 下，$\mathcal{T}_q : \Omega \to \Omega$ 是连续、一次正齐次的映射，并且存在常数 $M > 0$ 使

$$\left\|\mathcal{T}_q(E)\right\|_\infty \leqslant M \left\|E\right\|_\infty, \qquad \forall E \in \Omega.$$

证明 先证一次正齐次性。由于对任意 $t > 0$ 都有 $\alpha_r(tE) = \alpha_r(E)$，于是

$$\eta_0(tE) = t\eta_0(E),$$

并对块内递推作归纳可得

$$\eta_{r+1}(tE) = (I - \alpha_r(tE)H)\eta_r(tE) = t(I - \alpha_r(E)H)\eta_r(E) = t\eta_{r+1}(E),$$

对所有 $r = 0, \cdots, p-1$ 成立。故

$$\mathcal{T}_q(tE) = t\mathcal{T}_q(E), \qquad \forall t > 0,\ \forall E \in \Omega \setminus \{0\}.$$

再结合假设中的 $\mathcal{T}_q(0) = 0$，即得 $\mathcal{T}_q$ 在整个 Ω 上都是一次正齐次的。

下面证明有界性，并顺便得到在原点处的连续性。记

$$L := 1 + \overline{\alpha}\left\|H\right\|.$$

则对任意 $E = (x_0, \cdots, x_p) \in \Omega \setminus \{0\}$，由块内递推有

$$\left\|\eta_{r+1}\right\| \leqslant \left\|I - \alpha_r(E)H\right\| \left\|\eta_r\right\| \leqslant L\left\|\eta_r\right\|, \qquad r = 0, \cdots, p-1.$$

归纳得到

$$\left\|\eta_r\right\| \leqslant L^r \left\|\eta_0\right\| = L^r \left\|x_p\right\| \leqslant L^r \left\|E\right\|_\infty, \qquad r = 0, \cdots, p.$$

因此

$$\left\|\mathcal{T}_q(E)\right\|_\infty = \max_{0 \leqslant r \leqslant p} \left\|\eta_r\right\| \leqslant L^p \left\|E\right\|_\infty.$$

对 $E = 0$ 上式显然仍成立，所以可取 $M := L^p$。

最后证连续性。对 $\Omega \setminus \{0\}$ 上的点，连续性直接来自步长函数 α_r 的连续性以及有限次代数运算的连续性。对原点，若 $E_k \to 0$ 且 $E_k \in \Omega$，则

$$\left\|\mathcal{T}_q(E_k)\right\|_\infty \leqslant M \left\|E_k\right\|_\infty \to 0,$$

故 $\mathcal{T}_q(E_k) \to 0 = \mathcal{T}_q(0)$。于是 $\mathcal{T}_q$ 在 Ω 上连续。 ∎

定理 2.4 设某固定记忆梯度法满足假设 2.1，其严格凸二次模型块映射 $\mathcal{T}_q$ 由（5）给出。若对任意初始块状态 $E \in \Omega$ 都有

$$\mathcal{T}_q^k(E) \to 0,$$

则存在常数 $C > 0$ 与 $\rho \in (0, 1)$，使得

$$\left\|\mathcal{T}_q^k(E)\right\|_\infty \leqslant C\rho^k \left\|E\right\|_\infty, \qquad \forall E \in \Omega,\ \forall k \geqslant 0.$$

进一步，原始迭代序列 $\{x_j\}$ 也在该状态空间上对应地 R-线性收敛到原点。

证明 由命题 2.3，$\mathcal{T}_q$ 是定义在闭锥 Ω 上的连续正齐次映射。再结合逐点收敛假设，可直接应用引理 2.2，得到统一有限步收缩与块映射的全局 R-线性收敛。

对原始迭代序列，任取整数 $j \geqslant 0$，写为

$$j = kp + r, \qquad 0 \leqslant r \leqslant p-1.$$

由前向块记号有 $x_j = x_{k,r}$，从而

$$\|x_j\| \leqslant \|E_k\|_\infty = \|\mathcal{T}_q^k(E_0)\|_\infty \leqslant C\rho^k \|E_0\|_\infty .$$

令 $\hat{\rho} = \rho^{1/p} \in (0,1)$，则 $\rho^k \leqslant \hat{\rho}^j/\rho$，于是

$$\|x_j\| \leqslant \frac{C}{\rho}\,\hat{\rho}^j \|E_0\|_\infty .$$

故原始迭代序列 R-线性收敛到原点。 ∎

第四节 紧子集上的局部收敛

对于许多行为更为复杂的算法，找到适用于该算法自然状态的闭锥有时也是困难的，但很多算法随着迭代的进行，尾部状态会永远限制在某个紧子集中。针对这类问题，我们在本小节给出一个紧子集诱导的局部化统一步有限收缩与 R-线性收敛命题。

命题 2.5 设 X 是有限维赋范空间，$\Omega \subset X \setminus \{0\}$ 为开锥，映射 $T : \Omega \cup \{0\} \to \Omega \cup \{0\}$ 连续，满足 $T(0) = 0$ 且

$$T(tx) = tT(x), \qquad \forall t \geqslant 0,\ \forall x \in \Omega.$$

记单位球面为 $\mathbb{S} := \{x \in X : \|x\| = 1\}$，并设 $K \subset \Omega \cap \mathbb{S}$ 为紧集。若

(i) 对任意 $z \in K$，只要 $T(z) \neq 0$，就有 $T(z)/\|T(z)\| \in K$；

(ii) 对每个 $z \in K$，都有 $T^k(z) \to 0$,

则存在常数 $C_K > 0$、$\rho_K \in (0,1)$，使得对所有属于锥

$$\mathrm{cone}(K) := \{tz :\ t \geqslant 0,\ z \in K\}$$

的初值 x 都有

$$\|T^k(x)\| \leqslant C_K \rho_K^k \|x\|, \qquad \forall k \geqslant 0.$$

从而还存在整数 $N \geqslant 1$ 与常数 $\theta \in (0,1)$ 使得

$$\|T^N(x)\| \leqslant \theta \|x\|, \qquad \forall x \in \mathrm{cone}(K).$$

证明 对任意 $z \in K$，由 $T^k(z) \to 0$ 可取整数 $n_z \geqslant 1$ 使

$$\|T^{n_z}(z)\| < \frac{1}{2}.$$

由于 T^{n_z} 连续，存在 z 在 K 中的开邻域 $U_z \subset K$，使得对所有 $y \in U_z$ 都有

$$\|T^{n_z}(y)\| < \frac{1}{2}.$$

$\{U_z\}_{z\in K}$ 构成 K 的开覆盖。由紧性，可取有限子覆盖 $U_{z_1}, \cdots, U_{z_s}$，并令

$$N_0 := \max_{1\leqslant i\leqslant s} n_{z_i}.$$

于是对任意 $y \in K$，都存在 $m(y) \in \{1, \cdots, N_0\}$ 使

$$\left\|T^{m(y)}(y)\right\| < \frac{1}{2}.$$

再令

$$M_0 := \max\left\{1,\ \sup_{z\in K}\|T(z)\|,\ \cdots,\ \sup_{z\in K}\left\|T^{N_0}(z)\right\|\right\} < \infty.$$

固定任意非零 $x \in \operatorname{cone}(K)$，令 $x = \|x\| z_0$ 且 $z_0 \in K$。若某一步已有 $T^{s_q}(x) = 0$，则后续估计显然成立；否则递归定义

$$z_q := \frac{T^{s_q}(x)}{\left\|T^{s_q}(x)\right\|}, \qquad s_q := \sum_{j=1}^{q} m_j,$$

其中 $m_j := m(z_{j-1}) \in \{1, \cdots, N_0\}$。由假设 (i)，所有 z_q 都属于 K。再由正齐次性，

$$\left\|T^{s_q}(x)\right\| = \left\|T^{m_q}\left(T^{s_{q-1}}(x)\right)\right\| = \left\|T^{m_q}(z_{q-1})\right\| \left\|T^{s_{q-1}}(x)\right\| \leqslant \frac{1}{2}\left\|T^{s_{q-1}}(x)\right\|.$$

归纳得到

$$\left\|T^{s_q}(x)\right\| \leqslant 2^{-q}\|x\|.$$

给定任意 $k \geqslant 1$，取唯一整数 $p \geqslant 0$ 使 $s_p \leqslant k < s_{p+1}$。由 $k - s_p < m_{p+1} \leqslant N_0$ 以及 $z_p \in K$，得到

$$\left\|T^{k}(x)\right\| = \left\|T^{k-s_p}\left(T^{s_p}(x)\right)\right\| \leqslant M_0\left\|T^{s_p}(x)\right\| \leqslant M_0 2^{-p}\|x\|.$$

另一方面，由 $m_j \leqslant N_0$ 可知

$$k < s_{p+1} \leqslant (p+1)N_0,$$

故 $p > k/N_0 - 1$。代回上式可得

$$\left\|T^{k}(x)\right\| \leqslant M_0 2^{-k/N_0+1}\|x\| = 2M_0\left(2^{-1/N_0}\right)^k\|x\|.$$

因此，只需令

$$C_K := 2M_0, \qquad \rho_K := 2^{-1/N_0} \in (0,1),$$

即可得到所述 R-线性估计。最后取 N 使 $C_K\rho_K^N < 1$，并令 $\theta := C_K\rho_K^N$，便得到统一 N 步收缩。 ∎

注 命题 2.5 可视作引理 2.2 的局部化版本：它不再要求整个 $\Omega\cap\mathbb{S}$ 都参与有限覆盖，只要求归一化轨道被困在一个紧子集 K 中。而定理 2.4 是一个要求更强的抽象模板：它要求所研究的步长在给定闭锥状态空间上处处有定义，并且可以诱导出闭锥上的连续自映射。对 BB 方法或带秩压缩的 LMSD / harmonic-LMSD 理想化算法的严格凸二次模型，这一点可以直接验证。而对这类方法的一般非线性模型，就需要结合命题 2.5 的局部化版本以及后文的局部扰动分析来使用，不能不加区分地直接套用定理 2.4。

第五节 一些例子

我们以固定步长梯度法及 BB 方法为例，给出上述定理的具体应用实例。

对于固定步长梯度法，我们熟知以下命题：

命题 2.6 考虑严格凸二次模型（4）以及固定步长梯度法

$$x_{k+1}=x_k-\alpha g_k,\qquad g_k=Hx_k, \tag{6}$$

其中常数步长满足

$$0<\alpha<\frac{2}{\lambda_n}. \tag{7}$$

则对任意初值 $x_0\in\mathbb{R}^n$，都有

$$x_k=(I-\alpha H)^k x_0,\qquad \|x_k\|\leqslant\delta(\alpha)^k\|x_0\|,$$

其中

$$\delta(\alpha):=\max_{1\leqslant i\leqslant n}\left|1-\alpha\lambda_i\right|\in(0,1).$$

因此，固定步长梯度法在欧氏范数下收敛到原点。

证明 由假设 1.1，不失一般性可设

$$H=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),\qquad 0<\lambda_1<\cdots<\lambda_n.$$

把 $x_k=(x_k^{(1)},\cdots,x_k^{(n)})^T$ 写成坐标形式，则（6）等价于

$$x_{k+1}^{(i)}=(1-\alpha\lambda_i)x_k^{(i)},\qquad i=1,\cdots,n.$$

于是

$$x_k^{(i)}=(1-\alpha\lambda_i)^k x_0^{(i)},\qquad i=1,\cdots,n,$$

也即

$$x_k=(I-\alpha H)^k x_0.$$

由（7）及 $0<\lambda_i\leqslant\lambda_n$，对每个 $i=1,\cdots,n$ 都有

$$-1<1-\alpha\lambda_i<1,$$

从而

$$\delta(\alpha) := \max_{1\leqslant i\leqslant n} \left|1-\alpha\lambda_i\right| < 1.$$

因此在欧氏范数下，

$$\left\|x_k\right\|^2 = \sum_{i=1}^{n} \left|x_k^{(i)}\right|^2 = \sum_{i=1}^{n} \left|1-\alpha\lambda_i\right|^{2k} \left|x_0^{(i)}\right|^2 \leqslant \delta(\alpha)^{2k} \sum_{i=1}^{n} \left|x_0^{(i)}\right|^2 = \delta(\alpha)^{2k} \left\|x_0\right\|^2.$$

取平方根即得

$$\left\|x_k\right\| \leqslant \delta(\alpha)^k \left\|x_0\right\|, \qquad k \geqslant 0.$$

故该方法对于任意初值逐点收敛。 ∎

若把 $T_\alpha(x) := (I-\alpha H)x$ 看成定义在闭锥 $\Omega = \mathbb{R}^n$ 上的映射，显然固定步长对 x 是连续零次的，则由命题 2.3 知 T_α 是连续一次正齐次的。命题 2.6 表明 $T_\alpha^k(x) \to 0$ 对每个初值都成立，因此应用引理 2.2 也可以再次推出统一有限步收缩和 R-线性收敛（事实上命题 2.6 已经表明了收敛是 Q-线性的，我们在此只以此为例子说明本节方法的实用性）。这说明本节的抽象工具在最基本的固定步长情形下已经完全适用。

再看 BB 方法，若把任意点对 $E = (x_0, x_1)$ 当作研究状态，则会出现并非真实 BB 轨道可达的人工状态：例如 $x_1 = x_0$ 但并未结束迭代，此时 $s = x_1 - x_0 = 0$，BB1/BB2 步长都没有定义（事实上 BB 方法要求第一个步长不为 0）。因此，此时更自然的做法不是把“点块”当作状态，而是使用连续两步梯度所构成的梯度对。

在二次模型（4）下，对任意非零梯度 g，记

$$\beta_{\mathrm{BB1}}(g) := \frac{g^T g}{g^T H g}, \qquad \beta_{\mathrm{BB2}}(g) := \frac{g^T H g}{g^T H^2 g}. \tag{8}$$

它们分别等于 BB1 与 BB2 由当前位移 $s = -\alpha g$ 和梯度差 $y = -\alpha H g$ 计算出的下一步步长。易知对任意 $g \neq 0$，都有

$$\frac{1}{\lambda_n} \leqslant \beta_{\mathrm{BB1}}(g),\, \beta_{\mathrm{BB2}}(g) \leqslant \frac{1}{\lambda_1}. \tag{9}$$

固定一种选择 $\beta \in \{\beta_{\mathrm{BB1}}, \beta_{\mathrm{BB2}}\}$。定义锥

$$\Omega_{\mathrm{BB}} := \left\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n : \exists \alpha \in \left[\frac{1}{\lambda_n}, \frac{1}{\lambda_1}\right] \text{使} v = (I-\alpha H)u\right\}. \tag{10}$$

若普通 BB 方法在某一步之后进入标准递推，即 $g_{k+1} = (I-\alpha_k H)g_k$ 且 $\alpha_{k+1} = \beta(g_k)$，则梯度对

$$Z_k := (g_k, g_{k+1})$$

满足

$$Z_{k+1} = \mathcal{T}_{\mathrm{BB}}(Z_k).$$

其中

$$\mathcal{T}_{\mathrm{BB}}(u,v) := \begin{cases} \big(v,(I-\beta(u)H)v\big), & (u,v)\in\Omega_{\mathrm{BB}},\ u\neq 0, \\ (0,0), & (u,v)=(0,0). \end{cases} \tag{11}$$

命题 2.7 Ω_{BB} 是 $(\mathbb{R}^n)^2$ 中的闭锥，且映射 $\mathcal{T}_{\mathrm{BB}}:\Omega_{\mathrm{BB}}\to\Omega_{\mathrm{BB}}$ 连续、一次正齐次。

证明 先证 Ω_{BB} 是闭锥。若 $(u,v)\in\Omega_{\mathrm{BB}}$，则存在 $\alpha\in[1/\lambda_n,1/\lambda_1]$ 使 $v=(I-\alpha H)u$，从而对任意 $t\geqslant 0$ 都有

$$tv=(I-\alpha H)(tu),$$

故 $t(u,v)\in\Omega_{\mathrm{BB}}$。再设 $(u_k,v_k)\in\Omega_{\mathrm{BB}}$ 且 $(u_k,v_k)\to(u,v)$。对每个 k，取 $\alpha_k\in[1/\lambda_n,1/\lambda_1]$ 使

$$v_k=(I-\alpha_k H)u_k.$$

由区间紧性可取子列（仍记为 α_k），使 $\alpha_k\to\alpha\in[1/\lambda_n,1/\lambda_1]$。取极限得

$$v=(I-\alpha H)u,$$

于是 $(u,v)\in\Omega_{\mathrm{BB}}$，故 Ω_{BB} 闭。

再证 $\mathcal{T}_{\mathrm{BB}}$ 是自映射。若 $(u,v)\in\Omega_{\mathrm{BB}}$ 且 $u\neq 0$，则由（9）知 $\beta(u)\in[1/\lambda_n,1/\lambda_1]$。于是

$$(I-\beta(u)H)v$$

恰好仍由第一分量 v 经过某个允许步长的一次二次梯度更新得到，从而 $\mathcal{T}_{\mathrm{BB}}(u,v)\in\Omega_{\mathrm{BB}}$。

由（8）可知 $\beta(tu)=\beta(u)$（$t>0$），于是对任意 $t\geqslant 0$ 都有

$$\mathcal{T}_{\mathrm{BB}}(tu,tv)=t\mathcal{T}_{\mathrm{BB}}(u,v),$$

即 $\mathcal{T}_{\mathrm{BB}}$ 一次正齐次。它在 $u\neq 0$ 的点处连续是显然的。对原点，令

$$M_0 := \max_{\alpha\in[1/\lambda_n,\,1/\lambda_1]} \|I-\alpha H\| < \infty.$$

若 $(u,v)\in\Omega_{\mathrm{BB}}$，则存在某个 $\alpha\in[1/\lambda_n,1/\lambda_1]$ 使 $v=(I-\alpha H)u$，故

$$\|v\|\leqslant M_0\|u\|.$$

又由（9），$\beta(u)\in[1/\lambda_n,1/\lambda_1]$，从而

$$\|(I-\beta(u)H)v\|\leqslant M_0\|v\|.$$

因此当 $(u,v)\to(0,0)$ 且 $(u,v)\in\Omega_{\mathrm{BB}}$ 时，$\mathcal{T}_{\mathrm{BB}}(u,v)\to(0,0)$。故 $\mathcal{T}_{\mathrm{BB}}$ 在 Ω_{BB} 上连续。 ∎

推论 2.8 (BB 方法的 R-线性收敛) 考虑严格凸二次模型（4）上的 BB1 或 BB2 方法。记经过一次预热之后的连续梯度对为

$$Z_k := (g_k, g_{k+1}), \qquad k \geqslant 1.$$

若方法不在有限步内终止，则 $Z_{k+1} = \mathcal{T}_{\mathrm{BB}}(Z_k)$，且由 Raydan 的经典收敛定理可知 $g_k \to 0$[3]，从而 $Z_k \to 0$。因此，将引理 2.2 用于闭锥 Ω_{BB} 与映射 $\mathcal{T}_{\mathrm{BB}}$，便得到常数 $C > 0$ 与 $\rho \in (0,1)$，使

$$\|Z_k\| \leqslant C\rho^k \|Z_1\|, \qquad k \geqslant 1.$$

特别地，由于 g_k 是梯度对 Z_k 的第一分量，故梯度序列 $\{g_k\}$ 也是 R-线性收敛的。这与 Dai–Liao 的经典结论一致[7]。

注 推论 2.8 说明第 2 章的抽象工具并不要求状态一定取为点块 E。对 BB 而言，最自然的状态是梯度对，而对后文的所讨论的方法，有时选择点块状态 E_k 更为方便。真正重要的是所选状态需要在算法递推下闭合，并且能够构造一个适合应用引理 2.2 的闭锥。

从上述过程可以看出，该证明框架也适用于周期长度固定的 CBB 方法[10]以及[6] 中部分情况下（如 $\rho(k)$ 及 $\nu(k)$ 固定或具有周期性）算法的 R-线性收敛性证明。

第三章　LMSD 的高维最小二乘解释与子空间结构

本章讨论严格凸二次模型中的 LMSD 方法的步长结构。首先回顾 Fletcher 对 Ritz-sweep 的原始构造[11]以及 Ferrandi–Hochstenbach 关于 LMSD 方法的割线最小二乘解释[19]，系统罗列并简要证明一些重要的性质与定理。在此基础上，把本文后续所需的代数对象写成统一块映射形式，为第 4 章的收敛性证明以及第 6 章的局部分析奠定基础。

第一节　BB1/BB2 的最小二乘解释

设

$$s_{k,0} = x_{k,1} - x_{k,0}, \qquad y_{k,0} = g_{k,1} - g_{k,0}.$$

BB1 与 BB2 分别可以由两个一维最小二乘问题

$$\eta_{k,0}^{\mathrm{BB1}} := \arg\min_{\eta\in\mathbb{R}} \left\|\eta s_{k,0} - y_{k,0}\right\|^2 = \frac{s_{k,0}^T y_{k,0}}{s_{k,0}^T s_{k,0}}, \qquad \alpha_k^{\mathrm{BB1}} = \frac{1}{\eta_{k,0}^{\mathrm{BB1}}}, \tag{12}$$

以及

$$\alpha_k^{\mathrm{BB2}} := \arg\min_{\alpha\in\mathbb{R}} \left\|s_{k,0} - \alpha y_{k,0}\right\|^2 = \frac{s_{k,0}^T y_{k,0}}{y_{k,0}^T y_{k,0}}. \tag{13}$$

导出。因此，BB1 是“用一个标量拟合 Hessian”的最小二乘解，而 BB2 是“用一个标量拟合逆 Hessian”的最小二乘解。这正是推广到更高维的自然出发点。

第二节　p 阶矩阵的最小二乘推广

文献[19] 中的 Propositions 3 (i)(ii) 给出了 LMSD 的最小二乘逼近解释：固定正整数 p，在第 k 个块起点，记最近 p 个割线对为

$$S_k = [s_{k,0}, s_{k,1}, \cdots, s_{k,p-1}] \in \mathbb{R}^{n\times p}, \qquad Y_k = [y_{k,0}, y_{k,1}, \cdots, y_{k,p-1}] \in \mathbb{R}^{n\times p}. \tag{14}$$

相应地定义

$$M_k = S_k^T S_k, \qquad B_k = S_k^T Y_k, \qquad D_k = Y_k^T Y_k. \tag{15}$$

把（12）与（13）中的标量分别提升为 $p\times p$ 矩阵，可得到原始矩阵最小二乘模型

$$M_k^{\mathrm{LS}} := \arg\min_{M\in\mathbb{R}^{p\times p}} \left\|Y_k - S_k M\right\|_F^2, \qquad N_k^{\mathrm{LS}} := \arg\min_{N\in\mathbb{R}^{p\times p}} \left\|S_k - Y_k N\right\|_F^2. \tag{16}$$

当 M_k 与 D_k 可逆时，其正规方程给出

$$M_k^{\mathrm{LS}} = M_k^{-1} B_k, \qquad N_k^{\mathrm{LS}} = D_k^{-1} B_k^T. \tag{17}$$

命题 3.1 当 $p=1$ 时，（17）精确退化为经典 BB1/BB2：

$$M_k^{\mathrm{LS}}=\frac{s_{k,0}^T y_{k,0}}{s_{k,0}^T s_{k,0}}=\frac{1}{\alpha_k^{\mathrm{BB1}}},\qquad N_k^{\mathrm{LS}}=\frac{s_{k,0}^T y_{k,0}}{y_{k,0}^T y_{k,0}}=\alpha_k^{\mathrm{BB2}}.$$

证明 此时 $S_k=s_{k,0}$、$Y_k=y_{k,0}$ 都是列向量，（17）直接退化为标量等式。∎

上述内容表明：经典 BB 与更高维的 pD-BB 步长可以放在同一个“矩阵最小二乘谱拟合”框架下理解。

第三节　严格凸二次情形下与原始公式的等价性

由假设 1.1，严格凸二次情形都可在标准形

$$q(x)=\frac{1}{2}x^T Hx,\qquad H=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),\qquad 0<\lambda_1<\cdots<\lambda_n \tag{18}$$

下讨论。此时 $Y_k=HS_k$，于是[19]中的 Propositions 3 (iii) 给出矩阵最小二乘模型与子空间 Ritz 值之间的严格等价关系。

定理 3.2 设 q 由（18）给出。令 $S_k=Q_kR_k$ 为 S_k 的薄 QR 分解，$Y_k=W_kT_k$ 为 Y_k 的薄 QR 分解，则

$$M_k^{\mathrm{LS}}=R_k^{-1}(Q_k^THQ_k)R_k,\qquad N_k^{\mathrm{LS}}=T_k^{-1}(W_k^TH^{-1}W_k)T_k. \tag{19}$$

因此：

(i) M_k^{LS} 的特征值精确等于 H 在子空间 $\mathrm{span}(S_k)$ 上的 Ritz 值；

(ii) N_k^{LS} 的特征值精确等于 H^{-1} 在子空间 $\mathrm{span}(Y_k)$ 上的 Ritz 值。

特别地，当 $p=2,3$ 时，最小二乘推导得到的 p 维步长与[20]中原始 2D/3D 公式完全一致。

证明 由 $Y_k=HS_k$ 与 $S_k=Q_kR_k$ 得

$$M_k=R_k^TR_k,\qquad B_k=S_k^THS_k=R_k^T(Q_k^THQ_k)R_k.$$

于是

$$M_k^{\mathrm{LS}}=(R_k^TR_k)^{-1}R_k^T(Q_k^THQ_k)R_k=R_k^{-1}(Q_k^THQ_k)R_k.$$

同理，由 $S_k=H^{-1}Y_k$ 与 $Y_k=W_kT_k$ 得

$$D_k=T_k^TT_k,\qquad B_k^T=Y_k^TH^{-1}Y_k=T_k^T(W_k^TH^{-1}W_k)T_k,$$

从而

$$N_k^{\mathrm{LS}}=(T_k^TT_k)^{-1}T_k^T(W_k^TH^{-1}W_k)T_k=T_k^{-1}(W_k^TH^{-1}W_k)T_k.$$

因此结论成立。∎

定理 3.2 说明，矩阵最小二乘是在严格凸二次情形下对[20]中原始公式的另一种推导方式。

第四节　二次模型中的块步长定义及其与 LMSD 的对应关系

我们基于[19]给出相应块步长的定义，作为后续讨论的基础。我们首先回顾[19]中的 Propositions 1–2。

命题 3.3　设 $q(x)=\frac{1}{2}x^{T}Hx$ 为严格凸二次函数，考虑某个固定块中的 p 步梯度迭代。记

$$G_k=[g_{k,0},g_{k,1},\cdots,g_{k,p-1}],\qquad S_k=[s_{k,0},s_{k,1},\cdots,s_{k,p-1}],$$
$$Y_k=[y_{k,0},y_{k,1},\cdots,y_{k,p-1}],$$

以及由块首梯度生成的 Krylov 矩阵

$$K_k=[g_{k,0},Hg_{k,0},\cdots,H^{p-1}g_{k,0}].$$

则有

$$\operatorname{span}(G_k)=\operatorname{span}(S_k)=\mathcal{K}_p(H,g_{k,0}),\qquad \operatorname{span}(Y_k)=H\operatorname{span}(G_k)=H\mathcal{K}_p(H,g_{k,0}).$$

因而无论从 G_k、S_k 还是 K_k 出发，得到的 Ritz 值都只由子空间 $\mathcal{K}_p(H,g_{k,0})$ 决定。同理，Y_k 上的 Y 型的 Ritz 值只由 $H\mathcal{K}_p(H,g_{k,0})$ 决定。

证明　由二次递推

$$g_{k,r+1}=(I-\alpha_{k,r}H)g_{k,r},\qquad r=0,\cdots,p-2,$$

可递归得到：对每个 $r=0,\cdots,p-1$，存在次数恰为 r 且最高次项系数非零的多项式 $P_{k,r}$，使得

$$g_{k,r}=P_{k,r}(H)g_{k,0}.$$

因此存在一个上三角且可逆的矩阵 $U_k\in\mathbb{R}^{p\times p}$ 使

$$G_k=K_kU_k,$$

从而 $\operatorname{span}(G_k)=\operatorname{span}(K_k)=\mathcal{K}_p(H,g_{k,0})$。

又因为每个块内位移都满足

$$s_{k,r}=-\alpha_{k,r}g_{k,r},\qquad \alpha_{k,r}>0,$$

故 $S_k=-G_k\Delta_k$，其中 $\Delta_k=\operatorname{diag}(\alpha_{k,0},\cdots,\alpha_{k,p-1})$ 可逆，于是 $\operatorname{span}(S_k)=\operatorname{span}(G_k)$。最后，

$$y_{k,r}=g_{k,r+1}-g_{k,r}=-\alpha_{k,r}Hg_{k,r},$$

从而 $Y_k=-HG_k\Delta_k$，故 $\operatorname{span}(Y_k)=H\operatorname{span}(G_k)=H\mathcal{K}_p(H,g_{k,0})$。

Ritz 值与 harmonic Ritz 值是子空间不变量，一旦上述列空间确定，它们与采用哪一组基无关。 ∎

引理 3.4 在假设 1.1 的标准形下，设

$$S(u) := \{i : u_i \neq 0\}, \qquad s(u) := |S(u)|,$$

即 $s(u)$ 是当前梯度所含的活跃不同特征值个数。则

$$\operatorname{rank}[u, Hu, \cdots, H^{p-1}u] = \min\{p, s(u)\}.$$

进一步，若同一块内的连续梯度由

$$g_{r+1} = (I - \alpha_r H)g_r, \qquad \alpha_r > 0,$$

生成，并令

$$G = [g_0, \cdots, g_{p-1}], \qquad S = [s_0, \cdots, s_{p-1}],$$

则

$$\operatorname{rank} G = \operatorname{rank} S = \operatorname{rank}[g_0, Hg_0, \cdots, H^{p-1}g_0] = \min\{p, s(g_0)\}.$$

因此，在精确严格凸二次模型和连续块历史下，只要 $s(g_0) \geqslant p$，历史束就不可能精确降到小于 p 维，少于 p 个 Ritz 步长只会在 $s(g_0) < p$ 时出现。

证明 在标准坐标下，

$$[u, Hu, \cdots, H^{p-1}u] = \operatorname{diag}(u_i)_{i\in S(u)} \begin{bmatrix} 1 & \lambda_i & \cdots & \lambda_i^{p-1} \end{bmatrix}_{i\in S(u)}.$$

去掉非零对角因子后，右端是由互异 λ_i 生成的 Vandermonde 矩阵，列秩为 $\min\{p, s(u)\}$。对连续块历史，由命题 3.3 的证明可知 $G = K_p(H, g_0)U$，其中 U 为可逆上三角矩阵；又 $S = -G\operatorname{diag}(\alpha_0, \cdots, \alpha_{p-1})$，而步长均为正，故三者秩相同。 ∎

注 (重复 Ritz 值的使用方式) 本文中的 "p 个 Ritz 值" 均按重数计数。若某一轮标准 p 维小问题产生两个相同的 Ritz 值，则相应倒数步长也在 sweep 中使用两遍。

我们给出二次问题的 S 型和 Y 型方法及块步长定义。

定义 3.1 (二次 S 型 Ritz-sweep) 设当前完整块的历史子空间为

$$\mathcal{S}_k := \operatorname{span}(S_k) = \mathcal{K}_p(H, g_{k,0}), \qquad d_k := \dim \mathcal{S}_k \leqslant p.$$

记 H 在 $\mathcal{S}_k$ 上的 Ritz 值为

$$\theta_{k,0}, \theta_{k,1}, \cdots, \theta_{k,d_k-1},$$

其中 Ritz 值按重数计数。当 $d_k = p$ 时，下一完整块依次使用 p 个步长

$$\alpha_{k+1,r}^{(S,q)} = \theta_{k,r}^{-1}, \qquad r = 0, \cdots, p-1.$$

当 $d_k < p$ 时，算法只在实际子空间上执行 d_k 步短 sweep。

定义 3.2 (二次 Y 型 Ritz-sweep)　设当前完整块的历史梯度差子空间为

$$\mathcal{Y}_k := \operatorname{span}(Y_k) = H\mathcal{K}_p(H, g_{k,0}), \qquad d_k := \dim \mathcal{Y}_k \leqslant p.$$

记 H^{-1} 在 $\mathcal{Y}_k$ 上的 Ritz 值为

$$\tau_{k,0}, \tau_{k,1}, \cdots, \tau_{k,d_k-1},$$

其中 Ritz 值按重数计数。当 $d_k = p$ 时，下一完整块依次使用 p 个步长

$$\alpha_{k+1,r}^{(Y,q)} = \tau_{k,r}, \qquad r = 0, \cdots, p-1.$$

当 $d_k < p$ 时，算法只在实际子空间上执行 d_k 步短 sweep。

注　上述 S 型与 Y 型方法分别对应带秩压缩的 LMSD 以及 harmonic-LMSD 方法。

第四章　二次模型的有限终止、逐点收敛与 R-线性收敛

本章进入本文的二次问题核心部分，在前一章的统一记号下，分别讨论有限终止、收敛性以及 R-线性收敛性质。由假设 1.1，本章所有严格凸二次问题证明都在标准形

$$q(x)=\frac{1}{2}x^{T}Hx,\qquad H=\operatorname{diag}(\lambda_1,\cdots,\lambda_n),\qquad 0<\lambda_1<\cdots<\lambda_n,$$

下进行，其中 $\lambda_1,\cdots,\lambda_n$ 是关于初始梯度活跃的不同特征值。

第一节　pD-BB 梯度法的有限终止定理

Yuan 在 $p=2,3$ 的低维严格凸二次问题上给出了有限终止性质[20]。更一般地，Fletcher 对 LMSD 方法的讨论，以及 Curtis–Guo 对 LMSD 的回顾，都说明了同一个基本机制：一旦一个块已经识别出全部活跃特征值，那么后续用这些特征值的倒数作步长就会构成消去多项式[11,13]。本文在这一小节用当前的记号给出一个简单证明。

引理 4.1　在上述假设下，H 关于 g_0 的极小多项式为

$$\pi(t)=\prod_{i=1}^{n}(t-\lambda_i),$$

其次数恰好为 n。因此

$$\dim\mathcal{K}_n(H,g_0)=n.$$

证明　由于 H 已经是对角矩阵，对任意多项式 P 都有

$$P(H)g_0=\left(P(\lambda_1)g_0^{(1)},\cdots,P(\lambda_n)g_0^{(n)}\right)^{T}.$$

因为 $g_0^{(i)}\neq 0$，故 $P(H)g_0=0$ 当且仅当 $P(\lambda_i)=0$ 对所有 $i=1,\cdots,n$ 成立。次数最低的首一多项式正是上式中的 $\pi(t)$。极小多项式次数等于 Krylov 子空间维数，故 $\dim\mathcal{K}_n(H,g_0)=n$。 ∎

引理 4.2　对严格凸二次函数 $q(x)=\frac{1}{2}x^{T}Hx$，梯度迭代 $x_{k+1}=x_k-\alpha_k g_k$ 满足

$$g_{k+1}=(I-\alpha_k H)g_k.$$

因此对每个 $j\geqslant 0$，都存在次数为 j 且最高次项系数非零的多项式 P_j，使得

$$g_j=P_j(H)g_0,\qquad P_j(t)=\prod_{\ell=0}^{j-1}(1-\alpha_\ell t).$$

证明 由 $g_k = Hx_k$ 及 $x_{k+1} = x_k - \alpha_k g_k$ 可得

$$g_{k+1} = Hx_{k+1} = Hx_k - \alpha_k Hg_k = (I - \alpha_k H)g_k.$$

反复迭代即可得到所述多项式表示。 ∎

引理 4.3

$$\mathcal{K}_n(H, g_0) = \mathrm{span}\{e_1, \cdots, e_n\} = \mathbb{R}^n,$$

其中 e_i 是标准基向量。特别地，整个空间 $\mathbb{R}^n$ 就是 H 的不变子空间，而 H 在该空间上的谱恰好是全部活跃特征值 $\{\lambda_1, \cdots, \lambda_n\}$。

证明 由极小多项式次数等于 n 且活跃特征值互异，向量组 $\{g_0, Hg_0, \cdots, H^{n-1}g_0\}$ 线性无关并张成整个 $\mathbb{R}^n$ 空间（$[g_0, Hg_0, \cdots, H^{n-1}g_0]$ 构成可逆范德蒙矩阵），于是等号成立。由于 H 在标准基下对角，其在 $\mathbb{R}^n$ 上的谱自然是 $\{\lambda_1, \cdots, \lambda_n\}$。 ∎

引理 4.4 对任意正的预热步长，前 n 个预热梯度与位移满足

$$\mathcal{K}_n(H, g_0) = \mathrm{span}\{g_0, g_1, \cdots, g_{n-1}\} = \mathrm{span}\{s_0, s_1, \cdots, s_{n-1}\}.$$

特别地，当历史窗口 $p \geqslant n$ 时，第一个完整块的历史位移矩阵满足

$$\mathrm{span}(S_0) = \mathcal{K}_n(H, g_0) = \mathbb{R}^n.$$

证明 由引理 4.2，对 $j = 0, \cdots, n-1$ 有 $g_j = P_j(H)g_0$，其中 P_j 的次数为 j 且最高次项系数非零，故 $\mathrm{span}\{g_0, \cdots, g_{n-1}\} = \mathcal{K}_n(H, g_0)$。又由于 $s_j = -\alpha_j g_j$ 且 $\alpha_j > 0$，位移与对应梯度共线，于是 $\mathrm{span}\{s_0, \cdots, s_{n-1}\} = \mathcal{K}_n(H, g_0)$。当 $p \geqslant n$ 时，第一个完整块的历史位移矩阵 S_0 的列空间就等于上述 Krylov 子空间，因此 $\mathrm{span}(S_0) = \mathbb{R}^n$。 ∎

定理 4.5 (有限终止定理) 设目标函数满足（18）且已规约至标准形，活跃的不同特征值个数为 n，历史窗口 $p \geqslant n$。若前 p 个预热步长任意给定且均为正，则在第一个完整块开始时，历史子空间恰为 $\mathcal{K}_n(H, g_0) = \mathbb{R}^n$。在精确秩压缩下，$S$ 型方法给出全部 n 个活跃特征值，Y 型方法给出这些活跃特征值的倒数，因此下一次 sweep 实际使用的步长集合在两种情形下都恰为 $\{\lambda_1^{-1}, \cdots, \lambda_n^{-1}\}$。随后再经过 n 步迭代，梯度变为零。换言之，算法在至多 $p + n \leqslant 2p$ 步内有限终止。

证明 由引理 4.4，当 $p \geqslant n$ 时，第一个完整块的历史子空间满足

$$\mathrm{span}(S_0) = \mathcal{K}_n(H, g_0) = \mathbb{R}^n.$$

对 S 型，由定义 3.1，第一个完整块提取到的 Ritz 值正是 H 在全空间 $\mathbb{R}^n$ 上的特征值，因此恰为 $\lambda_1, \cdots, \lambda_n$，下一块实际使用的步长便是 $\lambda_1^{-1}, \cdots, \lambda_n^{-1}$。

对 Y 型，由定义 3.2，第一个完整块提取到的 Ritz 值正是 H^{-1} 在全空间 $\mathbb{R}^n$ 上的特征值，也就是 $\lambda_1^{-1},\cdots,\lambda_n^{-1}$，而按定义，这些特征值本身就被直接用作下一块的步长。

于是两种情形下，预热结束后的当前梯度 g_p 都属于 $\mathbb{R}^n$，且接下来的 n 步都依次使用步长 $\lambda_1^{-1},\cdots,\lambda_n^{-1}$。故有

$$g_{p+n}=\prod_{j=1}^{n}\Big(I-\frac{1}{\lambda_j}H\Big)g_p=\frac{(-1)^n}{\prod_{j=1}^{n}\lambda_j}\,\pi(H)g_p,$$

其中 π 是引理 4.1 中的极小多项式。由于 $\pi(H)=0$，必有 $g_{p+n}=0$。故算法在至多 $p+n$ 步内终止。 ∎

推论 4.6 (i) 若标准形问题维数 $n=2$（原问题具有 2 个活跃谱），取 $p=2$，则算法在至多 4 步内终止；

(ii) 若标准形问题维数 $n=3$（原问题具有 3 个活跃谱），取 $p=3$，则算法在至多 6 步内终止。

（注：若原问题存在谱重数或初始梯度缺失某些特征分量，其等效的活跃谱个数 n 将小于原矩阵维数，算法会相应地更早终止。）

定理 4.5 清楚地表明：有限终止本质上来自“前 p 步识别出有效 Krylov 子空间，后 n 步用该子空间上的全部活跃特征值构造消去多项式”。特别地，若能够事先判断活跃谱个数并取 $p=n$，则算法可在至多 $2n$ 步内终止。

第二节　二次模型上的逐点收敛

本节证明严格凸二次问题精确秩压缩模型上的逐点收敛：从任意初始块状态出发，要么算法有限步终止，要么块起点梯度 $g_{k,0}$ 沿活跃谱逐分量趋于零。下一节再借助第二章的闭锥连续正齐次原理，把这一逐点收敛升级为 R-线性收敛。

按照精确秩压缩的二次模型算法，固定块长 $p\times p$ 小广义特征值公式只在 $K_p(H,u)$ 满列秩（即 $s(u)>p$）时使用，当 $s(u)\leqslant p$ 时，算法在实际 Krylov 子空间上提取按重数计的 $s(u)$ 个 Ritz 值并执行短 sweep。由引理 3.4，精确二次模型中的秩下降只能来自于剩余活跃谱已经不超过 p，亦即下面命题的有限终止分支。

命题 4.7 (正则区域外的有限终止)　设 q 满足（18），并考虑精确秩压缩的 S 型或 Y 型块步长。若某个块起点 $g_{k,0}$ 满足

$$m:=s(g_{k,0})\leqslant p,$$

则从该块起点出发，算法在至多 m 个有效步长后达到 $g=0$。因此，任何非有限终止的轨道必然在每个非零块起点满足 $s(g_{k,0})>p$。

证明　若 $m=0$，结论显然成立。设 $m\geqslant 1$，记 $J=S(g_{k,0})$。由引理 3.4，实

际 Krylov 子空间维数为 m，且

$$\mathcal{K}_m(H, g_{k,0}) = \operatorname{span}\{e_i : i \in J\}.$$

对 S 型方法，H 在该不变子空间上的 Ritz 值恰为 $\{\lambda_i : i \in J\}$；对 Y 型方法，H^{-1} 在 $H\,\mathcal{K}_m(H, g_{k,0}) = \operatorname{span}\{e_i : i \in J\}$ 上的 Ritz 值恰为 $\{\lambda_i^{-1} : i \in J\}$，而它们正是 Y 型直接采用的步长。因此无论 S 型还是 Y 型，下一短 sweep 的步长集合都是 $\{\lambda_i^{-1} : i \in J\}$。于是

$$\prod_{i \in J} \Big(I - \frac{1}{\lambda_i} H \Big) g_{k,0} = 0,$$

因为对每个仍活跃的第 i 个分量，乘子中含有因子 $1 - \lambda_i/\lambda_i = 0$，而非活跃分量本来就为零。 ∎

命题 4.7 说明，在精确严格凸二次模型上不存在“真实 Krylov 维数下降但算法仍需继续做渐近分析”的第三种情形：秩下降等价于剩余活跃谱已经不超过 p，对应有限终止分支。因此，真正需要做渐近收敛的分析对象只有 $s(g_{k,0}) > p$ 的正则非终止情形。此时每个 sweep 都恰好给出 p 个 Ritz 值。

一、几个引理

我们先收集后续证明所需的工具。第一个引理是 Fletcher[11] 附录中证明 LMSD 收敛性时使用的两步收缩判据，第二、第三个引理把块乘子的行列式比表示与 Cauchy–Binet 加权展开抽象成对所有 S、Y、ω 三种步长统一适用的代数模板，第四个引理是“乘子权重引理”，它把这一加权平均的极限控制问题转化为非负权重相对比的阶数比较，是逐点收敛证明的关键。

引理 4.8 (Fletcher 两步引理)　设 $\{a_k\}_{k \geqslant 0}$ 为非负实数列，并存在常数 $M \geqslant 1$ 使

$$a_{k+2} \leqslant M a_{k+1}, \qquad k \geqslant 0.$$

若对任意 $\varepsilon > 0$ 都存在常数 $c_\varepsilon \in (0, 1)$ 与整数 k_ε，使得对一切 $k \geqslant k_\varepsilon$ 有

$$a_k \geqslant \varepsilon \Longrightarrow a_{k+2} \leqslant c_\varepsilon a_{k+1},$$

则 $a_k \to 0$。

证明　任取 $\varepsilon > 0$。

第一步：存在 $N \geqslant k_\varepsilon$ 使 $a_N < \varepsilon$。 若不然，对一切 $k \geqslant k_\varepsilon$ 有 $a_k \geqslant \varepsilon$，于是题设给出 $a_{k+2} \leqslant c_\varepsilon a_{k+1}$，递推得到对每个 $j \in \mathbb{N}$

$$a_{k_\varepsilon + j + 1} \leqslant c_\varepsilon^{\,j}\, a_{k_\varepsilon + 1}.$$

由 $0 < c_\varepsilon < 1$，存在 j_0 使 $a_{k_\varepsilon + j_0 + 1} < \varepsilon$，矛盾。故集合 $\{n \geqslant k_\varepsilon : a_n < \varepsilon\}$ 非空。令 $N_1 := \min\{n \geqslant k_\varepsilon : a_n < \varepsilon\}$。

第二步：对 $n \geqslant N_1$ 有 $a_n < M^2\varepsilon$。 对 n 归纳。$n = N_1$ 时 $a_{N_1} < \varepsilon \leqslant M^2\varepsilon$ 成立。$n = N_1 + 1$ 时由全局递推 $a_{N_1+1} \leqslant M a_{N_1} < M\varepsilon \leqslant M^2\varepsilon$ 成立。设 $N_1 \leqslant n \leqslant m$ 时结论已成立，$m \geqslant N_1 + 1$；考察 a_{m+1}。

- 若 $a_{m-1} < \varepsilon$，则 $a_{m+1} \leqslant M a_m \leqslant M^2 a_{m-1} < M^2\varepsilon$；
- 若 $a_{m-1} \geqslant \varepsilon$（注意 $m-1 \geqslant N_1 \geqslant k_\varepsilon$），由条件 $a_{m+1} \leqslant c_\varepsilon a_m$，归纳假设 $a_m < M^2\varepsilon$ 与 $c_\varepsilon < 1$ 给出 $a_{m+1} < c_\varepsilon M^2\varepsilon < M^2\varepsilon$。

归纳完成。由 ε 任意小，得 $a_k \to 0$。 ∎

引理 4.9 (块乘子的行列式比表示) 设 $K \in \mathbb{R}^{n\times p}$ 列满秩，$A, B \in \mathbb{R}^{p\times p}$ 是对称正定矩阵，且广义特征值问题 $Av = \theta Bv$ 的 p 个特征值 $\theta_0, \cdots, \theta_{p-1}$ 全为正。则对任意 $t \in \mathbb{R}$，

$$\prod_{\ell=0}^{p-1}\Big(1 - \frac{t}{\theta_\ell}\Big) = \frac{\det(A - tB)}{\det(A)}. \tag{20}$$

特别地，对当前完整块的 Krylov 历史矩阵 $K = K_p(H, u)$（$u = g_{k,0}$，且 $s(u) > p$ 故 K 列满秩），块乘子有如下行列式比表示：

(1) S 型 (LMSD)，记 $\theta_\ell = \theta_\ell(u)$ 为 H 在 span(K) 上的 Ritz 值，则

$$\Psi_u^{(S)}(t) := \prod_{\ell=0}^{p-1}\Big(1 - \frac{t}{\theta_\ell}\Big) = \frac{\det\big(K^T(H - tI)K\big)}{\det\big(K^T H K\big)}; \tag{21}$$

(2) Y 型 (harmonic-LMSD)，记 $\tau_\ell = \tau_\ell(u)$ 为 H^{-1} 在 H span(K) 上的 Ritz 值（即 Y 型直接采用的步长），则

$$\Phi_u^{(Y)}(t) := \prod_{\ell=0}^{p-1}\big(1 - \tau_\ell t\big) = \frac{\det\big(K^T(H^2 - tH)K\big)}{\det\big(K^T H^2 K\big)}; \tag{22}$$

(3) ω 加权（参见 §四），记 θ_ℓ^ω 为问题（38）的根，则

$$\Psi_u^{(\omega)}(t) := \prod_{\ell=0}^{p-1}\Big(1 - \frac{t}{\theta_\ell^\omega}\Big) = \frac{\det\big(K^T\omega(H)(H - tI)K\big)}{\det\big(K^T\omega(H)HK\big)}. \tag{23}$$

证明 对一般式（20）：由于 A 正定可逆，

$$\det(A - tB) = \det(A)\det\big(I - tA^{-1}B\big) = \det(A)\prod_{\ell=0}^{p-1}(1 - t\,\mu_\ell),$$

其中 μ_ℓ 为 $A^{-1}B$ 的特征值。由 $Av = \theta Bv$ 等价于 $A^{-1}Bv = \theta^{-1}v$，故 $\mu_\ell = 1/\theta_\ell$，得到（20）。

对（21）：取 $A = K^T H K$、$B = K^T K$，则 θ_ℓ 满足 $K^T H K v = \theta\, K^T K v$，是 H 在 span(K) 上的 Ritz 值，由 $H \succ 0$ 全为正及（20）即得。

对（22）：Y 型乘子 $\Phi_u^{(Y)}(t) = \prod(1 - \tau_\ell t)$ 在 $t = \tau_\ell^{-1}$ 处取零，即把 $\xi = \tau_\ell^{-1}$ 视为根。由 H^{-1} 在 H span(K) 上的 Ritz 值，τ_ℓ 是 $K^T H K w = \tau\, K^T H^2 K w$ 的根，等

价地 $\xi=\tau_\ell^{-1}$ 是 $K^TH^2Kw=\xi K^THKw$ 的根。取 $A=K^TH^2K$、$B=K^THK$ 应用（20）：

$$\Phi_u^{(Y)}(t)=\prod_\ell(1-t/\xi_\ell)=\frac{\det(A-tB)}{\det(A)}=\frac{\det\big(K^T(H^2-tH)K\big)}{\det\big(K^TH^2K\big)}.$$

对（23）：取 $A=K^T\omega(H)HK$、$B=K^T\omega(H)K$，由 $\omega(H)\succ0$ 与 $H\succ0$，二者均为对称正定矩阵，θ_ℓ^{ω} 是 $Av=\theta Bv$ 的根，应用（20）即得。 ∎

引理 4.10 (Cauchy–Binet 加权展开) 设 $u\in\mathbb{R}^n$，$M=\operatorname{diag}(m_1,\cdots,m_n)$ 为对角矩阵（每个 $m_i\in\mathbb{R}$），$K=K_p(H,u)$。令 $V\in\mathbb{R}^{n\times p}$ 满足 $V_{ij}=\lambda_i^{j-1}$，记 $\mathcal{I}_p:=\{I\subset\{1,\cdots,n\}:|I|=p\}$，$V_I$ 为 V 的行下标在 I 中的 $p\times p$ 子矩阵。则

$$\det(K^TMK)=\sum_{I\in\mathcal{I}_p}\det(V_I)^2\prod_{i\in I}m_i\,u_i^2. \tag{24}$$

特别地，对 S、Y、ω 三种乘子的分子分母，（24）分别取

$$\begin{aligned}
&S\text{ 型：}\quad m_i=\lambda_i\ (\text{分母}),\quad m_i=\lambda_i-\lambda_r\ (\text{分子},\ t=\lambda_r);\\
&Y\text{ 型：}\quad m_i=\lambda_i^2\ (\text{分母}),\quad m_i=\lambda_i(\lambda_i-\lambda_r);\\
&\omega\text{ 型：}\quad m_i=\omega(\lambda_i)\lambda_i,\quad m_i=\omega(\lambda_i)(\lambda_i-\lambda_r).
\end{aligned}$$

证明 在标准形 $H=\operatorname{diag}(\lambda_i)$ 下，$K_p(H,u)$ 的第 i 行第 j 列为 $u_i\lambda_i^{j-1}$，故 $K=D_uV$，其中 $D_u=\operatorname{diag}(u_1,\cdots,u_n)$。于是 $K^TMK=V^TD_uMD_uV=V^T\operatorname{diag}(m_iu_i^2)V$。对 $V^T\operatorname{diag}(m_iu_i^2)V$ 应用 Cauchy–Binet 公式即得（24）。 ∎

把引理 4.9 与 4.10 合并，即得到块乘子在每个活跃 λ_r 处的统一加权平均表示：

$$\Psi_u^{(\bullet)}(\lambda_r)=\frac{\sum_{I\in\mathcal{I}_p}\beta_I^{(\bullet)}(u)\,\phi_I^{(r)}}{\sum_{I\in\mathcal{I}_p}\beta_I^{(\bullet)}(u)},\qquad \phi_I^{(r)}:=\prod_{i\in I}\Big(1-\frac{\lambda_r}{\lambda_i}\Big), \tag{25}$$

其中权重 $\beta_I^{(\bullet)}(u)$ 在三种情形分别为

$$\beta_I^{(S)}(u)=\det(V_I)^2\prod_{i\in I}\lambda_iu_i^2,\quad \beta_I^{(Y)}(u)=\det(V_I)^2\prod_{i\in I}\lambda_i^2u_i^2,\quad \beta_I^{(\omega)}(u)=\det(V_I)^2\prod_{i\in I}\omega(\lambda_i)\lambda_iu_i^2. \tag{26}$$

这一统一表达式说明，三种步长在“块乘子的非负加权平均”的意义下完全同构，差别仅在权重的常数因子 λ_i、λ_i^2、$\omega(\lambda_i)\lambda_i$ 的不同。

引理 4.11 (乘子权重引理) 设 $r\in\{2,\cdots,n\}$，$\mathcal{I}_p$ 为 $\{1,\cdots,n\}$ 的所有 p 元指标集，$\{\beta_I^{(k)}\}_{k\geqslant0,\,I\in\mathcal{I}_p}\subset[0,\infty)$，$\{\phi_I^{(r)}\}_{I\in\mathcal{I}_p}\subset\mathbb{R}$。考虑加权平均

$$\chi_k:=\frac{\sum_{I\in\mathcal{I}_p}\beta_I^{(k)}\phi_I^{(r)}}{\sum_{I\in\mathcal{I}_p}\beta_I^{(k)}}.$$

若 $\mathcal{I}_p=A_r\cup B_r\cup C_r$ 为不交并，且：

(i) 若 $I \in A_r$，则 $\phi_I^{(r)} = 0$；

(ii) 若 $I \in B_r$，则 $0 \leqslant \phi_I^{(r)} \leqslant \rho_r < 1$；

(iii) 对每个 $I \in C_r$，存在 $i(I) < r$、$J(I) \in A_r$ 与与 k 无关的常数 $c_I > 0$，使

$$\frac{\beta_I^{(k)}}{\beta_{J(I)}^{(k)}} = c_I \, \frac{(\xi_k^{(i(I))})^2}{(\xi_k^{(r)})^2},$$

其中 $\{\xi_k\} \subset \mathbb{R}^n$ 是某给定向量列；

若 $\xi_k^{(j)} \to 0$（$j < r$），且对某 $\varepsilon > 0$，集合 $S_\varepsilon := \{k \geqslant 0 : |\xi_k^{(r)}| \geqslant \varepsilon\}$ 是无限集，则

$$\limsup_{k \in S_\varepsilon} |\chi_k| \leqslant \rho_r < 1.$$

证明 按三类分组：

$$|\chi_k| \leqslant \frac{\sum_{I \in B_r} \beta_I^{(k)} \phi_I^{(r)}}{\sum_{I \in \mathcal{I}_p} \beta_I^{(k)}} + \frac{\sum_{I \in C_r} \beta_I^{(k)} |\phi_I^{(r)}|}{\sum_{I \in \mathcal{I}_p} \beta_I^{(k)}}.$$

第一项由 (ii) 不超过 ρ_r。对第二项，记 $M_r := \max_{I \in C_r} |\phi_I^{(r)}| < \infty$，则

$$|\chi_k| \leqslant \rho_r + M_r \, \frac{\sum_{I \in C_r} \beta_I^{(k)}}{\sum_{I \in \mathcal{I}_p} \beta_I^{(k)}}.$$

对每个 $I \in C_r$，由 $J(I) \in A_r$ 与 $|\xi_k^{(r)}| \geqslant \varepsilon$（$k \in S_\varepsilon$），

$$\frac{\beta_I^{(k)}}{\sum_L \beta_L^{(k)}} \leqslant \frac{\beta_I^{(k)}}{\beta_{J(I)}^{(k)}} = c_I \, \frac{(\xi_k^{(i(I))})^2}{(\xi_k^{(r)})^2} \leqslant \frac{c_I}{\varepsilon^2} \, (\xi_k^{(i(I))})^2.$$

由 $i(I) < r$ 与归纳假设，$\xi_k^{(i(I))} \to 0$；对有限个 $I \in C_r$ 求和即得结论。 ∎

注 (关于 Fletcher 附录中极限 Ritz 值论证的漏洞) Fletcher[11] 附录中证明 LMSD 收敛性的一个关键困难是：若试图从极限向量的谱支撑直接推出每个单独的 Ritz 值的渐近位置，则需要在 Krylov 维数下降的边界上单独处理这一步，并不能按原文表述直接成立。一个简单反例：取 $H = \mathrm{diag}(1, 2)$、$p = 2$，令

$$q_n = \frac{(1/n, 1)^T}{\sqrt{1 + 1/n^2}} \to e_2.$$

对每个有限 n，$\mathrm{span}\{q_n, H q_n\} = \mathbb{R}^2$，故由该二维 Krylov 子空间得到的两个 Ritz 值始终为 $\{1, 2\}$；它们不会随 $q_n \to e_2$ 一起落入子区间 $(\frac{4}{3}, 2)$ 之内。这说明在极限点出现 Krylov 维数下降时，单个 Ritz 值的渐近行为不是一个可靠的分析对象。本文下面的证明改为直接控制完整块作用在第 r 个特征分量上的乘子多项式

$$\Psi_u^{(\bullet)}(\lambda_r) = \prod_{\ell=0}^{p-1} \Big(1 - \frac{\lambda_r}{\theta_{k,\ell}}\Big),$$

通过引理 4.9–4.11 把它表为非负权重的加权平均，并直接控制其极限，从而绕开这一障碍。

二、S 型方法的逐点收敛

定理 4.12 (S 型方法的逐点收敛) 设 q 满足（18）并已规约至假设 1.1 的标准形。考虑严格凸二次问题上的精确秩压缩 S 型 LMSD（定义 3.1）。则从任意初始块状态出发，相应轨道要么有限步终止于原点，要么 $g_{k,0}\to 0$，进而 $E_k\to 0$。

证明 若某个块起点 $g_{k,0}$ 满足 $s(g_{k,0})\leqslant p$，则由命题 4.7 算法有限步终止。下面只需考虑非有限终止情形。此时每个非零块起点都满足 $s(g_{k,0})>p$。若某个活跃分量在有限块后变为零，则由二次递推它此后永久保持为零，删去这些分量等价于把问题压缩到更小的活跃不变子空间。这样的删减至多发生有限次。若删减后剩余活跃特征值个数不超过 p，仍由命题 4.7 有限终止。因此在非有限终止尾轨道上，可重新编号并不失一般性地假设所有剩余活跃分量始终非零且其个数 $n>p$。

记 $u_k:=g_{k,0}$。由命题 3.3，$\operatorname{span}(S_k)=\mathcal{K}_p(H,u_k)$，故 S 型步长所用的 Ritz 值 $\theta_{k,\ell}$ 是 H 在 $\mathcal{K}_p(H,u_k)$ 上的标准 Ritz 值，从而

$$\lambda_1\leqslant\theta_{k,\ell}\leqslant\lambda_n,\qquad \ell=0,\cdots,p-1. \tag{27}$$

设 $\Psi_k(t):=\prod_{\ell=0}^{p-1}(1-t/\theta_{k,\ell})$。块 k 生成的 Ritz 值在块 $k+1$ 中依次使用，因此

$$u_{k+2}=\Psi_k(H)u_{k+1},\qquad u_{k+2}^{(i)}=\Psi_k(\lambda_i)\,u_{k+1}^{(i)}. \tag{28}$$

第一分量。由（27），对每个 ℓ 有 $0\leqslant 1-\lambda_1/\theta_{k,\ell}\leqslant 1-\lambda_1/\lambda_n<1$，故

$$|u_{k+2}^{(1)}|\leqslant\left(1-\frac{\lambda_1}{\lambda_n}\right)^p|u_{k+1}^{(1)}|,$$

从而 $u_k^{(1)}\to 0$。

归纳。若所有分量都趋零则结论成立。否则设 $r>1$ 是满足 $u_k^{(j)}\to 0\,(j<r)$ 但 $u_k^{(r)}\not\to 0$ 的最小指标。则存在 $\bar\varepsilon>0$，使对每个 $\varepsilon\in(0,\bar\varepsilon]$，集合 $S_\varepsilon:=\{k:|u_k^{(r)}|\geqslant\varepsilon\}$ 都是无限集。固定这样的 ε，简记 $S:=S_\varepsilon$。

由（27）及乘积形式 $\Psi_k(\theta)=\prod(1-\theta/\theta_{k,\ell})$，存在仅依赖谱比值的常数 $M_r:=\max\{1,(\lambda_n/\lambda_1-1)^p\}\geqslant 1$，使

$$|u_{k+2}^{(r)}|\leqslant M_r|u_{k+1}^{(r)}|,\qquad k\geqslant 0. \tag{29}$$

将引理 4.9 应用于 S 型方法，再由引理 4.10，得到（25）的 S 型形式：

$$\Psi_k(\lambda_r)=\frac{\sum_{I\in\mathcal{J}_p}\beta_I^{(S)}(u_k)\phi_I^{(r)}}{\sum_{I\in\mathcal{J}_p}\beta_I^{(S)}(u_k)}, \tag{30}$$

其中 $\beta_I^{(S)}(u_k)=\det(V_I)^2\prod_{i\in I}\lambda_i u_{k,i}^2\geqslant 0$，$\phi_I^{(r)}=\prod_{i\in I}(1-\lambda_r/\lambda_i)$。

把 $\mathcal{J}_p$ 分成

$$A_r:=\{I:r\in I\},\quad B_r:=\{I\subset\{r{+}1,\cdots,n\}\},\quad C_r:=\{I:r\notin I,\ I\cap\{1,\cdots,r{-}1\}\neq\varnothing\}.$$

对 $I \in A_r$ 有 $\phi_I^{(r)} = 0$；对 $I \in B_r$ 有

$$0 \leqslant \phi_I^{(r)} \leqslant \left(1 - \frac{\lambda_r}{\lambda_n}\right)^p =: \rho_r < 1.$$

对 $I \in C_r$，任选 $i(I) \in I \cap \{1, \cdots, r-1\}$，令 $J(I) := (I \setminus \{i(I)\}) \cup \{r\} \in A_r$；由活跃特征值互异，相关 Vandermonde 子式非零，于是存在与 k 无关的常数 $c_I > 0$ 使

$$\frac{\beta_I^{(S)}(u_k)}{\beta_{J(I)}^{(S)}(u_k)} = \frac{\det(V_I)^2}{\det(V_{J(I)})^2} \cdot \frac{\lambda_{i(I)}}{\lambda_r} \cdot \frac{(u_k^{(i(I))})^2}{(u_k^{(r)})^2} = c_I \, \frac{(u_k^{(i(I))})^2}{(u_k^{(r)})^2}.$$

故 $\Psi_k(\lambda_r)$ 满足引理 4.11 的全部假设，从而

$$\limsup_{k \in S} |\Psi_k(\lambda_r)| \leqslant \rho_r < 1. \tag{31}$$

任选 $c_r(\varepsilon) \in (\rho_r, 1)$，则存在 k_ε，使对一切 $k \in S$ 且 $k \geqslant k_\varepsilon$，

$$|u_{k+2}^{(r)}| \leqslant c_r(\varepsilon) |u_{k+1}^{(r)}|.$$

令 $a_k := |u_k^{(r)}|$。结合（29）与上式，应用引理 4.8 得 $a_k \to 0$，与 r 的最小性矛盾。故 $u_k = g_{k,0} \to 0$。

最后，由（27）与块内递推 $g_{k,j+1} = (I - \alpha_{k,j} H) g_{k,j}$，每步乘子一致有界，故对每个 $r = 0, \cdots, p$ 都有 $\|g_{k,r}\| \leqslant C \|g_{k,0}\|$，从而 $g_{k,r} \to 0$。再由 $g = Hx$ 与 $H \succ 0$ 得 $E_k \to 0$。 ∎

三、Y 型方法的逐点收敛

定理 4.13 (Y 型方法的逐点收敛) 设 q 满足（18）并已规约至标准形。考虑严格凸二次问题上的精确秩压缩 Y 型方法（即 harmonic-LMSD，定义 3.2）。则从任意初始块状态出发，相应轨道要么有限步终止于原点，要么 $g_{k,0} \to 0$，进而 $E_k \to 0$。

证明 由命题 4.7，非正则分支有限终止。在非有限终止尾轨道上，与定理 4.12 证明开头相同，可以删去永久为零的分量并重新编号，使所有剩余活跃分量始终非零且活跃特征值个数 $n > p$。

记 $u_k := g_{k,0}$。由定义 3.2，Y 型直接采用 H^{-1} 在 $H\,\mathcal{K}_p(H, u_k)$ 上的 Ritz 值 $\tau_{k,\ell}$ 为下一块步长。由 Rayleigh 商的性质，$\tau_{k,\ell} \in [1/\lambda_n, 1/\lambda_1]$。设 $\Phi_k(t) := \prod_{\ell=0}^{p-1} (1 - \tau_{k,\ell} t)$，则

$$u_{k+2} = \Phi_k(H) u_{k+1}, \qquad u_{k+2}^{(i)} = \Phi_k(\lambda_i) u_{k+1}^{(i)}. \tag{32}$$

特别地，

$$|\Phi_k(\lambda_1)| \leqslant \left(1 - \frac{\lambda_1}{\lambda_n}\right)^p < 1,$$

故 $u_k^{(1)} \to 0$。

设 $r>1$ 为满足 $u_k^{(j)}\to 0$（$j<r$）但 $u_k^{(r)}\not\to 0$ 的最小指标，仍记 $S=S_\varepsilon$。由 $\tau_{k,\ell}\in[1/\lambda_n,1/\lambda_1]$ 可得粗上界：存在 $M_r\geqslant 1$ 使

$$|u_{k+2}^{(r)}|\leqslant M_r|u_{k+1}^{(r)}|,\qquad k\geqslant 0. \tag{33}$$

由引理 4.9 应用于 Y 型与引理 4.10，得到

$$\boldsymbol{\Phi}_k(\lambda_r)=\frac{\sum_{I\in\mathcal{J}_p}\beta_I^{(Y)}(u_k)\boldsymbol{\phi}_I^{(r)}}{\sum_{I\in\mathcal{J}_p}\beta_I^{(Y)}(u_k)}, \tag{34}$$

其中 $\beta_I^{(Y)}(u_k)=\det(V_I)^2\prod_{i\in I}\lambda_i^2u_{k,i}^2$，$\boldsymbol{\phi}_I^{(r)}$ 与定理 4.12 中相同。把 $\mathcal{J}_p$ 同样分成 A_r、B_r、C_r 三类。对 $I\in C_r$，仍取 $i(I)<r$、$J(I)\in A_r$，存在与 k 无关的常数 $\tilde{c}_I>0$ 使

$$\frac{\beta_I^{(Y)}(u_k)}{\beta_{J(I)}^{(Y)}(u_k)}=\frac{\det(V_I)^2}{\det(V_{J(I)})^2}\cdot\frac{\lambda_{i(I)}^2}{\lambda_r^2}\cdot\frac{(u_k^{(i(I))})^2}{(u_k^{(r)})^2}=\tilde{c}_I\,\frac{(u_k^{(i(I))})^2}{(u_k^{(r)})^2}.$$

Y 型权重相比 S 型只多出一个常数因子 $\lambda_{i(I)}/\lambda_r$，不改变指标交换处的阶数比较；按引理 4.11 仍得

$$\limsup_{k\in S}|\boldsymbol{\Phi}_k(\lambda_r)|\leqslant\rho_r<1.$$

结合粗界（33），应用引理 4.8 得 $|u_k^{(r)}|\to 0$，与 r 的最小性矛盾。故 $u_k=g_{k,0}\to 0$。块内梯度与误差的传递与 S 型证明完全相同。 ∎

第三节　二次模型上的 R-线性收敛

本节把 §二 的逐点收敛升级为闭锥状态空间上的 R-线性收敛。技术上的关键是把“历史束秩亏”吸收到一个闭锥的内部边界上，从而克服 Curtis–Guo[13] 的“历史束非退化”假设，使得 R-线性收敛成为真正无条件成立的结论。具体做法是把状态从历史矩阵改为连续两个块端点梯度

$$Z_k=(u_k,v_k):=(g_{k,0},g_{k+1,0}),$$

并把秩下降解释为闭锥边界上的低维活跃谱情形——在严格凸二次的精确模型中，这正是有限终止分支。

一、相容块端点梯度对闭锥

定义 4.1 (相容块端点梯度对闭锥)　固定块长 $p\geqslant 1$，并令

$$\mathcal{P}_p:=\left\{q(t)=\prod_{\ell=1}^{p}(1-\alpha_\ell t)\,:\,\alpha_\ell\in\left[\frac{1}{\lambda_n},\frac{1}{\lambda_1}\right]\right\}.$$

定义

$$\Omega_p:=\left\{(u,v)\in\mathbb{R}^n\times\mathbb{R}^n\,:\,\exists q\in\mathcal{P}_p,\ v=q(H)u\right\}.$$

称 Ω_p 为 p 步 Ritz-sweep 的相容块端点梯度对闭锥。这个集合比真实 LMSD 可达状态集合更宽，但它具有闭性，并且包含所有由合法谱步长块产生的块端点对。

引理 4.14 Ω_p 是 $\mathbb{R}^n \times \mathbb{R}^n$ 中的闭锥。并且若 $(u,v) \in \Omega_p$，则

$$u_i = 0 \Longrightarrow v_i = 0, \qquad i = 1, \cdots, n.$$

证明 锥性直接来自 $v = q(H)u \Rightarrow tv = q(H)(tu)$。下面证明闭性。设 $(u_j, v_j) \in \Omega_p$ 且 $(u_j, v_j) \to (u, v)$。对每个 j，取

$$q_j(t) = \prod_{\ell=1}^{p}(1 - \alpha_\ell^{(j)} t), \qquad \alpha_\ell^{(j)} \in \left[\frac{1}{\lambda_n}, \frac{1}{\lambda_1}\right],$$

使得 $v_j = q_j(H)u_j$。由区间紧性，取子列后可令每个 $\alpha_\ell^{(j)} \to \alpha_\ell$。于是 $q_j(\lambda_i) \to q(\lambda_i)$，其中

$$q(t) = \prod_{\ell=1}^{p}(1 - \alpha_\ell t) \in \mathcal{P}_p.$$

逐坐标取极限得 $v = q(H)u$，故 $(u,v) \in \Omega_p$。最后，若 $u_i = 0$，则 $v_i = q(\lambda_i)u_i = 0$。 ∎

引理 4.15 若从梯度 u 出发连续执行 p 步梯度迭代

$$w_{\ell+1} = (I - \alpha_\ell H)w_\ell, \qquad \ell = 0, \cdots, p-1,$$

并且每个步长都满足 $\alpha_\ell \in [1/\lambda_n, 1/\lambda_1]$，则 $(u, w_p) \in \Omega_p$。特别地，Cauchy 步长、BB1/BB2 步长以及 LMSD 步长均满足这个区间约束。

证明 直接有

$$w_p = \prod_{\ell=0}^{p-1}(I - \alpha_\ell H)u = q(H)u,$$

其中 $q(t) = \prod_{\ell=0}^{p-1}(1 - \alpha_\ell t) \in \mathcal{P}_p$，故 $(u, w_p) \in \Omega_p$。

对 Cauchy 步长，$\alpha_C = (g^T g)/(g^T H g)$ 是 Rayleigh 商倒数，故落在 $[1/\lambda_n, 1/\lambda_1]$。对 BB1，$\alpha_{BB1} = s^T s/(s^T H s)$；对 BB2，$\alpha_{BB2} = s^T H s/(s^T H^2 s)$，二者同样由 H 或 H^{-1} 的 Rayleigh 商给出，均在上述区间内。LMSD 步长是 Ritz 值的倒数，而 Ritz 值属于 $[\lambda_1, \lambda_n]$，故其倒数也在该区间内。 ∎

对 $u \in \mathbb{R}^n$，记

$$S(u) := \{i : u_i \neq 0\}, \qquad s(u) := |S(u)|,$$

并令

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

当 $s(u) > p$ 时，$r(u) = p$，$K_p(H,u)$ 满列秩，定义 S 型 Ritz-sweep 乘子

$$\Psi_u(t) := \frac{\det\big(K_p(H,u)^T (H - tI) K_p(H,u)\big)}{\det\big(K_p(H,u)^T H K_p(H,u)\big)}. \tag{35}$$

它等价于

$$\Psi_u(t) = \prod_{\ell=0}^{p-1} \left(1 - \frac{t}{\theta_\ell(u)}\right),$$

其中 $\theta_\ell(u)$ 是 H 在 $\mathcal{K}_p(H,u)$ 上的 p 个 Ritz 值。

当 $s(u) \leqslant p$ 时，秩压缩算法解释为在实际维数 $r(u) = s(u)$ 的 Krylov 子空间上取 Ritz 值。由于该子空间就是 $\operatorname{span}\{e_i : i \in S(u)\}$，这些 Ritz 值恰为 $\{\lambda_i : i \in S(u)\}$。因此短 sweep 的乘子在 $S(u)$ 上全部为零，又由引理 4.14，相容状态满足 $\operatorname{supp}(v) \subseteq S(u)$，故下一块端点为零。为书写简洁，下面在 $s(u) \leqslant p$ 时直接记

$$\Psi_u(H)v := 0.$$

若 $s(u) = p$，算法也给出全部活跃特征值，因而与有限终止分支一致。从而可以给出下面的分段定义。

定义闭锥上的块端点映射

$$T_p^S(u,v) := \begin{cases} \big(v, \Psi_u(H)v\big), & s(u) > p, \\ (v, 0), & s(u) \leqslant p. \end{cases} \tag{36}$$

引理 4.16 $T_p^S : \Omega_p \to \Omega_p$ 是连续的一次正齐次映射。

证明 先证自映射性。若 $s(u) > p$，则 Ritz 值满足 $\theta_\ell(u) \in [\lambda_1, \lambda_n]$，从而 $\Psi_u \in \mathcal{P}_p$。若 $(u,v) \in \Omega_p$，则 $(v, \Psi_u(H)v) \in \Omega_p$。若 $s(u) \leqslant p$，由引理 4.14 有 $\operatorname{supp}(v) \subseteq S(u)$，故 v 的活跃谱个数也不超过 p。取一个 $q \in \mathcal{P}_p$，使它的根包含 $\{\lambda_i : i \in \operatorname{supp}(v)\}$，即可得到 $q(H)v = 0$，于是 $(v,0) \in \Omega_p$。

正齐次性来自 $K_p(H, tu) = tK_p(H,u)$：当 $t > 0$ 时，Ritz 值与 Ψ_u 不变，故 $T_p^S(tu, tv) = tT_p^S(u,v)$；$t = 0$ 情形显然成立。

下面证明连续性。设 $(u_j, v_j) \to (u,v)$ 且 $(u_j, v_j) \in \Omega_p$。由于任意序列都可分解为 $s(u_j) > p$ 与 $s(u_j) \leqslant p$ 的子列，只需分别证明这些子列的第二分量都收敛到同一极限。

若 $s(u) > p$，则充分大的 j 也有 $s(u_j) > p$。对 $i \in S(u)$，式（35）的分母在极限处非零，故 $\Psi_{u_j}(\lambda_i) \to \Psi_u(\lambda_i)$。若 $i \notin S(u)$，则由引理 4.14 有 $v_i = 0$，从而 $v_{j,i} \to 0$。又由于所有 Ritz 值都在 $[\lambda_1, \lambda_n]$ 中，$\Psi_{u_j}(\lambda_i)$ 在有限谱点上一致有界，故 $\Psi_{u_j}(\lambda_i) v_{j,i} \to 0$。于是 $\Psi_{u_j}(H)v_j \to \Psi_u(H)v$。

下面考虑 $s(u) \leqslant p$。此时 $T_p^S(u,v)$ 的第二分量为零，需证子列的第二分量也趋于零。若沿某个子列 $s(u_j) \leqslant p$，该子列的第二分量本来就是 0。因此只需考察 $s(u_j) > p$ 的子列。对 $i \notin S(u)$，仍由 $v_{j,i} \to 0$ 与乘子一致有界得到第 i 个输出分量趋于零。对 $r \in S(u)$，证明 $\Psi_{u_j}(\lambda_r) \to 0$。

记 $V \in \mathbb{R}^{n\times p}$，$V_{ab} = \lambda_a^{b-1}$，并令 $\mathcal{I}_p$ 为所有 p 元指标集。由 Cauchy–Binet 展开

$$\Psi_{u_j}(\lambda_r) = \frac{\sum_{I\in\mathcal{I}_p} \beta_I(u_j)\phi_I^{(r)}}{\sum_{I\in\mathcal{I}_p} \beta_I(u_j)}, \tag{37}$$

其中

$$\beta_I(u_j) = \det(V_I)^2 \prod_{i\in I} \lambda_i u_{j,i}^2 \geqslant 0, \qquad \phi_I^{(r)} = \prod_{i\in I}\left(1 - \frac{\lambda_r}{\lambda_i}\right).$$

若 $r \in I$，则 $\phi_I^{(r)} = 0$。若 $r \notin I$，由于 $|S(u)| \leqslant p$ 且 I 是 p 元集合，计数得

$$|I \cap S(u)^c| \geqslant p - (|S(u)| - 1) \geqslant 1.$$

于是可取 $h \in I \cap S(u)^c$。令 $J = (I \setminus \{h\}) \cup \{r\}$。当 $\beta_I(u_j) > 0$ 时，充分大的 j 下 $\beta_J(u_j) > 0$，且

$$\frac{\beta_I(u_j)}{\beta_J(u_j)} = C_{I,J}\frac{u_{j,h}^2}{u_{j,r}^2} \to 0,$$

其中 $C_{I,J} > 0$ 与 j 无关。这里用到了 $u_{j,h} \to 0$ 而 $u_{j,r} \to u_r \neq 0$。于是

$$\frac{\beta_I(u_j)}{\sum_L \beta_L(u_j)} \leqslant \frac{\beta_I(u_j)}{\beta_J(u_j)} \to 0.$$

对有限个 I 求和，由（37）得 $\Psi_{u_j}(\lambda_r) \to 0$，连续性得证。 ∎

引理 4.17 对任意 $Z_0 = (u_0, u_1) \in \Omega_p$，由

$$Z_{k+1} = T_p^S(Z_k), \qquad Z_k = (u_k, u_{k+1}),$$

生成的序列满足 $Z_k \to 0$。

证明 Ω_p 比真实 LMSD 可达状态集合更宽，因此需要证明：闭锥块映射意义下的轨道，仍然满足逐点收敛。若某个 u_k 满足 $s(u_k) \leqslant p$，则由定义（36）有 $u_{k+2} = 0$，随后有限步终止。以下假设 $s(u_k) > p$ 对所有 k 成立。于是

$$u_{k+2} = \Psi_{u_k}(H)u_{k+1}, \qquad u_{k+2}^{(i)} = \Psi_{u_k}(\lambda_i)u_{k+1}^{(i)}.$$

把这个递推与定理 4.12 证明中得到的递推（28）进行对照，可看出二者结构完全相同：闭锥状态空间下的“块端点对” (u_k, u_{k+1}) 与真实 LMSD 轨道 $(g_{k,0}, g_{k+1,0})$ 在精确二次模型上服从同一条由块乘子 Ψ_{u_k} 决定的递推。因此定理 4.12 的证明可以逐字搬到这里：先用块乘子在第一谱分量上的一致收缩 $\leqslant (1-\lambda_1/\lambda_n)^p$ 得 $u_k^{(1)} \to 0$；再对 $r = 2, \cdots, n$ 归纳，借助引理 4.9 和 4.10 把 $\Psi_{u_k}(\lambda_r)$ 写成（25）的加权平均，应用乘子权重引理 4.11 得到 $\limsup_{k\in S} |\Psi_{u_k}(\lambda_r)| \leqslant \rho_r < 1$，再由 Fletcher 两步引理 4.8 推出 $u_k^{(r)} \to 0$。归纳得 $u_k \to 0$，故 $Z_k = (u_k, u_{k+1}) \to 0$。 ∎

二、R-线性收敛定理

定理 4.18 (精确秩压缩 LMSD 的无条件 R-线性收敛) 设 q 由（18）给出，并已按假设 1.1 化为标准形。考虑严格凸二次问题上的精确秩压缩 S 型 LMSD：若算法经过有限个预热步后进入相容闭锥 Ω_p，则算法要么有限步终止，要么块端点梯度对 $Z_k = (g_{k,0}, g_{k+1,0})$ R-线性收敛到零。进一步，所有块内梯度与迭代误差也 R-线性收敛到零。

证明 在 $K_p(H,u)$ 满列秩时按定义 3.1 使用标准 Ritz 值；当 $K_p(H,u)$ 的实际维数小于 p 时，按实际 Krylov 子空间提取短 sweep。若实际活跃谱个数等于 p，标准 p 步 sweep 本身给出全部活跃特征值并有限终止；若实际活跃谱个数小于 p，短 sweep 有限终止。

由引理 4.14 与 4.16，T_p^S 是闭锥 Ω_p 上的连续正齐次自映射。由引理 4.17，对任意 $Z \in \Omega_p$ 都有 $(T_p^S)^k(Z) \to 0$。由此应用第二章的引理 2.2，存在常数 $C > 0$ 与 $\rho \in (0,1)$ 使

$$\|(T_p^S)^k(Z)\| \leqslant C\rho^k \|Z\|, \qquad Z \in \Omega_p, \quad k \geqslant 0,$$

即得块端点梯度对的 R-线性收敛。

对真实 LMSD 轨道而言，只要预热阶段采用 Cauchy、BB1/BB2 或 Ritz-sweep 等区间内步长，引理 4.15 保证经过一个完整合法块后，相邻两个块端点梯度对属于 Ω_p。

块内每个步长都属于 $[1/\lambda_n, 1/\lambda_1]$，故存在常数

$$B := \max_{\alpha \in [1/\lambda_n, 1/\lambda_1]} \|I - \alpha H\| < \infty$$

使对 $r = 0, \cdots, p$ 有 $\|g_{k,r}\| \leqslant B^r \|g_{k,0}\|$，于是块内梯度也 R-线性收敛。最后由 $g = Hx$、$H \succ 0$ 知梯度范数与误差范数等价，因此迭代误差同样 R-线性收敛到零。 ∎

定理 4.19 (精确秩压缩 harmonic-LMSD 的无条件 R-线性收敛) 设 q 由（18）给出。考虑严格凸二次问题上的精确秩压缩 Y 型方法（即 harmonic-LMSD）。则与定理 4.18 相同，该方法要么有限步终止，要么块端点梯度对、块内梯度和迭代误差均 R-线性收敛到零。

证明 当 $s(u) > p$ 时，Y 型块乘子 Φ_u 由引理 4.9 的（22）给出；当 $s(u) \leqslant p$ 时，由命题 4.7，短 sweep 立即终止。定义 $\Phi_u(H)v = 0$，并令

$$T_p^Y(u,v) := \begin{cases} (v, \Phi_u(H)v), & s(u) > p, \\ (v, 0), & s(u) \leqslant p. \end{cases}$$

由 $\tau_\ell(u) \in [1/\lambda_n, 1/\lambda_1]$，$T_p^Y$ 仍是 Ω_p 上的一次正齐次自映射。

连续性证明完全平行于引理 4.16。Cauchy–Binet 权重从 $\beta_I^{(S)}(u) = \det(V_I)^2 \prod_{i\in I} \lambda_i u_i^2$ 变为 $\beta_I^{(Y)}(u) = \det(V_I)^2 \prod_{i\in I} \lambda_i^2 u_i^2$（参见（26）），新增的 λ_i 因子是与迭代无关的正常数，不改变退化边界处指标交换的阶数比较。逐点收敛由定理 4.13 给出，再由第二章的引理 2.2 即得块端点对的 R-线性收敛。块内梯度与迭代误差的传递与 S 型证明相同。 ∎

注 定理 4.18–4.19 说明了本文结论与 Curtis–Guo[13] 的关系。Curtis–Guo 的证明固定使用 m 列历史矩阵，并在其 Assumption 3.4(ii) 中要求 G_k 列满秩以及 $\|R_k^{-1}\| \leqslant \rho \|g_{k,1}\|^{-1}$。本文 §二–三 不再要求这一历史束非退化界，但相应地证明对象也限定为严格凸二次下的秩压缩理论模型，该模型在严格凸二次问题中与原始 LMSD 方法是完全一致的。可见本定理结论确实是对 Curtis–Guo[13] 中结论的推广。

本节的 R-线性收敛对块长 p 没有结构性要求，$p = 1$ 时同样适用。事实上，第二章 §五 已经把 BB1/BB2 作为闭锥原理（引理 2.2）的一个简单实例处理过：以连续两步梯度对 (g_k, g_{k+1}) 为状态，把 BB 步长 $\beta(u) \in [1/\lambda_n, 1/\lambda_1]$ 视作 u 上的 Rayleigh 商或 Harmonic Rayleigh 商，则相应轨道闭锥 Ω_{BB} 是闭的，递推映射 $\mathcal{T}_{\mathrm{BB}}$ 连续正齐次（命题 2.7），再结合 Raydan[3] 的逐点收敛即得 R-线性结论（推论 2.8）。

把命题 2.7–推论 2.8 与定理 4.18 并置可以看出 $p = 1$ 与 $p \geqslant 2$ 在闭锥几何上的两点差异：

(1) Ω_{BB} 是 $\alpha \in [1/\lambda_n, 1/\lambda_1]$ 上 $u \mapsto (I - \alpha H)u$ 这族线性映射的图像之并，几何结构相对简单，而 Ω_p（$p \geqslant 2$）含降秩子层作为内部边界（引理 4.14）。

(2) Ω_{BB} 没有降秩分支（退化为原点），而 Ω_p 上 T_p^S 在降秩边界附近的连续性正是 LMSD 证明的核心技术内容（引理 4.16）。

由此可见，闭锥框架同时覆盖 BB 与 LMSD，二者可作为同一速率提升机制（逐点收敛 $\to R$-线性收敛）的两个实例，体现了闭锥框架的有效性和普适性。

最后，作为对上述收敛性理论的补充，类似[4]，我们补充一个简单的良态情形：若所有可能出现的步长都被自然地限制在 $[1/\lambda_{\max}, 1/\lambda_{\min}]$ 内，且问题条件数足够好，则收敛会进一步强化为 Q-线性。

定理 4.20 考虑严格凸二次模型

$$q(x) = \frac{1}{2} x^T H x, \qquad H \succ 0.$$

记 H 的最大和最小特征值分别为 $\lambda_{\max}$ 与 $\lambda_{\min}$。若满足 $\lambda_{\max} < 2\lambda_{\min}$，则对于任何满足

$$\alpha_k \in [1/\lambda_{\max},\ 1/\lambda_{\min}]$$

的步长选取策略，梯度法迭代序列 $\{x_k\}$ 都 Q-线性收敛到原点。

证明 在严格凸二次情形下，

$$g_k = Hx_k, \qquad x_{k+1} = x_k - \alpha_k g_k = (I - \alpha_k H)x_k.$$

取 2-范数可得

$$\|x_{k+1}\|_2 \leqslant \|I - \alpha_k H\|_2 \, \|x_k\|_2 .$$

由于 H 为对称正定矩阵，矩阵 $I - \alpha_k H$ 的 2-范数等于其谱半径，即

$$\|I - \alpha_k H\|_2 = \max_{\lambda_i \in \sigma(H)} |1 - \alpha_k \lambda_i| .$$

当 $\alpha_k \in [1/\lambda_{\max}, 1/\lambda_{\min}]$ 时，上式的最大值由端点控制：

$$|1 - \alpha_k \lambda_i| \leqslant \max\left\{ \left|1 - \frac{\lambda_{\max}}{\lambda_{\min}}\right|, \left|1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right| \right\} = \frac{\lambda_{\max}}{\lambda_{\min}} - 1.$$

由假设 $\lambda_{\max} < 2\lambda_{\min}$，令

$$\bar{c} := \frac{\lambda_{\max}}{\lambda_{\min}} - 1,$$

则 $0 < \bar{c} < 1$，并且对所有 k 都有

$$\|x_{k+1}\|_2 \leqslant \bar{c} \, \|x_k\|_2 .$$

这正是 Q-线性收敛。 ∎

根据 pD-BB 步长的代数性质，其步长 α_k 由历史梯度的 Ritz 值生成，并且恒有

$$\alpha_k \in [1/\lambda_{\max},\ 1/\lambda_{\min}].$$

因此，定理 4.20 表明：在良态假设下，pD-BB 方法的收敛会进一步强化为 Q-线性。

第四节 ω 加权 Ritz-sweep 方法族

§二 中的工具引理 4.9–4.11 把块乘子的代数结构归纳为两条：(a) 在任一活跃特征值 λ_r 处，块乘子可以写成两个对称正定矩阵的行列式之比；(b) 由 Cauchy–Binet 展开，这一比值等于 λ_i 上的非负权重平均，并且在指标交换下，权重之比是与迭代无关的正常数。事实上，引理 4.9 与 4.10 中已经把 ω 加权情形作为一般情形写出，对应的权重形如（26）的第三式。因此把广义特征值问题的权重 I 换成任何与 H 交换的正算子 $\omega(H)$，整套证明逐字照搬即可。本节正式把这一观察抽象为一族单参数步长规则，其中 $\omega \equiv 1$ 即 LMSD（S 型），$\omega(t) = t$ 即 harmonic-LMSD（Y 型），它们恰是这一族的两个最简特例。

定义 4.2 (ω 加权 Ritz-sweep)　设 $\omega:\{\lambda_1,\cdots,\lambda_n\}\to(0,\infty)$ 为活跃谱上的任意正函数，并令 $d_k:=\dim\mathcal{K}_p(H,g_{k,0})$。$\omega$-加权 Ritz 值是对称广义特征值问题

$$\big(K_k^T\,\omega(H)H\,K_k\big)v\;=\;\theta\,\big(K_k^T\,\omega(H)\,K_k\big)v \tag{38}$$

的 d_k 个特征值 $\theta^{\omega}_{k,0},\cdots,\theta^{\omega}_{k,d_k-1}$。$\omega$-加权 Ritz-sweep 梯度法在下一个有效 sweep 中使用步长 $\alpha^{\omega}_{k+1,r}=(\theta^{\omega}_{k,r})^{-1}$。在正则非终止分支上 $d_k=p$，使用标准 $p\times p$ 矩阵的特征值。

引理 4.21　ω-加权 Ritz 值满足 $\theta^{\omega}_{k,\ell}\in[\lambda_1,\lambda_n]$。

证明　设 $v\neq0$，令 $u:=K_kv$。在标准形下，

$$\frac{v^T\,K_k^T\,\omega(H)H\,K_k\,v}{v^T\,K_k^T\,\omega(H)\,K_k\,v}=\frac{\sum_i\omega(\lambda_i)\lambda_iu_i^2}{\sum_i\omega(\lambda_i)u_i^2}$$

是 λ_i 的非负权重加权平均，故落在 $[\lambda_1,\lambda_n]$ 内。∎

定理 4.22 (ω 族在精确秩压缩模型上的 R-线性收敛)　设 $\omega:[\lambda_1,\lambda_n]\to(0,\infty)$ 在活跃谱上恒正。在严格凸二次精确秩压缩模型下，ω-加权 Ritz-sweep 方法或者有限步终止，或者块端点梯度对、块内梯度和迭代误差均 R-线性收敛到零。

证明　当 $s(g_{k,0})\leqslant p$ 时，实际 Krylov 子空间为 $\operatorname{span}\{e_i:i\in S(g_{k,0})\}$。因 $\omega(H)$ 是对角正定矩阵，ω-加权 Ritz 值恰为活跃特征值集合 $\{\lambda_i:i\in S(g_{k,0})\}$，下一短 sweep 用其倒数构造消去多项式，立即终止。这与 S 型、Y 型情形完全一致。

在正则非终止分支 $s(g_{k,0})>p$ 上，由引理 4.9 的（23），ω-加权块乘子有

$$\Psi^{\omega}_u(t)=\frac{\det\big(K_p(H,u)^T\omega(H)(H-tI)K_p(H,u)\big)}{\det\big(K_p(H,u)^T\omega(H)H\,K_p(H,u)\big)}=\prod_{\ell=0}^{p-1}\Big(1-\frac{t}{\theta^{\omega}_{k,\ell}}\Big).$$

由引理 4.10 的 ω 型权重（26），对每个活跃 r 有

$$\Psi^{\omega}_u(\lambda_r)=\frac{\sum_{I\in\mathcal{J}_p}\beta_I^{(\omega)}(u)\phi_I^{(r)}}{\sum_{I\in\mathcal{J}_p}\beta_I^{(\omega)}(u)},\qquad\beta_I^{(\omega)}(u)=\det(V_I)^2\prod_{i\in I}\omega(\lambda_i)\lambda_iu_i^2\geqslant0.$$

$\omega(\lambda_i)$ 是与迭代无关的正常数，因此对每个 $I\in C_r$ 的指标交换 $i(I)<r$、$J(I)\in A_r$，比值

$$\frac{\beta_I^{(\omega)}(u_k)}{\beta_{J(I)}^{(\omega)}(u_k)}=c_I^{\omega}\,\frac{(u_k^{(i(I))})^2}{(u_k^{(r)})^2}$$

仍满足乘子权重引理 4.11 假设 (iii)。结合引理 4.21 给出的 $\theta^{\omega}_{k,\ell}\in[\lambda_1,\lambda_n]$（由此 $\Psi^{\omega}_u\in\mathcal{P}_p$ 与 $|\Psi^{\omega}_u(\lambda_1)|<1$）以及上面的退化边界控制，将定理 4.12 的证明逐字照搬即得 $g_{k,0}\to0$ 与 $E_k\to0$。

把状态改为 $(u_k,u_{k+1})\in\Omega_p$，相应的 ω 块端点映射 T_p^{ω} 在 Ω_p 上的连续正齐次性证明与引理 4.16 完全平行（仅是把 $\beta_I^{(S)}$ 换成 $\beta_I^{(\omega)}$）。再应用第二章的引理 2.2，得到块端点梯度对的 R-线性收敛。块内梯度与迭代误差的传递与定理 4.18 证明的最后一段相同。∎

注 (常见权重的实例)　取 $\omega \equiv 1$ 即定理 4.18 中的 S 型 LMSD；取 $\omega(t) = t$ 即定理 4.19 中的 Y 型 harmonic-LMSD。其它实用选择包括：

- $\omega(t) = t^a,\ a \in \mathbb{R}$：单参数族，定理 4.22 对所有这些 a 一致地成立；
- $\omega(t) = e^{-\beta t},\ \beta > 0$：在已知 λ_n 上界时，加大对小特征值的权重；
- $\omega(t) = (t+\mu)^{-2},\ \mu > 0$：常见的正则化权重。

对所有上述 ω，由定理 4.22，相应的 ω-加权 Ritz-sweep 方法在精确秩压缩下都是无条件 R-线性收敛的。

把上述方法以算法形式总结如下。该实现统一覆盖 LMSD（$\omega \equiv 1$）、harmonic-LMSD（$\omega(t) = t$）以及任意正权 ω，并以秩压缩的方式处理历史束秩亏：当 G 的有效秩 $r_k < p$ 时，算法只生成 r_k 个步长并执行短 sweep，对应于精确二次模型上的有限终止分支。

算法 1 严格凸二次问题上的 ω-加权 Ritz-sweep 梯度法

Require: 初始点 x_0，块长 p，正权重 $\omega : \sigma(H) \to (0,\infty)$（默认 $\omega \equiv 1$）；停止阈值 tol

计算 $g_0 = \nabla q(x_0) = Hx_0$
for $j = 0, \cdots, p-1$ **do**
　取预热步长 $\alpha_j \in [\lambda_n^{-1}, \lambda_1^{-1}]$（如 $\alpha_j = 1/\lambda_n$）
　$x_{j+1} \leftarrow x_j - \alpha_j g_j,\ \ g_{j+1} \leftarrow g_j - \alpha_j H g_j$
end for
全局索引 $k \leftarrow p$；最近 p 个梯度的滚动窗口 $G = [g_{k-p}, \cdots, g_{k-1}]$
while $\|g_k\| > \mathrm{tol} \cdot \|g_0\|$ **do**
　对 G 作 thin QR 得 $G = Q_g R_g$；按相对阈值确定有效秩 $r_k \leqslant p$，记 $Q_b \leftarrow Q_g(:, 1:r_k)$
　形成两个 $r_k \times r_k$ 对称矩阵

$$A_k = Q_b^T\, \omega(H)\, H\, Q_b, \qquad B_k = Q_b^T\, \omega(H)\, Q_b$$

$\triangleright\ \omega \equiv 1$ 时 $B_k = I_{r_k}$；$\omega(t) = t$ 时 $A_k = Q_b^T H^2 Q_b,\ \ B_k = Q_b^T H Q_b$

　求解对称广义特征值问题 $A_k z = \theta B_k z$，得到 r_k 个特征值 $\theta_0, \cdots, \theta_{r_k-1} \in [\lambda_1, \lambda_n]$
　步长 $\alpha_\ell \leftarrow 1/\theta_\ell,\ \ell = 0, \cdots, r_k - 1$（按预定顺序排列，例如降序）
　for $r = 0, \cdots, r_k - 1$ **do**
　　$x_{k+1} \leftarrow x_k - \alpha_r g_k,\ \ g_{k+1} \leftarrow g_k - \alpha_r H g_k$
　　把 g_{k+1} 加入 G，丢弃最旧一列
　　$k \leftarrow k + 1$
　end for
end while
return x_k

注 取 $\omega\equiv 1$、$p\geqslant 1$ 即 Fletcher[11] 的 LMSD（S 型）；取 $\omega(t)=t$ 即 Curtis–Guo[13] 与 Ferrandi–Hochstenbach[19] 讨论的 harmonic-LMSD（Y 型）；取 $\omega\equiv 1$、$p=1$ 退化为 BB1 方法，取 $\omega(t)=t$、$p=1$ 退化为 BB2 方法（参见 §五 中作为闭锥原理的例子给出的 R-线性证明）。算法 1 在 thin QR 之后用 r_k 而非固定的 p 作为有效广义特征值问题维数，把秩压缩进行具体实现，当 G 满列秩时它就退化回标准固定 $p\times p$ 实现。

第五节 λ_1-分量的显式收缩界

引理 2.2 由有限覆盖紧致性给出闭锥上的收敛率 $\rho\in(0,1)$，但该 ρ 不是关于 κ 与 p 的显式公式。下面给出最小活跃特征值对应特征分量的逐块显式收缩界，作为闭锥逐点收敛与 Cauchy–Binet 表示的直接推论和收敛速率的一个简单的定量补充。

推论 4.23 在定理 4.18 的条件下，块起点梯度的 λ_1 特征分量满足

$$\left|g^{(1)}_{K,0}\right| \;\leqslant\; \left(1-\frac{1}{\kappa}\right)^{Kp}\left|g^{(1)}_{0,0}\right|, \qquad K\geqslant 0, \tag{39}$$

其中 $\kappa=\lambda_n/\lambda_1$。同样的不等式对 harmonic-LMSD（$Y$ 型）以及 §四 的 ω 族都成立。这与 Fletcher[11] 附录中对单步 sweep 给出的 $|g_1^{k+1}|\leqslant(1-\lambda_1/\lambda_n)^m|g_1^k|$ 形式相同（其中 Fletcher 的 m 即本文的 p）。

证明 按假设 1.1，H 已对角化，故每个特征分量独立演化。对块 $K+1$ 内的 p 步迭代，

$$g^{(1)}_{K+2,0} \;=\; \left[\prod_{r=0}^{p-1}\bigl(1-\alpha_{K+1,r}\lambda_1\bigr)\right] g^{(1)}_{K+1,0}.$$

由定义 3.1，$\alpha_{K+1,r}$ 是 H 在某 Krylov 子空间上的 Ritz 值的倒数，故 $\alpha_{K+1,r}\in[\lambda_n^{-1},\lambda_1^{-1}]$，预热步长（Cauchy、BB1、BB2 或 ω-Ritz）由引理 4.21 落在同一区间。从而每个因子满足 $|1-\alpha_{K+1,r}\lambda_1|\leqslant 1-\lambda_1/\lambda_n=1-1/\kappa$，逐块的 λ_1 分量收缩因子至多为 $(1-1/\kappa)^p$，关于 K 一致。从 $g_{0,0}$ 经 K 块到 $g_{K,0}$ 即得（39）。Y 型与 ω 型情形完全平行，三种规则的步长区间均为 $[\lambda_n^{-1},\lambda_1^{-1}]$。 ∎

注 求一个关于 κ 与 p 的紧的显式 ρ 是一个开放问题。本文第八章第三节将对 $p=1$ 的经典 BB 步长给出一个独立的遍历动力系统证明，说明其在严格凸二次问题上的最坏渐近指数常数为

$$\frac{\kappa-1}{\kappa+1}.$$

不过，这一结果目前只覆盖经典 BB 情形，对于行为更加复杂的一般 pD-BB 方法，是否存在类似的紧显式速率，仍是开放问题。

第五章　一般问题下的对称化算法

本节开始转向一般非线性问题，并给出一种适用于一般非线性问题的对称化步长，以保证算法的稳定性。

第一节　一般非线性问题下的对称化改进步长

在一般非线性情形下，$B_k = S_k^T Y_k$ 不再对称。如果直接取 M_k^{LS} 或 N_k^{LS} 的特征值作为步长，则会遇到两个问题：其一，小矩阵可能出现复特征值；其二，特征值关于历史数据的连续依赖性不如对称广义特征值问题稳定。因此，和 Ferrandi–Hochstenbach[19] 的对称化处理方法类似，本文对原始矩阵最小二乘模型做对称化处理，并将其进一步写成更便于进行局部扰动估计的小规模广义特征值问题。与[19] 不同的是，本文的对称化操作更加直接。

令

$$C_k := \mathrm{sym}(S_k^T Y_k) = \frac{1}{2}\left(S_k^T Y_k + Y_k^T S_k\right). \tag{40}$$

在 S 型与 Y 型两种情况下，分别考虑广义特征值问题

$$C_k z = \theta M_k z, \tag{41}$$

以及

$$C_k w = \tau D_k w. \tag{42}$$

二者都是小规模实对称广义特征值问题。

定义 5.1 (pD(S) 步长)　设 $r_k = \mathrm{rank}(S_k)$。在 rank-revealing QR 压缩后，广义特征值问题（41）共有 r_k 个实特征值，记为

$$\hat{\theta}_{k,0}^{(S)}, \cdots, \hat{\theta}_{k,r_k-1}^{(S)}.$$

若 $\hat{\theta}_{k,r}^{(S)} > 0$，则相应步长为

$$\hat{\alpha}_{k,r}^{pD(S)} = \frac{1}{\hat{\theta}_{k,r}^{(S)}}. \tag{43}$$

若 $\hat{\theta}_{k,r}^{(S)} \leqslant 0$，则该步长由假设 5.1 中的参考步长替代。最后对所有候选步长作上下界截断，并按实际采用的步长顺序在一个块内依次使用。

定义 5.2 (pD(Y) 步长)　设 $r_k = \mathrm{rank}(Y_k)$。在 rank-revealing QR 压缩后，广义特征值问题（42）共有 r_k 个实特征值，记为

$$\hat{\tau}_{k,0}^{(Y)}, \cdots, \hat{\tau}_{k,r_k-1}^{(Y)}.$$

若 $\hat{\tau}_{k,r}^{(Y)} > 0$，则相应步长为

$$\hat{\alpha}_{k,r}^{pD(Y)} = \hat{\tau}_{k,r}^{(Y)}. \tag{44}$$

若 $\hat{\tau}_{k,r}^{(Y)} \leqslant 0$，则该步长由假设 5.1 中的参考步长替代。最后对所有候选步长作上下界截断，并按实际采用的步长顺序在一个块内依次使用。

对于严格凸二次问题，上述步长显然退化为第 3 章的情形，有如下命题成立。

命题 5.1 若目标函数满足（18），则对任意 p 有

(i) $C_k = S_k^T H S_k$，故（41）的广义特征值正是 H 在 span(S_k) 上的 Ritz 值；

(ii) $C_k = Y_k^T H^{-1} Y_k$，故（42）的广义特征值正是 H^{-1} 在 span(Y_k) 上的 Ritz 值；

(iii) 当 $p = 2, 3$ 时，定义 5.1–5.2 与[20]中的 2D(S)/2D(Y)/3D(S)/3D(Y) 公式完全一致。

证明 在二次情形下 $Y_k = H S_k$，故 $B_k = S_k^T H S_k$ 本身对称，于是 $C_k = B_k$。对任意非零 z，有

$$\frac{z^T C_k z}{z^T M_k z} = \frac{(S_k z)^T H (S_k z)}{\|S_k z\|^2},$$

这正是 H 在子空间 span(S_k) 上的 Rayleigh 商。因此（41）的广义特征值就是 H 在该子空间上的 Ritz 值。同理，由 $S_k = H^{-1} Y_k$，对任意非零 w 有

$$\frac{w^T C_k w}{w^T D_k w} = \frac{(Y_k w)^T H^{-1} (Y_k w)}{\|Y_k w\|^2},$$

于是（42）的广义特征值是 H^{-1} 在 span(Y_k) 上的 Ritz 值。结论随即成立。 ∎

第二节 二维显式公式

当 $p = 2$ 时，对称化表述可以写成闭式表达。记

$$S_k = [s_{k,0}, s_{k,1}], \qquad Y_k = [y_{k,0}, y_{k,1}],$$

并记

$$a_k = s_{k,0}^T s_{k,0}, \quad b_k = s_{k,0}^T s_{k,1}, \quad c_k = s_{k,1}^T s_{k,1},$$

$$u_k = s_{k,0}^T y_{k,0}, \quad p_k = s_{k,0}^T y_{k,1}, \quad q_k = s_{k,1}^T y_{k,0}, \quad w_k = s_{k,1}^T y_{k,1},$$

$$d_k = y_{k,0}^T y_{k,0}, \quad e_k = y_{k,0}^T y_{k,1}, \quad f_k = y_{k,1}^T y_{k,1},$$

以及对称化交叉项

$$v_k = \frac{p_k + q_k}{2}.$$

则

$$M_k=\begin{pmatrix} a_k & b_k \\ b_k & c_k \end{pmatrix},\qquad C_k=\begin{pmatrix} u_k & v_k \\ v_k & w_k \end{pmatrix},\qquad D_k=\begin{pmatrix} d_k & e_k \\ e_k & f_k \end{pmatrix}.$$

进一步定义

$$A_k^{(S)}=a_kc_k-b_k^2,\qquad B_k^{(S)}=c_ku_k+a_kw_k-2b_kv_k,$$

$$A_k^{(Y)}=d_kf_k-e_k^2,\qquad B_k^{(Y)}=f_ku_k+d_kw_k-2e_kv_k,$$

$$\Delta_k=u_kw_k-v_k^2,\qquad \Gamma_k^{(S)}=\left(B_k^{(S)}\right)^2-4A_k^{(S)}\Delta_k,\qquad \Gamma_k^{(Y)}=\left(B_k^{(Y)}\right)^2-4A_k^{(Y)}\Delta_k.$$

命题 5.2 (二维显式公式)　若 $A_k^{(S)}>0$，则 2D(S) 型广义特征值为

$$\theta_{k,1}^{(S)}=\frac{B_k^{(S)}+\sqrt{\Gamma_k^{(S)}}}{2A_k^{(S)}},\qquad \theta_{k,2}^{(S)}=\frac{B_k^{(S)}-\sqrt{\Gamma_k^{(S)}}}{2A_k^{(S)}},\tag{45}$$

相应步长为

$$\alpha_{k,1}^{2D(S)}=\frac{1}{\theta_{k,1}^{(S)}},\qquad \alpha_{k,2}^{2D(S)}=\frac{1}{\theta_{k,2}^{(S)}}.\tag{46}$$

若 $A_k^{(Y)}>0$，则 2D(Y) 型特征值为

$$\tau_{k,1}^{(Y)}=\frac{B_k^{(Y)}-\sqrt{\Gamma_k^{(Y)}}}{2A_k^{(Y)}},\qquad \tau_{k,2}^{(Y)}=\frac{B_k^{(Y)}+\sqrt{\Gamma_k^{(Y)}}}{2A_k^{(Y)}},\tag{47}$$

相应步长为

$$\alpha_{k,1}^{2D(Y)}=\tau_{k,1}^{(Y)},\qquad \alpha_{k,2}^{2D(Y)}=\tau_{k,2}^{(Y)}.\tag{48}$$

证明　把（41）与（42）分别写成

$$\det(C_k-\theta M_k)=0,\qquad \det(C_k-\tau D_k)=0,$$

直接展开即可得到两个二次方程：

$$A_k^{(S)}\theta^2-B_k^{(S)}\theta+\Delta_k=0,\qquad A_k^{(Y)}\tau^2-B_k^{(Y)}\tau+\Delta_k=0.$$

求根后即得结论。 ∎

因此，在二维情形中，这种对称化不仅理论上更统一，在书写和实现上也都更加直接。

注　对于 $p=3$，步长由三阶广义特征值问题的三个正根给出；对于 $p=4$，则对应四阶广义特征值问题。

原则上，这些根都可用 Cardano 或 Ferrari 公式写成显式根式表达，但从理论和数值两方面看，这样做都没有必要：算法实现直接调用小规模实对称广义特征值求解器更稳定，也更符合现代数值计算习惯。因此正文不再对三维及四维的步长公式进行闭式推导，相关公式见附录。

第三节　一些约定

为了使算法在所有历史数据上都有定义，本文采用如下约定与保护机制。

假设 5.1　当 S_k 或 Y_k 不满秩或数值上近秩亏时，先秩压缩到判定的有效秩 $r_k \leqslant p$，然后在压缩后的 r_k 维空间上计算广义特征值与候选步长。若某个候选特征值非正，则把对应步长替换为参考步长（例如最近一次 BB1 步长或 Cauchy 步长）。最后对所有候选步长作上下界截断。

上述假设保证了算法在全局范围内始终有定义。

必须强调，上述约定并不仅仅是为了算法鲁棒性而添加的修补，而是处理非线性问题时该算法特性所决定的必要步骤。即便对于强凸乃至严格凸二次问题，如果不加干预也会面临以下失效风险。

例 5.1　考虑严格凸二次函数 $f(x) = \frac{1}{2}x^T \operatorname{diag}(1,2,3,4,5)x$。若某次迭代处的梯度 g_k 仅在前两个分量上非零，由梯度下降法的更新规则，后续的所有 s_j 和 y_j 将永远落在由前两个坐标轴张成的二维子空间中。此时，即使设定历史窗口长度为 $p = 3$ 或更大，最近 p 个位移构成的矩阵 S_k 的实际秩最大也只能是 2。在第 4 章的精确秩压缩二次模型中，这会被解释为二维实际 Krylov 子空间上的短 sweep，并在识别出两个活跃谱后有限终止。但在实际固定维小矩阵公式中，如果不进行秩压缩而直接在 p 维空间求解，矩阵 $M_k = S_k^T S_k$ 必然奇异，导致相应的 pD(S) 小广义特征值问题奇异或退化，部分广义特征值发散并对应零步长，从而引发严重的数值崩溃。

例 5.2　考虑二维目标函数 $f(x,y) = \frac{1}{2}(x^2+y^2) + 5\log\cosh(x-y)$。其 Hessian 矩阵的特征值介于 $m = 1$ 与 $L = 11$ 之间，是严格凸函数。设初始点 $x_0 = (0,0)^T$，第一步迭代至 $x_1 = (2,-2)^T$，第二步迭代至 $x_2 = (4,-3.9)^T$。此时位移向量为 $s_1 = (2,-2)^T$，$s_2 = (2,-1.9)^T$。显然 $S_k = [s_1, s_2]$ 满秩，梯度差近似为 $y_1 \approx (7,-7)^T$，$y_2 \approx (2,-1.9)^T$。计算对称化矩阵 $C_k = \operatorname{sym}(S_k^T Y_k)$：$C_{11} = s_1^T y_1 \approx 28$，$C_{22} = s_2^T y_2 \approx 7.61$，$C_{12} = \frac{1}{2}(s_1^T y_2 + s_2^T y_1) \approx \frac{1}{2}(7.8 + 27.3) = 17.55$。其行列式 $\det(C_k) \approx 28 \times 7.61 - 17.55^2 = 213.08 - 308.0025 < 0$。说明 C_k 是不定矩阵，存在负特征值，因此上述约定在代数上是必要的。

注　若要在理论上全局规避负步长，仅有强凸性是不够的。我们至少需要进一步假设 Hessian 满足 Lipschitz 连续性（即 $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leqslant M\|x-y\|$），并且严格限制算法的步长上限，使得块内的总位移 $\|s_1\| + \|s_2\|$ 足够小。此时 Hessian 的变化量被压缩，才能确保对角线元素的正定性大于交叉项带来的偏差。然而，在实际算法中，强行限制步长上限违背了 BB 方法“允许大步长以加速收敛”的核心初衷。因此，保留上述约定的保护机制是在实际求解过程中兼顾效率与全局收敛性的最佳选择。对于 Hessian 满足局部 Lipschitz 连续性假设的情形，我们将

在后文第 6 章给出更加详细的分析及结论。

第四节　纯算法与 GLL 全局化算法

在上述对称化步长的基础上，我们先给出块式实现。为简洁起见，仅写 S 型版本，Y 型版本只需把 (M_k, C_k) 替换为 (D_k, C_k) 并修改相应步长即可。

算法 2 改进 pD(S)-BB 步长生成器

Require: 最近 p 个割线对 $\{(s_j, y_j)\}$，步长上下界 $[\underline{\alpha}, \overline{\alpha}]$

Ensure: 一个短块内的步长序列 $\alpha_0, \cdots, \alpha_{r_k-1}$

组装历史矩阵 $S = [s_1, \cdots, s_p]$，$Y = [y_1, \cdots, y_p]$
对 S 作 rank-revealing QR/SVD 压缩，得到有效秩 r_k 及压缩后的 S, Y
计算 $M = S^T S$，$C = \frac{1}{2}(S^T Y + Y^T S)$
求解对称广义特征值问题 $Cz = \theta M z$，得到 r_k 个实特征值 $\hat{\theta}_0, \cdots, \hat{\theta}_{r_k-1}$
for $r = 0, \cdots, r_k - 1$ **do**
    if $\hat{\theta}_r > 0$ **then**
        令候选步长 $\widehat{\alpha}_r = 1/\hat{\theta}_r$
    else
        用参考步长替代 $\widehat{\alpha}_r$
    end if
    令 $\alpha_r = \min\{\overline{\alpha}, \max\{\underline{\alpha}, \widehat{\alpha}_r\}\}$
end for
按预定顺序排列 $\alpha_0, \cdots, \alpha_{r_k-1}$
return $\alpha_0, \cdots, \alpha_{r_k-1}$

为与前文第 4 章中的内容区分，下面先给出实际计算时更自然的 “带秩压缩的纯实现”，再引入 GLL 检测下的全局化版本。

算法 3 带秩压缩的纯 pD-BB 梯度法

Require: 初始点 x_0，窗口长度 p

计算 $g_0 = \nabla f(x_0)$

for $j = 0, \cdots, p-1$ **do**

    取一个正的预热步长 α_j

    更新 $x_{j+1} = x_j - \alpha_j g_j$，计算 $g_{j+1} = \nabla f(x_{j+1})$

    存储割线对 $s_j = x_{j+1} - x_j$，$y_j = g_{j+1} - g_j$

end for

全局索引 $k \leftarrow p$；最近 p 个割线对的滚动窗口

while 停止准则未满足 **do**

    用算法 2 基于最近 p 个割线对计算短块步长 $\alpha_0, \cdots, \alpha_{r_k-1}$

    for $r = 0, \cdots, r_k - 1$ **do**

        更新 $x_{k+1} = x_k - \alpha_r g_k$，计算 $g_{k+1} = \nabla f(x_{k+1})$

        生成新割线对 $s_k = x_{k+1} - x_k$，$y_k = g_{k+1} - g_k$，加入滚动窗口并丢弃最旧割线对

        $k \leftarrow k + 1$

    end for

end while

下面给出 GLL 检测下的 pD-BB 算法。为简洁起见，仍然仅写 S 型版本。

算法 4 GLL 全局化的改进 pD-BB 梯度法

Require: 初始点 x_0，窗口长度 p，参数 $\underline{\alpha}, \overline{\alpha}, \beta, \sigma, M$

计算 $g_0 = \nabla f(x_0)$

for $j = 0, \cdots, p-1$ **do**

  取预热试探步长 $\tilde{\alpha}_j \in [\underline{\alpha}, \overline{\alpha}]$

  取最小 $h \geqslant 0$ 使

$$f(x_j - \beta^h \tilde{\alpha}_j g_j) \leqslant \max_{0 \leqslant \ell \leqslant \min\{M-1, j\}} f(x_{j-\ell}) - \sigma \beta^h \tilde{\alpha}_j \left\| g_j \right\|^2$$

  令 $\alpha_j = \beta^h \tilde{\alpha}_j$，更新 $x_{j+1} = x_j - \alpha_j g_j$，计算 $g_{j+1} = \nabla f(x_{j+1})$

  存储割线对 $s_j = x_{j+1} - x_j$，$y_j = g_{j+1} - g_j$

end for

全局索引 $k \leftarrow p$；最近 p 个割线对的滚动窗口

while 停止准则未满足 **do**

  用算法 2 计算短块试探步长 $\bar{\alpha}_0, \cdots, \bar{\alpha}_{r_k - 1}$

  for $r = 0, \cdots, r_k - 1$ **do**

    设 $\tilde{\alpha}_r = \min\{\overline{\alpha}, \max\{\underline{\alpha}, \bar{\alpha}_r\}\}$

    取最小 $h \geqslant 0$ 使

$$f(x_k - \beta^h \tilde{\alpha}_r g_k) \leqslant \max_{0 \leqslant \ell \leqslant \min\{M-1, k\}} f(x_{k-\ell}) - \sigma \beta^h \tilde{\alpha}_r \left\| g_k \right\|^2$$

    令 $\alpha_k = \beta^h \tilde{\alpha}_r$

    更新 $x_{k+1} = x_k - \alpha_k g_k$，计算 $g_{k+1} = \nabla f(x_{k+1})$

    生成新割线对 $s_k = x_{k+1} - x_k$，$y_k = g_{k+1} - g_k$，加入滚动窗口并丢弃最旧割线对

    $k \leftarrow k + 1$

  end for

end while

注 算法 3 与算法 4 有同样的“块首生成、块内依次执行”的队列骨架。前者是带压缩、无 GLL 的纯方法实现，后者则是在此基础上额外加入步长检测与 GLL 非单调接受准则的全局化版本。

第六章　GLL 全局化下 pD(S)-BB 算法的收敛与 R-线性收敛

本节通过块映射的视角把分析真正落到一般非线性算法上。简便起见，下面的局部扰动估计、块映射比较与局部 R-线性收敛都只对第 5 章给出的 S 型算法展开，对于 Y 型，只需要引入与之平行的步长扰动估计和局部比较引理便可完全类似的得到相应结论。本节证明分成三步：先用 GLL 非单调线搜索得到全局收敛，再在最优解邻域内比较真实块映射与局部二次模型块映射，最后把已经得到的二次块收敛性质传递到一般非线性算法。

第一节　GLL 检测下 pD(S)-BB 算法的全局收敛性

假设 6.1　函数 $f \in C^1(\mathbb{R}^n)$ 有下界，并且初始水平集

$$\Omega_0 := \{x \in \mathbb{R}^n : \ f(x) \leqslant f(x_0)\}$$

有界。此外，∇f 在某个包含 Ω_0 的邻域上 Lipschitz 连续。

命题 6.1　设假设 6.1 与 5.1 成立。则算法 4 要么在有限步终止于驻点，要么满足

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{49}$$

若进一步假设 f 强凸，则整个迭代序列收敛到唯一极小点 x^*。

证明　这是标准 GLL 型非单调梯度法全局收敛定理的一个特例。由于算法 4 在预热阶段和后续块执行阶段都采用同一套 GLL 回溯规则，故整条迭代序列的每一步都满足下面三条性质：

(i) 搜索方向始终是 $-g_k$；

(ii) 回溯前的试探步长始终落在固定区间 $[\underline{\alpha}, \overline{\alpha}]$ 内；

(iii) 接受准则始终是 GLL 非单调条件。

Raydan 对 GBB 方法的证明[5] 真正用到的也是这三点，而并不必要求试探步长来自经典 BB 的具体表达式。进一步，若 f 强凸，则驻点唯一，从而全序列收敛到唯一极小点 x^*。 ∎

第二节　关于无 GLL 检测时算法全局收敛性的反例

经典 BB 方法在一般非二次强凸问题上若不配合非单调线搜索，便不成立一般性的全局收敛结论。Raydan 引入 GLL 全局化也正是为了解决这一点[5]。对于本文讨论的纯 pD(S)-BB 算法 3，同样应当作出谨慎判断。

命题 6.2 设 $\varphi:\mathbb{R}\to\mathbb{R}$ 是一维 C^1 强凸函数，并且纯 BB1 迭代

$$\xi_{k+1}=\xi_k-\alpha_k\varphi'(\xi_k),\qquad \alpha_k=\frac{(\xi_k-\xi_{k-1})^2}{(\xi_k-\xi_{k-1})(\varphi'(\xi_k)-\varphi'(\xi_{k-1}))}$$

从某组初始点出发不收敛。对任意 $n\geqslant 2$ 与任意 $\gamma>0$，定义高维函数

$$F(x)=\varphi(x_1)+\frac{\gamma}{2}\sum_{i=2}^{n}x_i^2. \tag{50}$$

考虑任意带秩压缩的纯 pD(S)-BB 实现。若所有预热点都位于一维子空间

$$\mathcal{E}:=\operatorname{span}\{e_1\},$$

则全部后续迭代点都保持在 $\mathcal{E}$ 内，且整个高维 pD(S)-BB 方法精确退化为同一条一维 BB1 轨道。因此，任何一维 BB1 的非收敛反例都可直接提升为高维 pD(S)-BB 的非收敛反例。特别地，Burdakov–Dai–Huang 在 2019 年构造的 BB1 四点循环反例[15]可直接提升为任意维的纯 pD(S)-BB 反例。

证明 由（50）的梯度结构，若 $x=(\xi,0,\cdots,0)^T\in\mathcal{E}$，则

$$\nabla F(x)=\big(\varphi'(\xi),0,\cdots,0\big)^T\in\mathcal{E}.$$

因此，沿负梯度方向更新不会离开 $\mathcal{E}$，只要预热点属于 $\mathcal{E}$，全部迭代点与梯度都始终属于该一维不变子空间。于是最近 p 个历史向量 s_j,y_j 都与 e_1 共线，历史矩阵 S_k,Y_k 的实际秩恒为 1。在秩压缩之后，S 型广义特征值问题退化为

$$(s^Ty)z=\theta(s^Ts)z,$$

因此唯一正步长为

$$\alpha=\frac{1}{\theta}=\frac{s^Ts}{s^Ty},$$

这正是 BB1，故高维算法在 $\mathcal{E}$ 上与一维 BB1 迭代完全一致。若一维轨道不收敛，则对应的高维轨道也不收敛。最后，把 φ 取为 Burdakov–Dai–Huang 所构造的一维反例[15]即可得到高维四点循环。 ∎

命题 6.2 说明：高维 S 型公式在秩 1 情形下自然退化为 BB1，因此在没有 GLL 或其他全局化机制时就不可能拥有“对所有强凸非二次函数全局收敛”的一般性结论，从而说明了 GLL 全局化的必要性。

第三节　局部假设与块映射

以下所有局部分析都在一次固定平移与正交变换后的坐标系中进行。也就是说，可先把局部极小点移到原点，再用一次固定的正交变换把 H_* 对角化。由于本文算法只依赖欧氏内积、范数、梯度以及由历史向量构造的小规模广义特征值问题，这种一次性的坐标变换不改变算法的形式。于是以下不失一般性地设 $f(0)=0$，并在该局部坐标系中讨论。

假设 6.2 函数 f 在原点的某个邻域 U 内属于 $C^{2,1}$，且

$$\nabla f(0) = 0, \qquad H_* := \nabla^2 f(0) \succ 0.$$

在上述固定坐标系中，可进一步把

$$H_* = \nabla^2 f(0) = \mathrm{diag}(\mu_1, \cdots, \mu_n), \qquad 0 < \mu_1 \leqslant \cdots \leqslant \mu_n$$

写成对角形式。缩小邻域后，可取常数 $0 < m \leqslant L$ 使

$$mI \preceq \nabla^2 f(x) \preceq LI, \qquad \forall x \in U. \tag{51}$$

于是

$$m\|x\| \leqslant \|\nabla f(x)\| \leqslant L\|x\|, \qquad \forall x \in U. \tag{52}$$

固定一个块长 p。局部理论继续沿用块记号：第 k 个完整块记为

$$E_k = (x_{k,0}, x_{k,1}, \cdots, x_{k,p}), \qquad s_{k,r} = x_{k,r+1} - x_{k,r}, \qquad r = 0, \cdots, p-1.$$

对任意块状态

$$E = (x_0, x_1, \cdots, x_p) \in (\mathbb{R}^n)^{p+1},$$

定义历史位移束

$$S(E) = [x_1 - x_0,\ x_2 - x_1,\ \cdots,\ x_p - x_{p-1}],$$

并在 $(\mathbb{R}^n)^{p+1}$ 上取块范数

$$\|E\|_\infty := \max_{0 \leqslant r \leqslant p} \left\|x_r\right\|.$$

考虑局部二次模型

$$q(x) = \frac{1}{2} x^T H_* x. \tag{53}$$

由状态 E 的相邻差分可分别构造真实割线与二次模型割线

$$s_r = x_{r+1} - x_r, \qquad y_r^f(E) = \nabla f(x_{r+1}) - \nabla f(x_r),$$
$$y_r^q(E) = H_*(x_{r+1} - x_r), \qquad r = 0, \cdots, p-1.$$

把它们代入本文的 S 型步长生成公式，得到真实函数与二次模型下的整块步长

$$\alpha_0^f(E), \cdots, \alpha_{p-1}^f(E), \qquad \alpha_0^q(E), \cdots, \alpha_{p-1}^q(E).$$

由于下一完整块从当前块的最后一个点开始，我们定义块内递推

$$\eta_0^f = x_p, \qquad \eta_{r+1}^f = \eta_r^f - \alpha_r^f(E)\nabla f(\eta_r^f), \qquad r = 0, \cdots, p-1,$$

以及

$$\eta_0^q = x_p, \qquad \eta_{r+1}^q = \eta_r^q - \alpha_r^q(E) H_* \eta_r^q, \qquad r = 0, \cdots, p-1.$$

相应地定义真实块映射与二次模型块映射

$$\mathcal{T}_f(E) = (\eta_0^f, \eta_1^f, \cdots, \eta_p^f), \qquad \mathcal{T}_q(E) = (\eta_0^q, \eta_1^q, \cdots, \eta_p^q). \tag{54}$$

记块单位球面为

$$\mathbb{S}_\infty := \{E \in (\mathbb{R}^n)^{p+1} : \|E\|_\infty = 1\}.$$

我们做如下局部正则化假设。

假设 6.3 存在开锥 $\Omega_{\mathrm{reg}} \subset (\mathbb{R}^n)^{p+1} \setminus \{0\}$、紧集 $K \subset \Omega_{\mathrm{reg}} \cap \mathbb{S}_\infty$ 与半径 $\delta_{\mathrm{reg}} > 0$。记

$$\mathcal{U}_{\mathrm{reg}} := \{0\} \cup \{E \in \Omega_{\mathrm{reg}} : 0 < \|E\|_\infty < \delta_{\mathrm{reg}},\ E/\|E\|_\infty \in K\}.$$

假设满足：

(i) 在 $\mathcal{U}_{\mathrm{reg}} \setminus \{0\}$ 上，秩压缩、负特征值替代等保护机制均不触发，二次步长 $E \mapsto \alpha_i^q(E)$ 由定义 3.1 给出，真实步长 $E \mapsto \alpha_i^f(E)$ 由定义 5.1 给出，且存在常数 $0 < \underline{\alpha} \leqslant \overline{\alpha} < \infty$ 使

$$\underline{\alpha} \leqslant \alpha_i^f(E), \alpha_i^q(E) \leqslant \overline{\alpha}, \qquad i = 0, \cdots, p-1;$$

此外，归一化历史束在 K 上保持一致秩正则：存在 $\gamma_S > 0$，使得对所有 $z \in K$，压缩后的位移矩阵 $S(z)$ 满足 $\sigma_{\min}(S(z)) \geqslant \gamma_S$。

(ii) $\mathcal{T}_f(\mathcal{U}_{\mathrm{reg}}) \subset \mathcal{U}_{\mathrm{reg}}$ 且 $\mathcal{T}_q(\mathcal{U}_{\mathrm{reg}}) \subset \mathcal{U}_{\mathrm{reg}}$；

(iii) 对每个 $z \in K$，满足 $\mathcal{T}_q^k(z) \to 0$。

假设 6.3 把一般非线性局部分析中仍需控制的“正则尾轨道”写成严格的数学假设。在一般非线性情形下，由于历史束降秩、非正特征值和近秩亏等触发保护机制的情形，假设 6.3 这样的局部正则性条件是必要的。与 Curtis–Guo[13] 的 Assumption 3.4(ii) 相比，这里同样要求尾部历史束远离退化边界。本文的表述把这类非退化控制写成归一化紧集 K 上的奇异值下界，并额外要求保护机制最终不触发。在该假设下，当块状态位于 $\mathcal{U}_{\mathrm{reg}}$ 时，$\mathcal{T}_f(E)$ 与 $\mathcal{T}_q(E)$ 分别对应真实算法与局部二次模型在下一完整块上的状态。

第四节　局部二次逼近与步长扰动估计

引理 6.3 在假设 6.2 下，存在邻域 $U_0 \subset U$ 与常数 $c_0 > 0$，使得对任意 $x, y \in U_0$ 有

$$\left\|\nabla f(x) - H_* x\right\| \leqslant c_0 \|x\|^2, \tag{55}$$

以及

$$\left\|\big(\nabla f(x) - \nabla f(y)\big) - H_*(x-y)\right\| \leqslant c_0\big(\|x\| + \|y\|\big)\|x-y\|. \tag{56}$$

证明 由假设 6.2，函数 f 在原点附近属于 $C^{2,1}$，其 Hessian 在原点附近局部 Lipschitz 连续。故可取原点的一个充分小的凸邻域 $U_0 \subset U$ 与常数 $L > 0$，使得对任意 $z \in U_0$ 都有

$$\left\|\nabla^2 f(z) - H_*\right\| \leqslant L \|z\| .$$

又因为 0 是局部极小点，故 $\nabla f(0) = 0$。

对任意 $x \in U_0$，由积分形式的 Taylor 公式，

$$\nabla f(x) = \int_0^1 \nabla^2 f(tx)\, x \,\mathrm{d}t.$$

减去 $H_* x$ 后取范数，得到

$$\begin{aligned}
\left\|\nabla f(x) - H_* x\right\| &= \left\| \int_0^1 \left(\nabla^2 f(tx) - H_*\right) x \,\mathrm{d}t \right\| \\
&\leqslant \int_0^1 \left\|\nabla^2 f(tx) - H_*\right\| \|x\| \,\mathrm{d}t \\
&\leqslant \int_0^1 Lt \|x\|^2 \,\mathrm{d}t = \frac{L}{2} \|x\|^2 .
\end{aligned}$$

对任意 $x, y \in U_0$，同样有

$$\nabla f(x) - \nabla f(y) = \int_0^1 \nabla^2 f\big(y + t(x-y)\big)(x-y) \,\mathrm{d}t.$$

减去 $H_*(x-y)$ 后取范数，得到

$$\begin{aligned}
\left\|\big(\nabla f(x) - \nabla f(y)\big) - H_*(x-y)\right\| &= \left\| \int_0^1 \left(\nabla^2 f(y + t(x-y)) - H_*\right)(x-y) \,\mathrm{d}t \right\| \\
&\leqslant \int_0^1 \left\|\nabla^2 f(y + t(x-y)) - H_*\right\| \|x-y\| \,\mathrm{d}t \\
&\leqslant L \|x-y\| \int_0^1 \|(1-t)y + tx\| \,\mathrm{d}t \\
&\leqslant L \|x-y\| \int_0^1 \big((1-t)\|y\| + t\|x\|\big) \,\mathrm{d}t \\
&= \frac{L}{2}\big(\|x\| + \|y\|\big) \|x-y\| .
\end{aligned}$$

取 $c_0 = L/2$ 即得结论。 ∎

引理 6.4 在假设 6.2 与 6.3 下，存在模函数 $\omega_1 : [0,\infty) \to [0,\infty)$，满足 $\omega_1(t) \to 0$ （$t \downarrow 0$），使得只要 $\rho := \|E\|_\infty$ 充分小，就有

$$\left|\alpha_i^f(E) - \alpha_i^q(E)\right| \leqslant \omega_1\big(\|E\|_\infty\big), \qquad i = 0, \cdots, p-1. \tag{57}$$

证明 记

$$E = (x_0, \cdots, x_p), \qquad s_i = x_{i+1} - x_i, \qquad S = S(E) = [s_0, \cdots, s_{p-1}],$$

以及

$$Y^f(E) = [y_0^f(E), \cdots, y_{p-1}^f(E)], \qquad Y^q(E) = [y_0^q(E), \cdots, y_{p-1}^q(E)] = H_* S.$$

由引理 6.3 的（56），对每个 $i = 0, \cdots, p-1$ 都有

$$\left\| y_i^f(E) - y_i^q(E) \right\| \leqslant c_0 \left(\|x_{i+1}\| + \|x_i\| \right) \|s_i\| \leqslant 2c_0 \rho \|s_i\|.$$

令

$$R(E) := Y^f(E) - Y^q(E),$$

则存在只依赖于 p 和 c_0 的常数 $c_1 > 0$，使得

$$\|R(E)\| \leqslant c_1 \rho \|S\|.$$

对本文当前的 S 型步长，分母矩阵不随模型改变：

$$M(E) := S^T S.$$

而分子矩阵满足

$$C_f(E) - C_q(E) = \mathrm{sym}\left(S^T Y^f(E)\right) - \mathrm{sym}\left(S^T H_* S\right) = \mathrm{sym}\left(S^T R(E)\right),$$

故存在常数 $c_2 > 0$ 使得

$$\left\| C_f(E) - C_q(E) \right\| \leqslant c_2 \rho \|S\|^2.$$

现在在可能包含秩压缩后的有效子空间上讨论广义特征值。设

$$M(E) = V \Lambda V^T$$

为其标准特征分解，其中 $\Lambda \succ 0$ 为对角阵，V 为正交矩阵，并令

$$W := V \Lambda^{-1/2}.$$

则

$$W^T M(E) W = I.$$

于是二次模型的小广义特征值问题

$$C_q(E) z = \theta M(E) z$$

等价于标准对称特征值问题

$$\widehat{C}_q(E) u = \theta u, \qquad \widehat{C}_q(E) := W^T C_q(E) W.$$

又由于

$$\widehat{C}_q(E) = (SW)^T H_* (SW), \qquad (SW)^T (SW) = I,$$

其特征值正是 H_* 在 span(S) 上的 Ritz 值，因而全部落在 $[m, L]$ 内。对真实模型同理定义

$$\widehat{C}_f(E) := W^T C_f(E) W.$$

由假设 6.3(i) 的一致秩正则性，若 $E \in \mathcal{U}_{\mathrm{reg}}$ 且 $\rho = \|E\|_\infty$，则压缩后的位移矩阵满足

$$\sigma_{\min}(S(E)) \geqslant \gamma_S \rho, \qquad \|S(E)\| \leqslant c_S \rho$$

对某个常数 $c_S > 0$ 成立。因此

$$\|W\|_2 \leqslant \frac{1}{\gamma_S \rho}.$$

结合 $\left\|C_f(E) - C_q(E)\right\| \leqslant c_2 \rho \|S\|^2$ 得到

$$\left\|\widehat{C}_f(E) - \widehat{C}_q(E)\right\| = \left\|W^T (C_f(E) - C_q(E)) W\right\| \leqslant \frac{c_2 c_S^2}{\gamma_S^2} \rho.$$

把常数重新记为 c_2。于是由 Weyl 扰动定理，二次模型与真实模型对应的特征值满足

$$|\theta_i^f(E) - \theta_i^q(E)| \leqslant c_2 \rho, \qquad i = 0, \cdots, p-1.$$

而 S 型步长满足 $\alpha_i = 1/\theta_i$，且在区间 $[m, L]$ 上映射 $\theta \mapsto 1/\theta$ 是 Lipschitz 连续的，故存在常数 $c_3 > 0$ 使

$$\left|\alpha_i^f(E) - \alpha_i^q(E)\right| \leqslant c_3 |\theta_i^f(E) - \theta_i^q(E)| \leqslant c_3 c_2 \rho.$$

取 $\omega_1(t) = c_2 c_3 t$ 即得（57）。 ∎

引理 6.5 在假设 6.2、6.3 下，存在模函数 $\omega : [0, \infty) \to [0, \infty)$，满足 $\omega(t) \to 0$ $(t \downarrow 0)$，使得当 $\rho := \|E\|_\infty$ 充分小时有

$$\left\|\mathcal{T}_f(E) - \mathcal{T}_q(E)\right\|_\infty \leqslant \omega\big(\|E\|_\infty\big)\, \|E\|_\infty . \tag{58}$$

证明 记 $\rho = \|E\|_\infty$。由前面的步长界与（51），存在与 E 无关的常数 $\alpha_{\max} > 0$，使得块内所有步长都满足

$$0 < \alpha_i^f(E), \alpha_i^q(E) \leqslant \alpha_{\max}, \qquad i = 0, \cdots, p-1.$$

又由于 ∇f 在 U 上局部 Lipschitz，存在常数 $L_f > 0$ 使

$$\|\nabla f(x)\| \leqslant L_f \|x\|, \qquad x \in U.$$

由块映射定义，

$$\eta_0^f = \eta_0^q = x_p,$$

并且对 $i = 0, \cdots, p-1$ 有

$$\eta_{i+1}^f = \eta_i^f - \alpha_i^f(E) \nabla f(\eta_i^f), \qquad \eta_{i+1}^q = \eta_i^q - \alpha_i^q(E) H_* \eta_i^q.$$

因此存在常数

$$K := \max\{1 + \alpha_{\max} L_f,\ 1 + \alpha_{\max} \left\|H_*\right\|\}$$

使得

$$\left\|\eta_{i+1}^f\right\| \leqslant K\left\|\eta_i^f\right\|, \qquad \left\|\eta_{i+1}^q\right\| \leqslant K\left\|\eta_i^q\right\|.$$

由于 $\left\|\eta_0^f\right\| = \left\|\eta_0^q\right\| = \|x_p\| \leqslant \rho$，递推可得

$$\left\|\eta_i^f\right\| \leqslant K^p\rho, \qquad \left\|\eta_i^q\right\| \leqslant K^p\rho, \qquad i = 0, \cdots, p.$$

记 $c_5 := K^p$。

下面比较两种块内递推。设

$$\delta_i := \left\|\eta_i^f - \eta_i^q\right\|.$$

则对 $i = 0, \cdots, p-1$，

$$\begin{aligned}\eta_{i+1}^f - \eta_{i+1}^q = &\left(I - \alpha_i^q(E)H_*\right)(\eta_i^f - \eta_i^q)\\ &- \left(\alpha_i^f(E) - \alpha_i^q(E)\right)H_*\eta_i^f\\ &- \alpha_i^f(E)\left(\nabla f(\eta_i^f) - H_*\eta_i^f\right).\end{aligned}$$

取范数，并利用引理 6.4 与引理 6.3，得到

$$\delta_{i+1} \leqslant M\delta_i + c_6\omega_1(\rho)\rho + c_7\rho^2,$$

其中

$$M := \sup_{\alpha\in[0,\alpha_{\max}]}\left\|I - \alpha H_*\right\| < \infty,$$

而常数 $c_6, c_7 > 0$ 与 E 无关。令

$$\gamma(\rho) := c_6\omega_1(\rho) + c_7\rho.$$

则 $\gamma(\rho) \to 0$ 当 $\rho \downarrow 0$，并且

$$\delta_{i+1} \leqslant M\delta_i + \gamma(\rho)\rho, \qquad \delta_0 = 0.$$

由于块长 p 固定，展开这个线性递推得到

$$\max_{0\leqslant i\leqslant p}\delta_i \leqslant \Big(\sum_{j=0}^{p-1} M^j\Big)\gamma(\rho)\rho.$$

记

$$\omega(\rho) := \Big(\sum_{j=0}^{p-1} M^j\Big)\gamma(\rho).$$

则 $\omega(\rho) \to 0$ 当 $\rho \downarrow 0$，并且

$$\left\|\mathcal{T}_f(E) - \mathcal{T}_q(E)\right\|_\infty = \max_{0\leqslant i\leqslant p}\delta_i \leqslant \omega(\rho)\rho.$$

即得（58）。 ∎

引理 6.6 在假设 6.2、6.3 下，存在半径 $\delta_L \in (0, \delta_{\mathrm{reg}})$ 与常数 $L_{\mathcal{T}} > 0$，使得只要

$$E, \widetilde{E} \in \mathcal{U}_{\mathrm{reg}}, \qquad \|E\|_\infty, \left\|\widetilde{E}\right\|_\infty \leqslant \delta_L,$$

就有

$$\left\|\mathcal{T}_f(E) - \mathcal{T}_f(\widetilde{E})\right\|_\infty \leqslant L_{\mathcal{T}} \left\|E - \widetilde{E}\right\|_\infty.$$

证明 在 $\mathcal{U}_{\mathrm{reg}} \setminus \{0\}$ 上，由于特征值关于实对称矩阵元素局部 Lipschitz，因此步长函数 $E \mapsto \alpha_i^f(E)$ 也局部 Lipschitz。缩小 δ_L 后，可取统一常数 $L_\alpha > 0$，使得对 $i = 0, \cdots, p-1$ 及所有满足 $\|E\|_\infty, \left\|\widetilde{E}\right\|_\infty \leqslant \delta_L$ 的 $E, \widetilde{E} \in \mathcal{U}_{\mathrm{reg}}$ 都有

$$|\alpha_i^f(E) - \alpha_i^f(\widetilde{E})| \leqslant L_\alpha \left\|E - \widetilde{E}\right\|_\infty.$$

又因为 ∇f 在 U 上局部 Lipschitz，缩小 δ_L 后，可取常数 $L_g > 0$ 使

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L_g \|x - y\|, \qquad \|\nabla f(x)\| \leqslant L_g \|x\|,$$

对所有 $\|x\|, \|y\| \leqslant c_5 \delta_L$ 成立，其中 c_5 是引理 6.5 中给出的块内有界常数。

记

$$\mathcal{T}_f(E) = (\eta_0, \eta_1, \cdots, \eta_p), \qquad \mathcal{T}_f(\widetilde{E}) = (\tilde{\eta}_0, \tilde{\eta}_1, \cdots, \tilde{\eta}_p),$$

并设

$$\Delta_i := \left\|\eta_i - \tilde{\eta}_i\right\|.$$

由定义有 $\eta_0 = x_p$、$\tilde{\eta}_0 = \tilde{x}_p$，故

$$\Delta_0 \leqslant \left\|E - \widetilde{E}\right\|_\infty.$$

对 $i = 0, \cdots, p-1$，块内递推给出

$$\eta_{i+1} - \tilde{\eta}_{i+1} = \eta_i - \tilde{\eta}_i - \big(\alpha_i^f(E) - \alpha_i^f(\widetilde{E})\big)\nabla f(\eta_i) - \alpha_i^f(\widetilde{E})\big(\nabla f(\eta_i) - \nabla f(\tilde{\eta}_i)\big).$$

取范数并利用步长上界、梯度局部 Lipschitz 以及 $\left\|\eta_i\right\| \leqslant c_5 \delta_L$，可得

$$\Delta_{i+1} \leqslant (1 + \overline{\alpha} L_g)\Delta_i + L_\alpha L_g c_5 \delta_L \left\|E - \widetilde{E}\right\|_\infty.$$

由于块长 p 固定，展开这一递推可得到

$$\max_{0 \leqslant i \leqslant p} \Delta_i \leqslant L_{\mathcal{T}} \left\|E - \widetilde{E}\right\|_\infty$$

对某个与 δ_L 无关的常数 $L_{\mathcal{T}} > 0$ 成立。即得所证。 ∎

推论 6.7 对任意固定整数 $N \geqslant 1$，存在模函数 ω_N 满足 $\omega_N(t) \to 0$（$t \downarrow 0$）时，使得对足够小的 $\|E\|_\infty$ 有

$$\left\|\mathcal{T}_f^N(E) - \mathcal{T}_q^N(E)\right\|_\infty \leqslant \omega_N\big(\|E\|_\infty\big)\, \|E\|_\infty. \tag{59}$$

证明 我们通过对迭代次数 k $(1 \leqslant k \leqslant N)$ 进行归纳来证明。当 $k = 1$ 时，由引理 6.5 知结论成立。

假设对某个 $k \geqslant 1$，存在模函数 $\omega_k(\cdot)$ 使得 $\left\|\mathcal{T}_f^k(E) - \mathcal{T}_q^k(E)\right\|_\infty \leqslant \omega_k(\|E\|_\infty)\|E\|_\infty$。考察 $k+1$ 步的偏差，通过加入交叉项 $\mathcal{T}_f(\mathcal{T}_q^k(E))$ 并利用三角不等式，有：

$$\begin{aligned}\left\|\mathcal{T}_f^{k+1}(E) - \mathcal{T}_q^{k+1}(E)\right\|_\infty &= \left\|\mathcal{T}_f(\mathcal{T}_f^k(E)) - \mathcal{T}_q(\mathcal{T}_q^k(E))\right\|_\infty \\ &\leqslant \left\|\mathcal{T}_f(\mathcal{T}_f^k(E)) - \mathcal{T}_f(\mathcal{T}_q^k(E))\right\|_\infty + \left\|\mathcal{T}_f(\mathcal{T}_q^k(E)) - \mathcal{T}_q(\mathcal{T}_q^k(E))\right\|_\infty .\end{aligned} \tag{60}$$

对（60）右端第一项，由引理 6.6，块映射 $\mathcal{T}_f$ 在充分小邻域内局部 Lipschitz。设相应 Lipschitz 常数为 $L_{\mathcal{T}} > 0$。再由归纳假设，该项可放缩为

$$\left\|\mathcal{T}_f(\mathcal{T}_f^k(E)) - \mathcal{T}_f(\mathcal{T}_q^k(E))\right\|_\infty \leqslant L_{\mathcal{T}}\left\|\mathcal{T}_f^k(E) - \mathcal{T}_q^k(E)\right\|_\infty \leqslant L_{\mathcal{T}}\omega_k\big(\|E\|_\infty\big)\|E\|_\infty .$$

对（60）右端第二项，将其看作是从初始状态 $\tilde{E} = \mathcal{T}_q^k(E)$ 出发执行一步块映射的偏差。由二次局部模型的算子有界性，存在常数 $C_k > 0$ 使得 $\left\|\mathcal{T}_q^k(E)\right\|_\infty \leqslant C_k\|E\|_\infty$。由引理 6.5，有：

$$\begin{aligned}\left\|\mathcal{T}_f(\tilde{E}) - \mathcal{T}_q(\tilde{E})\right\|_\infty &\leqslant \omega\big(\left\|\tilde{E}\right\|_\infty\big)\left\|\tilde{E}\right\|_\infty \\ &\leqslant \omega\big(C_k\|E\|_\infty\big)C_k\|E\|_\infty .\end{aligned}$$

将两项合并，得到 $k+1$ 步的偏差界为：

$$\left\|\mathcal{T}_f^{k+1}(E) - \mathcal{T}_q^{k+1}(E)\right\|_\infty \leqslant \underbrace{\Big[L_{\mathcal{T}}\omega_k\big(\|E\|_\infty\big) + C_k\,\omega\big(C_k\|E\|_\infty\big)\Big]}_{:=\omega_{k+1}(\|E\|_\infty)}\|E\|_\infty .$$

显然，当 $\|E\|_\infty \downarrow 0$ 时，$\omega_k \to 0$ 且 $\omega \to 0$，故 $\omega_{k+1}(\|E\|_\infty) \to 0$。由有限次归纳法，当递推至 $k = N$ 时即证得（59）。 ∎

第五节 GLL 全局化方法的 R-线性收敛

命题 6.8 在假设 6.3 下，存在整数 $N \geqslant 1$ 与常数 $\theta \in (0,1)$，使得对所有 $E \in \mathcal{U}_{\mathrm{reg}}$ 都有

$$\left\|\mathcal{T}_q^N(E)\right\|_\infty \leqslant \theta\|E\|_\infty . \tag{61}$$

证明 由假设 6.3(i)，在 $\mathcal{U}_{\mathrm{reg}} \setminus \{0\}$ 上二次步长 $E \mapsto \alpha_i^q(E)$ 由 3.1 给出，故其关于 E 连续、零次齐次并且一致有界。于是 $\mathcal{T}_q$ 在 $\mathcal{U}_{\mathrm{reg}}$ 上连续，且一次正齐次。再由假设 6.3(ii)，若 $z \in K$ 且 $\mathcal{T}_q(z) \neq 0$，则归一化方向 $\mathcal{T}_q(z)/\left\|\mathcal{T}_q(z)\right\|_\infty$ 仍属于 K。结合假设 6.3(iii)，命题 2.5 直接给出常数 $C_q > 0$、$\rho_q \in (0,1)$，使

$$\left\|\mathcal{T}_q^k(E)\right\|_\infty \leqslant C_q\rho_q^k\|E\|_\infty , \qquad \forall E \in \mathcal{U}_{\mathrm{reg}},\ \forall k \geqslant 0.$$

取 N 使 $C_q\rho_q^N < 1$，并令 $\theta := C_q\rho_q^N$，便得到（61）。 ∎

定理 6.9 设假设 6.2、6.3 成立。则存在 $\mathcal{U} \subset \mathcal{U}_{\mathrm{reg}}$、整数 $N \geqslant 1$ 与常数 $\hat{\theta} \in (0,1)$，使得对所有 $E \in \mathcal{U}$ 都有

$$\left\|\mathcal{T}_f^N(E)\right\|_\infty \leqslant \hat{\theta}\,\|E\|_\infty . \tag{62}$$

证明 由命题 6.8，存在整数 $N \geqslant 1$ 与常数 $\theta \in (0,1)$ 使

$$\left\|\mathcal{T}_q^N(E)\right\|_\infty \leqslant \theta\,\|E\|_\infty , \qquad \forall E \in \mathcal{U}_{\mathrm{reg}}.$$

取任意 $\hat{\theta} \in (\theta,1)$。由推论 6.7，存在 $\rho_0 \in (0,\delta_{\mathrm{reg}})$ 使得当 $E \in \mathcal{U}_{\mathrm{reg}}$ 且 $\|E\|_\infty \leqslant \rho_0$ 时，

$$\left\|\mathcal{T}_f^N(E) - \mathcal{T}_q^N(E)\right\|_\infty \leqslant (\hat{\theta}-\theta)\,\|E\|_\infty .$$

于是对所有满足 $E \in \mathcal{U}_{\mathrm{reg}}$ 且 $\|E\|_\infty \leqslant \rho_0$ 的状态都有

$$\left\|\mathcal{T}_f^N(E)\right\|_\infty \leqslant \left\|\mathcal{T}_q^N(E)\right\|_\infty + \left\|\mathcal{T}_f^N(E) - \mathcal{T}_q^N(E)\right\|_\infty \leqslant \hat{\theta}\,\|E\|_\infty .$$

令

$$\mathcal{U} := \{E \in \mathcal{U}_{\mathrm{reg}} :\ \|E\|_\infty < \rho_0\},$$

即得（62）。 ∎

不难看出，只要 *p*D(S)-BB 的尾轨道落在假设 6.3 所定义的 $\mathcal{U}_{\mathrm{reg}}$，上述定理便可使用。记此时块状态序列为

$$E_{\ell+1}^{\mathrm{fb}} = \mathcal{T}_f(E_\ell^{\mathrm{fb}}), \qquad \ell \geqslant 0.$$

由上一定理可进一步得到其局部 *R*-线性收敛结论。

定理 6.10 在假设 6.2、6.3 下，存在常数 $\delta > 0$、$N \geqslant 1$、$\hat{\theta} \in (0,1)$、$C_b > 0$、$\rho_b \in (0,1)$ 以及 $C > 0$、$\rho \in (0,1)$，使得只要初始块状态满足

$$E_0^{\mathrm{fb}} \in \mathcal{U}_\delta := \{E \in \mathcal{U}_{\mathrm{reg}} :\ \|E\|_\infty \leqslant \delta\},$$

则固定块长 *p*D(S)-BB 模型尾轨道 $\{E_\ell^{\mathrm{fb}}\}$ 满足：

(i) 对任意 $\ell \geqslant 0$ 都有统一的 N 步收缩

$$\left\|E_{\ell+N}^{\mathrm{fb}}\right\|_\infty \leqslant \hat{\theta}\left\|E_\ell^{\mathrm{fb}}\right\|_\infty ; \tag{63}$$

(ii) 在块指标意义下，状态序列 *R*-线性收敛：

$$\left\|E_\ell^{\mathrm{fb}}\right\|_\infty \leqslant C_b\rho_b^\ell \left\|E_0^{\mathrm{fb}}\right\|_\infty , \qquad \forall \ell \geqslant 0; \tag{64}$$

(iii) 在原始迭代指标意义下，误差序列也局部 *R*-线性收敛

$$\left\|x_k^{\mathrm{fb}}\right\| \leqslant C\rho^k \left\|E_0^{\mathrm{fb}}\right\|_\infty , \qquad \forall k \geqslant 0. \tag{65}$$

证明 由定理 6.9，存在整数 $N \geqslant 1$、常数 $\hat{\theta} \in (0,1)$ 与半径 $\rho_0 \in (0, \delta_{\mathrm{reg}})$，使得（62）对所有 $E \in \mathcal{U}_{\mathrm{reg}}$ 且 $\|E\|_\infty \leqslant \rho_0$ 成立。

另一方面，由假设 6.3(i) 中的步长一致上界与（52），存在常数

$$B := (1 + \overline{\alpha} L)^p$$

使得

$$\left\|\mathcal{T}_f(E)\right\|_\infty \leqslant B \|E\|_\infty, \qquad \forall E \in \mathcal{U}_{\mathrm{reg}}.$$

取

$$\delta := \min\left\{ \frac{\rho_0}{B^{N-1}}, \frac{\delta_{\mathrm{reg}}}{B^{N-1}} \right\}.$$

若 $E_0^{\mathrm{fb}} \in \mathcal{U}_\delta$，则由假设 6.3(ii) 与上式，对任意 $r = 0, \cdots, N-1$ 都有

$$\mathcal{T}_f^r(E_0^{\mathrm{fb}}) \in \mathcal{U}_{\mathrm{reg}}, \qquad \left\|\mathcal{T}_f^r(E_0^{\mathrm{fb}})\right\|_\infty \leqslant B^r \delta \leqslant \rho_0.$$

于是（62）可应用于 E_0^{fb} 以及其前 $N-1$ 个后继状态。对任意 $\ell \geqslant 0$ 作归纳：若 $\left\|E_\ell^{\mathrm{fb}}\right\|_\infty \leqslant \delta$，则同样有

$$\left\|E_{\ell+r}^{\mathrm{fb}}\right\|_\infty \leqslant B^r \delta \leqslant \rho_0, \qquad r = 0, \cdots, N-1,$$

并由（62）得

$$\left\|E_{\ell+N}^{\mathrm{fb}}\right\|_\infty \leqslant \hat{\theta} \left\|E_\ell^{\mathrm{fb}}\right\|_\infty \leqslant \hat{\theta}\delta < \delta.$$

这证明了（63）对所有 $\ell \geqslant 0$ 均成立。

下面证明（64）。对任意 $\ell = mN + r$（其中 $m \geqslant 0$、$0 \leqslant r \leqslant N-1$），反复使用（63）得

$$\left\|E_{mN}^{\mathrm{fb}}\right\|_\infty \leqslant \hat{\theta}^m \left\|E_0^{\mathrm{fb}}\right\|_\infty,$$

而块间统一有界性给出

$$\left\|E_\ell^{\mathrm{fb}}\right\|_\infty = \left\|\mathcal{T}_f^r(E_{mN}^{\mathrm{fb}})\right\|_\infty \leqslant B^r \left\|E_{mN}^{\mathrm{fb}}\right\|_\infty \leqslant B^r \hat{\theta}^m \left\|E_0^{\mathrm{fb}}\right\|_\infty.$$

令

$$K := \max\{1, B, \cdots, B^{N-1}\}, \qquad \rho_b := \hat{\theta}^{1/N} \in (0,1), \qquad C_b := K/\hat{\theta},$$

并由 $m \geqslant \ell/N - 1$，即可得到

$$\left\|E_\ell^{\mathrm{fb}}\right\|_\infty \leqslant C_b \rho_b^\ell \left\|E_0^{\mathrm{fb}}\right\|_\infty, \qquad \forall \ell \geqslant 0,$$

即为（64）。

最后把块指标估计转回原始迭代指标。任意 $k \geqslant 0$ 可写成 $k = \ell p + j$，其中 $\ell \geqslant 0$、$0 \leqslant j \leqslant p-1$。该步对应的误差是从块状态 E_ℓ^{fb} 出发经过 j 次块内纯迭代得到的某个中间向量，因此

$$\left\|x_k^{\mathrm{fb}}\right\| \leqslant B \left\|E_\ell^{\mathrm{fb}}\right\|_\infty.$$

再结合（64）与 $\ell \geqslant k/p - 1$，得到

$$\left\|x_k^{\mathrm{fb}}\right\| \leqslant BC_b\rho_b^{\ell}\left\|E_0^{\mathrm{fb}}\right\|_\infty \leqslant \frac{BC_b}{\rho_b}\left(\rho_b^{1/p}\right)^k\left\|E_0^{\mathrm{fb}}\right\|_\infty.$$

令

$$\rho := \rho_b^{1/p} \in (0,1), \qquad C := BC_b/\rho_b,$$

便得到（65）。证毕。 ∎

注 在假设 6.3(i) 下，若某个块状态 $E \in \mathcal{U}_{\mathrm{reg}}$，则从该状态出发执行一个完整块时，纯实现中的秩压缩、负特征值替代与步长截断都不会触发。因此，该完整块产生的下一块状态恰好等于固定块长模型给出的 $\mathcal{T}_f(E)$。特别地，若带压缩纯实现的一条尾轨道满足

$$E_\ell^{\mathrm{pure}} \in \mathcal{U}_{\mathrm{reg}}, \qquad \forall \ell \geqslant \ell_0,$$

且令固定块长尾轨道以相同块状态起步，即 $E_{\ell_0}^{\mathrm{fb}} = E_{\ell_0}^{\mathrm{pure}}$，则

$$E_\ell^{\mathrm{fb}} = E_\ell^{\mathrm{pure}}, \qquad \forall \ell \geqslant \ell_0.$$

现在，我们可以得到本文最关心的问题。

定理 6.11 (GLL 最终全接受与局部 R-线性收敛) 设假设 6.1、6.2、5.1、6.3 成立，并进一步假设 f 强凸。记 $\{E_\ell^{\mathrm{gll}}\}$ 为算法 4 产生的块状态序列，则对采用 GLL 检测全局化的对称化 pD(S)-BB 算法有：

(i) 全序列全局收敛到唯一极小点 x^*；

(ii) 若进一步存在块指标 ℓ_{reg} 使

$$E_\ell^{\mathrm{gll}} \in \mathcal{U}_{\mathrm{reg}}, \qquad \forall \ell \geqslant \ell_{\mathrm{reg}},$$

则存在 $\ell_0 \geqslant \ell_{\mathrm{reg}}$ 与常数 $C > 0$、$\rho \in (0,1)$，使得以 $E_{\ell_0}^{\mathrm{gll}}$ 为初始块状态所诱导的固定块长尾轨道 $\{E_\ell^{\mathrm{fb}}\}_{\ell \geqslant \ell_0}$ 以及与之逐块一致的带压缩纯实现尾轨道 $\{E_\ell^{\mathrm{pure}}\}_{\ell \geqslant \ell_0}$ 都满足局部 R-线性收敛；

(iii) 在上述尾轨道上，搜索方向、试探步长区间以及梯度几何衰减满足 Dai[8] 中 Theorem 3.1 的假设。因此，沿该局部尾轨道，GLL 非单调线搜索从某个有限时刻起持续接受全试探步长；

(iv) 因而，全局化实现从某个有限时刻起与固定块长尾轨道一致，并且 R-线性收敛。

证明 由强凸性与命题 6.1，第 (i) 点成立。

现在额外假设存在 ℓ_{reg} 使 $E_\ell^{\mathrm{gll}} \in \mathcal{U}_{\mathrm{reg}}$ 对所有 $\ell \geqslant \ell_{\mathrm{reg}}$ 成立。由第 (i) 点知，全局化算法产生的迭代点满足 $x_k \to x^*$，从而相应块状态满足 $\left\|E_\ell^{\mathrm{gll}}\right\|_\infty \to 0$。取定理 6.10 中给出的半径 $\delta > 0$，则存在某个 $\ell_0 \geqslant \ell_{\mathrm{reg}}$ 使

$$E_{\ell_0}^{\mathrm{gll}} \in \mathcal{U}_\delta.$$

以该块状态作为初始块状态启动固定块长尾轨道，记所得轨道为 $\{E_\ell^{\mathrm{fb}}\}_{\ell\geqslant\ell_0}$。由定理 6.10，该尾轨道满足局部 R-线性收敛，并且由于 $\{E_\ell^{\mathrm{pure}}\}_{\ell\geqslant\ell_0}$ 与 $\{E_\ell^{\mathrm{fb}}\}_{\ell\geqslant\ell_0}$ 逐块一致，因此前者也满足同样的局部 R-线性收敛。这证明了第 (ii) 点。

再由（52），误差的几何衰减与梯度范数的几何衰减等价，因此上述带压缩纯实现的尾轨道也满足

$$\|g_k\| \leqslant C_g \rho^k$$

形式的局部 R-线性估计。另一方面，该尾轨道上的搜索方向始终为

$$d_k = -g_k,$$

试探步长由算法截断在固定区间

$$\alpha_k \in [\underline{\alpha}, \overline{\alpha}]$$

内，并且

$$g_k^T d_k = -\|g_k\|^2, \qquad \|d_k\|^2 = \|g_k\|^2.$$

因此，沿该局部尾轨道，Dai[8] 中 Theorem 3.1 所要求的梯度方向条件、步长有界条件以及有界间隔几何收缩条件均得到满足。由该定理的最终接受结论可知，从某个有限时刻起，GLL 非单调线搜索会持续接受全试探步长。这证明了第 (iii) 点。

一旦后续所有试探步长都被完全接受，全局化实现便从某个有限时刻起与纯实现的尾轨道逐块一致，再结合第 (ii) 点即知全局化实现从该时刻起与固定块长尾轨道一致，并继承同样的局部 R-线性收敛性质。这证明了第 (iv) 点。 ∎

定理 6.11 表明：对 GLL 全局化算法而言，一旦其尾轨道进入并停留在假设 6.3 所描述的正则区域内，则相应的固定块长尾轨道、带压缩纯实现尾轨道以及最终全接受后的 GLL 全局化尾轨道完全一致，有同一 R-线性收敛率。

第七章　数值实验

本章针对前述理论结果给出数值实验。第一部分（§一–§八）针对二次模型的精确秩压缩 LMSD 与 ω 加权 Ritz-sweep 族，包括验证有限终止、R-线性包络、处理大规模问题的性能以及与既有方法的比较等内容。第二部分（§九–§十五）将算法实现推广到一般非线性问题，考察第六章 GLL 全局化版本（算法 4）在传统强凸光滑测试问题上的实际行为以及局部 R-线性收敛的经验拟合。

第一部分：二次模型实验

本部分针对前述二次模型理论结果给出数值实验：（一）$p \geqslant n$ 时的有限终止；（二）与 BB1、BB2 以及共轭梯度法（CG）的比较；（三）历史束最小奇异值的数值衰减；（四）ω 加权 Ritz-sweep 在不同谱分布下的表现；（五）n 增大时的算法表现；（六）R-线性包络的经验拟合；（七）与 ABBmin、cyclic-BB 等成熟自适应 BB 方法的对比。所有实验均在严格凸二次模型上执行，与本文讨论的理论范围一致。

第一节　实验设置

取 $H = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，$0 < \lambda_1 < \cdots < \lambda_n$；除特别说明外，谱在 $[1, \lambda_n]$ 上对数等距分布。初始梯度由确定性公式

$$g_{0,0}^{(i)} = \sin(0.37(i+\sigma)) + 0.3\cos(1.17(i+\sigma)) + 0.5\sin(0.11(i+\sigma)) + 1 \qquad (66)$$

给出，其中 $\sigma \in \mathbb{Z}$ 为偏移量。除非另有说明，停止准则取 $\|g_k\| \leqslant 10^{-8}\|g_0\|$；有限终止实验取更严的 $10^{-13}\|g_0\|$。ω-Ritz-sweep 类算法均以常数预热步长 $\alpha_{\mathrm{warm}} = 1/\lambda_n$ 进行前 p 步；BB1、BB2 以 Cauchy 步起步；CG 从 $x_0 = 0$ 求解 $Hx = g_{0,0}$ 起步。

第二节　有限终止现象

本节通过低维可视化与高维计数两类例子直接验证定理 4.5。

（1）二维与三维可视化。

考虑二维严格凸二次函数

$$f(x) = \frac{1}{2}(10x_1^2 + x_2^2), \qquad x_0 = (8, 8)^T.$$

图 7.1 给出等高线与轨迹，并以固定步长梯度法（$\alpha = 0.15$）作为对照。根据定理 4.5，2D 算法应在 $p + n = 2 + 2 = 4$ 步内终止，图中清楚地体现这一点。同时，

由推论 2.8 的 $p = 1$ 闭锥实例化与第四章 §四 的等价性论证，本文二维改进对称化公式在严格凸二次情形下与原始 2D 公式产生相同的步长，因此该图同时验证了改进公式在二次模型上的正确性。

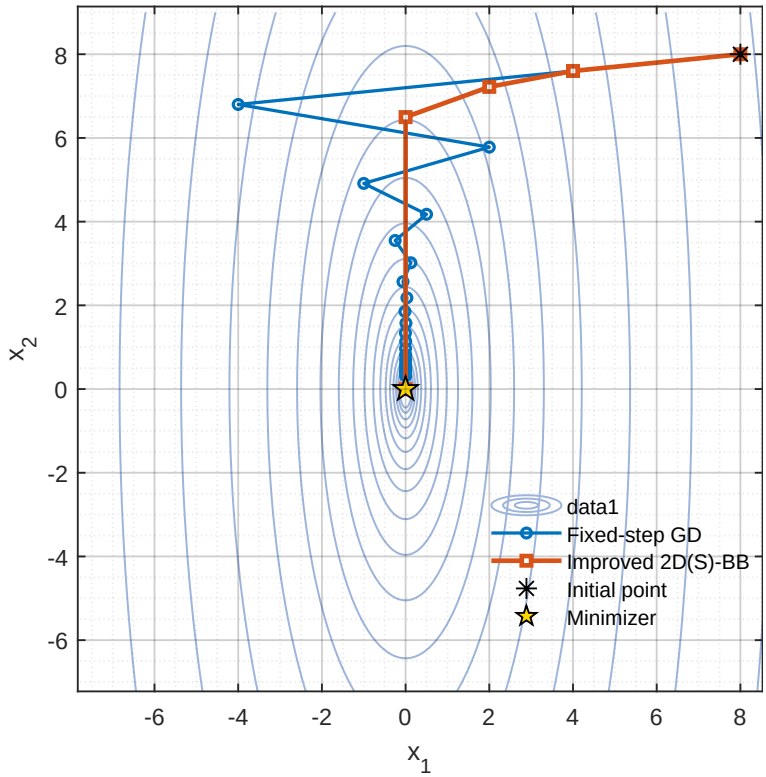


图 7.1　二维严格凸二次函数上的有限终止现象。2D(S)-BB 在数步内到达极小点，固定步长梯度法仅呈现缓慢锯齿式下降。

类似地，考虑三维严格凸二次函数

$$f(x) = \frac{1}{2}(10x_1^2 + 5x_2^2 + x_3^2), \qquad x_0 = (0.2, 0.3, 0.5)^T.$$

图 7.2 左图给出三维轨迹，右图给出目标函数值随迭代次数的下降曲线。3D 算法在数步内到达极小点，2D 算法也显著快于固定步长梯度法。

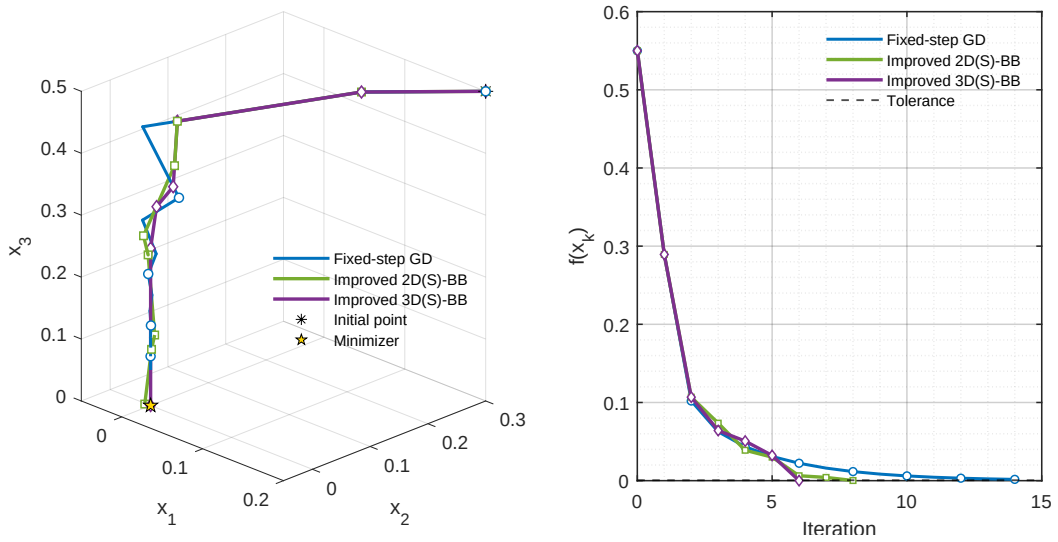


图 7.2　三维严格凸二次函数上的轨迹与历史函数值。3D(S)-BB 在数步内有限终止，2D(S)-BB 明显优于固定步长梯度法。

（2）高维有限终止上界的精确取到。

取 $n = 8$，谱 $(1, 1.7, 2.9, 5.1, 9, 16, 28, 50)$，$\sigma = 3$。表 7.1 给出秩压缩 ω-Ritz-sweep 方法在 $\omega(t) = t^a$、$a \in \{-1, 0, 1\}$ 下，$p \in \{8, 10, 12\}$ 时的梯度评估数。定理 4.5 给出的上界 $p + n$ 在该实现中精确取到，并且与权重 ω 的具体形式无关。

p	$\omega(t) = t^a$ 中 a	迭代次数	理论上界 $p + n$
8	-1, 0, 1	16	16
10	-1, 0, 1	18	18
12	-1, 0, 1	20	20

表 7.1　有限终止实验中三个权重给出相同的迭代次数且均与理论上界相同，因为预热后第一个完整块的 sweep 已经包含全部活跃特征值的倒数步长。

第三节　与 BB1、BB2、共轭梯度法的比较

在两个严格凸二次问题上比较：（A）对角对数谱 $n = 30$，$p = 4$，特征值在 $[1, 500]$，$\sigma = 0$；（B）$n = 20$ 的稠密对称正定问题 $H_d = Q\,\mathrm{diag}(\lambda_1, \cdots, \lambda_n)Q^T$，$\lambda_i$ 在 $[1, 100]$ 等距，Q 由 $n - 1$ 个角度为 $0.41i$ 的平面旋转复合得到。

方法	对角 n=30, κ=500	稠密 n=20, κ=100
CG（Krylov 最优基线）	45	21
BB1	269	119
BB2	272	111
LMSD（$\omega \equiv 1, p$=4）	187	88
harm-LMSD（$\omega(t)$=t, p=4）	222	85
逆权（$\omega(t)$=1/t, p=4）	183	85
$\omega(t)$=t^2, p=4	221	85
LMSD（$\omega \equiv 1, p$=10）	148	—
harm-LMSD（p=10）	150	—

表 7.2　达到 $\|g_k\| \leqslant 10^{-8}\|g_0\|$ 的梯度评估数。$p = 4$ 时 LMSD 类比 BB1/BB2 快约 30%，$p = 10$ 再降约 20%。$p = 10$ 在稠密 $n = 20$ 实验中被省略，因为 p 接近 n 已几乎进入有限终止区间。

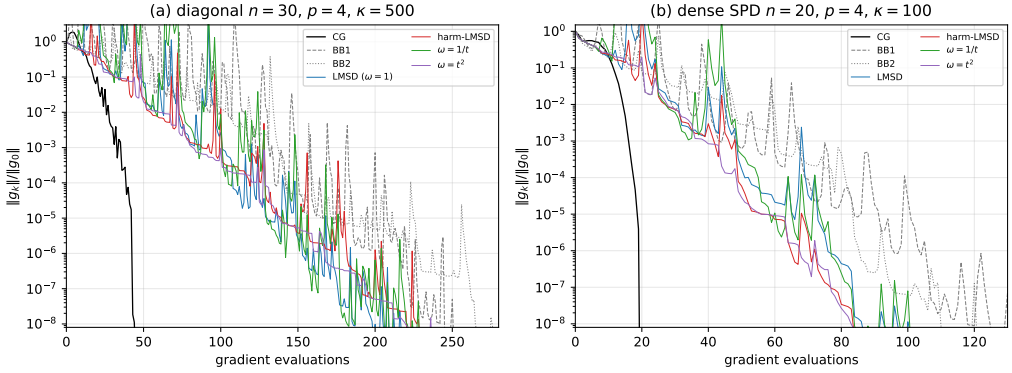


图 7.3　表 7.2 对应的梯度范数下降曲线。ω-Ritz-sweep 族（包括 LMSD、harm-LMSD 与 $\omega(t) = t^a$ 的若干 a）显著快于 BB1/BB2。

第四节　历史束最小奇异值的数值衰减

在 $n = 10$，$p = 6$，$\kappa = 100$，$\sigma = 0$ 的对角对数谱问题上监测 $\mu_k := \sigma_{\min}(G_k)/\|g_{k,0}\|$。表 7.3 显示 $\sigma_{\min}(G_k)$ 在 10 个外块内跨 9 个数量级衰减，与之相应的，$|\mathcal{A}(w_k)|$（按 10^{-12} 阈值计的活跃支撑数）从 9 降到 6。

块 k	μ_k	$\sigma_{\min}(G_k)$	$\|g_{k,0}\|$	$\|\mathcal{A}(w_k)\|$	$1/\mu_k$
1	9.05×10^{-5}	3.22×10^{-4}	3.56	9	1.10×10^{4}
3	1.50×10^{-4}	3.69×10^{-5}	0.246	9	6.66×10^{3}
5	4.25×10^{-3}	3.65×10^{-4}	0.086	9	2.35×10^{2}
8	1.91×10^{-4}	7.96×10^{-10}	4.17×10^{-6}	9	5.24×10^{3}
10	2.01×10^{-9}	1.70×10^{-12}	8.46×10^{-4}	6	4.97×10^{8}

表 7.3　LMSD（$\omega \equiv 1, p = 6$）轨道上的历史束最小奇异值。该轨道上没有有限的统一 ρ 能使 $\|R_k^{-1}\| \leqslant \rho\|g_{k,1}\|^{-1}$ 长期成立，相当于 Curtis–Guo[13] 历史束非退化假设在该轨道上失效。

第五节　不同谱下的 ω-加权表现

定理 4.22 把 ω-加权 Ritz-sweep 族作为统一的分析对象。本节通过若干典型谱说明，在不同条件数与不同聚类结构的谱下，没有任何固定 ω 一致占优，算法族里的不同算法步长表现大致类似。

考虑三个聚类谱：

- cluster-6：$n = 30$，6 个 5-簇，簇心在 $[1, 100]$ 上对数等距，簇宽 5%；
- low-cluster：$n = 20$，10 个特征值在 $[1, 1.5]$ 等距，10 个在 $[10, 100]$ 对数等距；
- two-cluster：$n = 20$，10 个在 $[1, 2]$，10 个在 $[50, 60]$，均对数等距。

在每个谱下取 $p \in \{4, 6, 8\}$，对 $\omega(t) = t^a$，$a \in \{-1, 0, 1/2, 1, 2\}$，记录达到 $\|g_k\| \leqslant 10^{-8}\|g_0\|$ 的迭代数（表 7.4）。各权重之间的极差从 3% 到 28% 不等：$\omega(t) = t^{1/2}$ 在 cluster-6 ($p = 4, 6$) 上最快，但在 two-cluster ($p = 8$) 上反而慢于 $\omega(t) = t^2$，没有任何固定权重一致最优。这也是为何本文把 ω 族作为统一分析工具、而并未对其具体性能优劣进行分析与解释。

谱	p	t^{-1}	1	$t^{1/2}$	t	t^2	极差
cluster-6	4	86	78	67	81	74	28%
cluster-6	6	76	76	60	75	69	27%
cluster-6	8	66	67	66	65	65	3%
low-cluster	4	62	76	71	65	73	23%
low-cluster	6	68	69	63	69	63	10%
low-cluster	8	69	65	65	72	73	12%
two-cluster	4	39	39	37	37	37	5%
two-cluster	6	38	36	36	35	35	9%
two-cluster	8	47	46	45	45	44	7%

表 7.4　不同聚类谱下 ω-加权 Ritz-sweep 方法的迭代数。极差从 3% 到 28% 不等，没有任何权重一致最优。

图 7.4 在另一组对数等距谱（$n = 30$、$\kappa = 500$）上给出 $a \in \{-2, -1, -1/2, 0, 1/2, 1, 2, 3\}$ 共 8 个权重对应的梯度范数下降曲线。各曲线总体相近，没有任何固定 a 一致占优，与表 7.4 的结论一致。

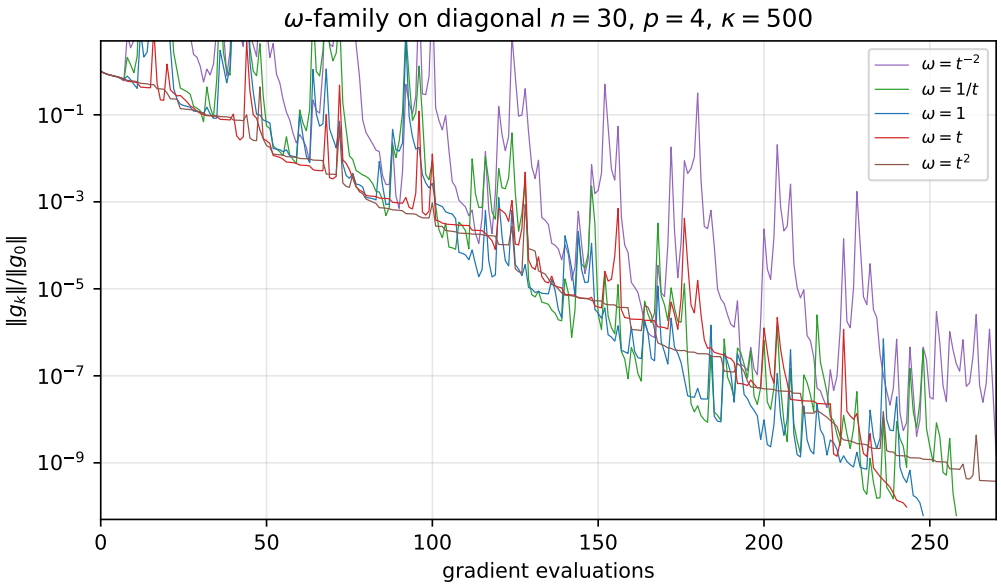


图 7.4　在 $n = 30$、$\kappa = 500$ 的对角对数谱上，$\omega(t) = t^a$ 的 8 个 a 值对应的梯度范数下降曲线，各曲线总体相近。

第六节　处理大规模问题的性能

在对角对数谱 $\kappa = 1000$，$n \in \{100, 1000, 10000\}$ 上运行典型方法，下表为相应方法所需的迭代数与时间。

方法	$n = 100$		$n = 1000$		$n = 10000$	
	迭代	秒	迭代	秒	迭代	秒
CG	168	0.002	294	0.001	308	0.016
BB1	438	0.001	371	0.002	511	0.026
BB2	404	0.001	419	0.001	403	0.021
LMSD (p=4)	417	0.007	487	0.031	444	0.312
LMSD (p=8)	355	0.005	369	0.025	360	0.277
harm. (p=4)	393	0.005	366	0.018	433	0.289
harm. (p=8)	330	0.006	379	0.030	393	0.303

表 7.5　达到 $\|g_k\| \leqslant 10^{-8}\|g_0\|$ 的迭代数与时间。

第七节　经验 R-线性包络

定理 4.18–4.19 与 4.22 说明每条非终止正则轨道上 $\|g_k\|$ 以某个速率 $\rho \in (0,1)$ R-线性收敛。本节给出三种角度的经验验证：（一）单条轨道上的拟合速率；（二）随机初始条件下的统计稳定性；（三）不同问题维数 n、条件数 κ、权重上 ω 的整体分布。

（1）单条轨道上的拟合速率。

对每条轨道，对 $\log(\|g_k\|/\|g_0\|)$ 在后半段做最小二乘拟合，得到经验逐步速率 $\hat{\rho}$（表 7.6）。所有实例上拟合速率都远低于 $1-1/\kappa$，可以猜测其为一个速率上界。

配置	$\hat{\rho}$ (ω=1/t)	$\hat{\rho}$ ($\omega\equiv1$)	$\hat{\rho}$ (ω=t)	$1-1/\kappa$
n=30, κ=100, p=4	0.794	0.817	0.838	0.990
n=30, κ=100, p=8	0.898	0.857	0.823	0.990
n=30, κ=500, p=4	0.901	0.902	0.911	0.998
n=30, κ=1000, p=4	0.931	0.941	0.937	0.999
n=40, κ=1000, p=4	0.937	0.956	0.928	0.999

表 7.6　每步拟合速率 $\hat{\rho}$ 及其与 $1-1/\kappa$ 的比较。

（2）随机初始条件下的稳定性。

确定性初始梯度（66）是单一代表性选择。改用 10 个独立 $\mathcal{N}(0, I_n)$ 随机初始向量（种子 $0,\cdots,9$）后，迭代数与 $\hat{\rho}$ 的均值与标准差列于表 7.7：迭代数相对标准差在 4%–12%，$\hat{\rho}$ 相对标准差在 1%–4%。表 7.6 中确定性初始的对应数据都落在随机均值的 1–2 倍标准差内。

配置	ω	迭代均值	迭代标准差	$\hat{\rho}$ 均值	$\hat{\rho}$ 标准差
n=30, κ=100, p=4	1/t	98.9	5.7	0.8176	0.0321
	$\equiv 1$	96.6	5.6	0.8213	0.0316
	t	94.5	3.7	0.8079	0.0152
n=30, κ=500, p=4	1/t	221.5	26.8	0.9101	0.0169
	$\equiv 1$	211.1	19.7	0.9129	0.0122
	t	211.1	19.3	0.9128	0.0132
n=40, κ=1000, p=4	1/t	318.6	32.7	0.9338	0.0118
	$\equiv 1$	319.7	28.2	0.9406	0.0072
	t	313.4	14.8	0.9414	0.0061

表 7.7　10 个高斯随机初始梯度（种子 $0,\cdots,9$）下的迭代数与拟合速率 $\hat{\rho}$。

（3）不同问题维数 n、条件数 κ、权重上 ω 的整体分布。

为说明 R-线性行为对配置选择的稳健性，在 $n\in\{20,30,40\}$、$\kappa\in\{100,1000\}$、$p\in\{4,8\}$、$\sigma\in\{0,7,13\}$ 共 36 个配置以及 $\omega\in\{1/t,1,t\}$ 共 108 次拟合上测得 $\hat{\rho}$ 中位数 ≈ 0.87，均值 ≈ 0.87，所有 $\hat{\rho}$ 都低于最坏情形 $1-1/\kappa$，绝大多数低于 0.96（图 7.6）。三个权重的分布相近，$\omega(t)=t$ 的速率略小（更快）。图 7.5 给出两组典型配置上的轨迹和包络的可视化：原始轨迹在半对数坐标下显示明显的非单调震荡，但其上包络近似为直线。

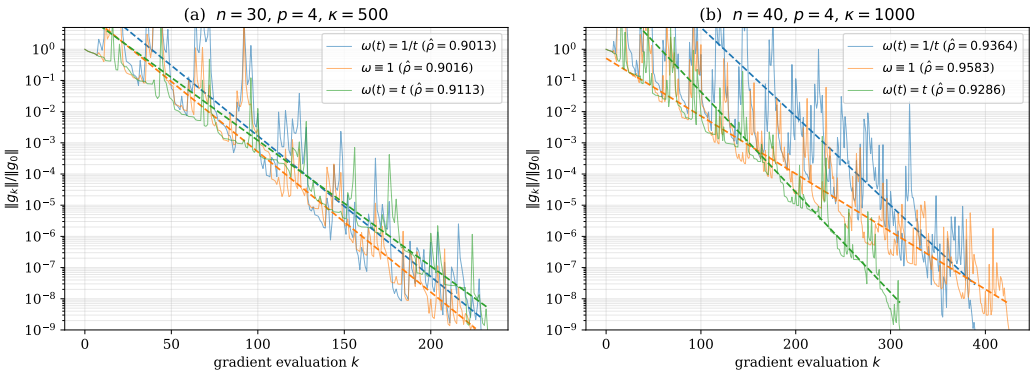


图 7.5　在两组 n、κ 的对角对数谱上，LMSD（$\omega \equiv 1$）与 harm-LMSD（$\omega(t) = t$）轨迹在半对数坐标下显示明显的非单调震荡，但其上包络近似为直线。该现象与第四章 §三 的闭锥 R-线性收敛结论一致。

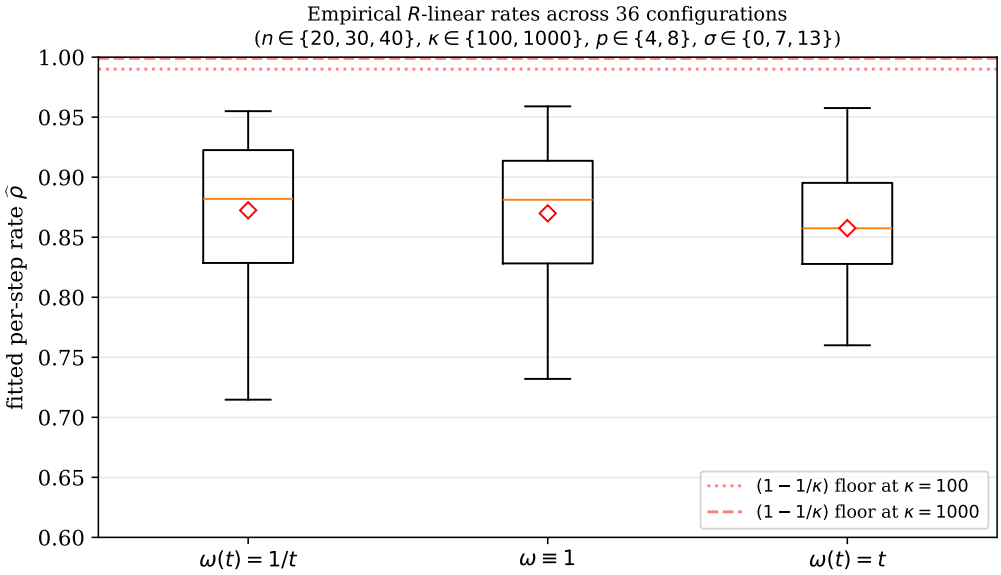


图 7.6　36 组配置（$\kappa \in \{10^2, 10^3\}$、$p \in \{4, 6, 8\}$、$\sigma \in \{0, \cdots, 5\}$）下三个权重 $\omega \in \{1, t, t^{-1}\}$ 拟合得到的经验速率 $\hat{\rho}$ 的箱线图，共 108 个数据点。$\hat{\rho}$ 全部落在 $[0.78, 0.96]$，远小于推论 4.23 中最坏情形 $1 - 1/\kappa \in \{0.99, 0.999\}$。这说明本文给出的显式收缩界确实保守，全范数的紧的实际速率仍是开放问题。

第八节　与自适应 BB 变体的对比

最后与两个成熟自适应 BB 方法对比：ABBmin[18]，自适应在 BB1 与 BB2 之间根据历史最小值切换；以及 cyclic-BB[10]，以更新 m 步固定 BB1 步长循环。每行为 10 个随机高斯初始梯度（种子 $0, \cdots, 9$）的均值 ± 标准差。

谱	BB1	BB2	ABBmin	cyclic-BB	LMSD p=4	LMSD p=8	harm. p=4
diag n=30, κ=100	123.1±10.4	120.1±8.5	108.4±4.7	124.6±9.4	96.6±5.6	92.2±0.9	94.5±3.7
diag n=30, κ=1000	397.0±41.1	419.1±74.2	320.4±17.4	389.5±49.3	273.3±14.3	202.4±12.7	294.2±24.7
diag n=100, κ=1000	423.7±35.6	431.7±80.6	345.9±14.7	406.7±24.4	379.3±31.4	323.4±8.9	369.3±27.6
two-cluster n=20	52.6±3.0	50.2±2.5	38.7±3.2	47.3±7.4	40.8±3.9	44.4±0.5	36.8±0.4
low-cluster n=20	110.5±13.7	118.5±14.6	89.8±8.1	99.9±10.0	65.5±3.7	61.4±3.2	66.6±3.3

表 7.8　达到 $\|g_k\| \leqslant 10^{-8}\|g_0\|$ 的迭代数。LMSD 类（$p \in \{4, 8\}$）在所有 5 个测试实例上均胜出（其中 LMSD $p = 8$ 在 4 行最优，harm. $p = 4$ 在 two-cluster 上最优）。ABBmin 是除 LMSD 类之外表现最好的算法，特别是在对数谱条件数大的情形。

值得注意的是，LMSD $p = 8$ 在表 7.8 上的标准差很小（$0.9, 12.7, 8.9, 0.5, 3.2$），这表明更长的记忆能抑制初始条件带来的轨道间变化。

第二部分：非二次模型实验

第一部分在严格凸二次模型上验证了行列式与 Cauchy–Binet 表示等给出的收敛性质以及闭锥原理给出的若干渐近行为。本部分把视野推广到一般非线性问题，重点考察第六章给出的对称化 pD(S)-BB 算法（算法 3 与其 GLL 全局化版本算法 4）。实验说明了三件事情：（一）GLL 全局化版本相对纯实现具有更稳健的收敛行为；（二）对称化 pD(S)-BB 在传统非线性测试问题上与 BB-Raydan、Polak–Ribière CG、L-BFGS 等基线方法是有竞争力的；（三）局部进入 $\mathcal{U}_{\mathrm{reg}}$ 之后，GLL 全局化轨道呈现与定理 6.11 预期一致的局部 R-线性渐近行为。

第九节　非二次实验设置与对比基线

记号沿用第七章。停止准则取 $\|\nabla f(x_k)\| \leqslant 10^{-6}\|\nabla f(x_0)\|$，最大迭代数 $N_{\max} = 10000$。GLL 参数取 $\beta = 0.5$、$\sigma = 10^{-4}$、$M = 10$、$\underline{\alpha} = 10^{-15}$、$\overline{\alpha} = 10^{15}$，与[5] 一致。$p$D(S)-BB 取 $p = 5$（除非另有说明），预热使用恒定步长 $\alpha_0 = 10^{-3}$。

对比的非二次算法包括：

- BB-Raydan[5]：BB1 步长 + GLL 非单调线搜索；
- CG-PR：Polak–Ribière 共轭梯度法 + Wolfe 线搜索（$\sigma = 0.1$）；
- L-BFGS-m：Nocedal[23] 的 L-BFGS，$m \in \{3, 5\}$；
- pD(S)-BB-pure：算法 3 的纯实现版本，无 GLL 检测。

第十节　精确 Hessian 公式与改进对称化公式的比较

本节考察一个明确非二次但保持强凸光滑的测试函数族

$$f(x)=\frac{1}{2}x^{T}Ax+\beta\sum_{i=1}^{n}\log\cosh\big((Ux)_i\big),\tag{67}$$

其中 $A=\mathrm{diag}(10^{6},10^{6-6/99},\cdots,1)$（$n=100$）、$U$ 为固定随机正交矩阵（随机种子 42）、$\beta=200$、初始点 $x_0=10\,\mathrm{randn}(n,1)$。该函数 Hessian

$$\nabla^2 f(x)=A+\beta U^{T}\ \mathrm{diag}\big(1-\tanh((Ux)_i)^2\big)\,U$$

在 $\beta>0$ 与 $A\succ 0$ 下保持 $\beta\cdot\lambda_{\min}(UU^{T})+\lambda_1(A)$ 的下界，因此整体强凸光滑。随机正交 U 引入真实的变量耦合，$\log\cosh$ 项提供可控的非二次扰动。具体实现中采用数值稳定的等价公式 $\log\cosh(z)=|z|-\log 2+\log(1+\mathrm{e}^{-2|z|})$，避免高条件数下的数值上溢。

在该问题上比较两种 pD(S)-BB 步长生成方式：

(i) **基于精确 Hessian 投影的公式**（基线）：在每个完整块上将 $H_{*}=\nabla^2 f(x_k)$（精确 Hessian）投影到最近 p 个位移张成的子空间上，取其特征值的倒数作为步长。

(ii) **改进对称化公式**：使用本文 §六 中算法 3 的对称化广义特征值问题 $C_k z=\theta M_k z$，无需精确 Hessian。

两者在严格凸二次问题上等价（见第四章 §四 末段），但在非二次问题上会出现差异。两种实现共用同一种“块队列”机制：每个块开始时一次性生成 p 个步长并依次执行，只有当块内 p 个步长全部用完后，才更新历史窗口并生成下一块步长。

图 7.7 给出 $p=2$、$p=3$ 两种情形下的下降曲线对比，表 7.9 汇总迭代步数、最终函数值与包络拟合因子。

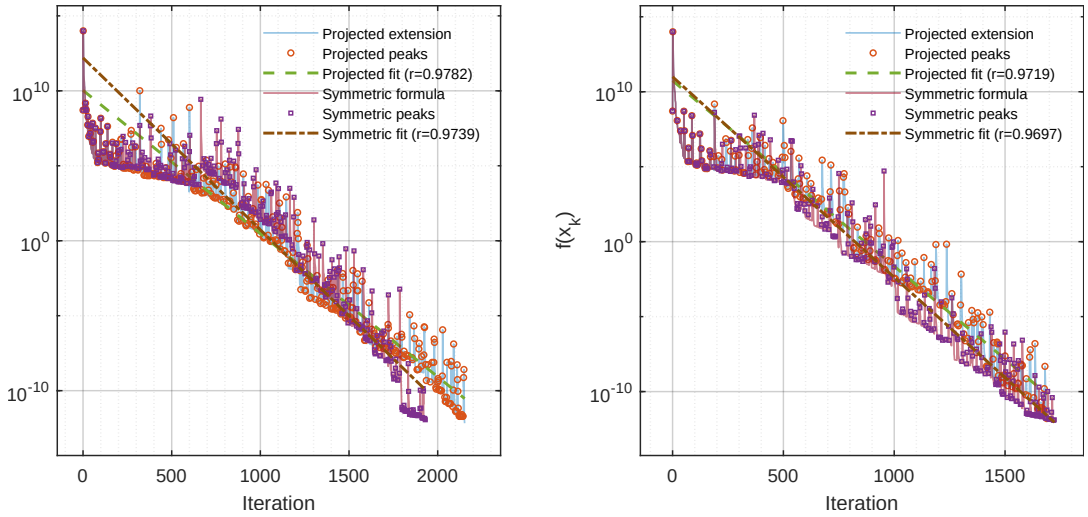


图 7.7　高维强凸光滑非二次问题（67）上的对比。左图为 $p=2$，右图为 $p=3$，两种方法都呈现明显的非单调下降与局部 R-线性包络。该测试族上，改进对称化公式在 $p=2$ 与 $p=3$ 情形中均显示更快的下降。

表 7.9　高维强凸非二次实验结果汇总。

方法	步长公式	迭代数	最终函数值	包络因子
$p=2$	精确 Hessian 投影	2114	9.56×10^{-13}	0.9771
$p=2$	改进对称化公式	**1931**	9.44×10^{-13}	**0.9739**
$p=3$	精确 Hessian 投影	1964	9.29×10^{-13}	0.9744
$p=3$	改进对称化公式	**1723**	9.66×10^{-13}	**0.9697**

从表 7.9 与图 7.7 可以看到：在该测试族上，改进对称化公式在 $p=2$ 与 $p=3$ 两个情形中均表现出更少的迭代次数与更小的包络因子。更重要的是，改进公式的定义本身不需要显式 Hessian，且避免了非对称小矩阵在一般非线性问题上可能带来的复特征值问题。

第十一节　Trigonometric 测试问题

Fletcher–Powell[21] 的三角函数测试问题：随机生成解 $x_*\in[-\pi,\pi]^n$，初始点 $x_0=x_*+\eta\xi$，其中 $\xi\sim\mathcal{N}(0,I_n)$、$\eta=10^{-1}$，每个 n 取 5 个随机种子求平均。表 7.10 给出 $n=50,100$ 的梯度评估数（5 次平均）。

方法	$n=50$	$n=100$
pD(S)-BB-GLL（$p=3$）	2081	1044
pD(S)-BB-GLL（$p=5$）	1081	858
pD(S)-BB-GLL（$p=7$）	708	**676**
pD(S)-BB-pure（$p=5$）	div†	div†
BB-Raydan	1461	803
CG-PR	1570	2581
L-BFGS-3	860	1023
L-BFGS-5	**536**	800

表 7.10　Trigonometric 测试问题的迭代数（5 次随机种子的平均）。† 超出 $N_{\max}=10000$ 仍未达停止准则。

第十二节　Chained Rosenbrock 测试问题

Toint[22] 的链式 Rosenbrock 函数

$$f(x)=\sum_{i=1}^{n-1}\left[100(x_{i+1}-x_i^2)^2+(1-x_i)^2\right],$$

初始点 $x_0=0$。这是非二次但形式简单的非凸问题（在远离极小点的区域目标函数沿狭长曲谷分布），是 BB-类方法相对薄弱的测试。

方法	$n=50$	$n=100$
pD(S)-BB-GLL（$p=3$）	3513	4958
pD(S)-BB-GLL（$p=5$）	2713	**2791**
pD(S)-BB-GLL（$p=7$）	1827	3023
BB-Raydan	2726	3301
CG-PR	1431	2609
L-BFGS-5	**574**	**1072**

表 7.11　Chained Rosenbrock 测试问题。L-BFGS-5 在两个维度上显著占优。pD(S)-BB-GLL($p=7$) 在 $n=100$ 上略劣于 $p=5$，反映在病态非凸问题上历史窗口过大可能引入过早曲率失真。

第十三节　Convex 2 测试问题

Raydan[5] 的可分严格凸非二次问题

$$f(x)=\sum_{i=1}^{n}\left(\mathrm{e}^{x_i}-x_i\right),$$

初始点 $x_0=(1,2,3,\cdots,n)/n$。该函数 Hessian $\operatorname{diag}(\mathrm{e}^{x_i})$ 在最优点 $x_*=0$ 附近退化为 I_n（条件数 1），但在初始点处条件数达到 $\mathrm{e}^{x_n}/\mathrm{e}^{x_1}=\mathrm{e}^{(n-1)/n}\approx\mathrm{e}$，在大 n 下仍然是良态非二次问题。

方法	$n = 10^3$	$n = 10^5$
pD(S)-BB-GLL（$p = 5$）	24	21
pD(S)-BB-GLL（$p = 10$）	40	36
BB-Raydan	**8**	**8**
CG-PR	14	14
L-BFGS-5	15	15

表 7.12　Convex 2 测试问题。BB-Raydan 在两个维度上均最快（仅 8 步），其原因是 Convex 2 是可分严格凸问题，BB1 步长在每个坐标方向上都接近最优。pD(S)-BB-GLL 反而较慢——历史窗口内的小广义特征值问题在可分良态问题上不提供额外信息，反而引入额外开销，这是历史窗口方法在该问题类上不占优的体现。该实验说明 pD(S)-BB 的优势不在所有非线性问题上一致。

第十四节　局部 R-线性渐近行为的经验拟合

定理 6.11 给出 GLL 全局化下 pD(S)-BB 在 $\mathcal{U}_{\mathrm{reg}}$ 内的局部 R-线性收敛。具体做法为：在尾部 50% 的轨道中识别出局部极大值点 k（即满足 $\|\nabla f(x_{k-1})\| < \|\nabla f(x_k)\| > \|\nabla f(x_{k+1})\|$ 的位置），对这些峰值的对数做最小二乘线性回归得到经验包络因子 $\hat{\rho}_{\mathrm{peaks}} = \exp(\mathrm{slope})$ 与拟合优度 R^2。表 7.13 给出 §十一–§十三 中各测试问题在 pD(S)-BB-GLL（$p = 5$）下的尾轨道经验包络速率。

问题	$\hat{\rho}_{\mathrm{peaks}}$	R^2	拟合点数	备注
Trigonometric $n = 50$	0.9862	+0.659	74	包络拟合
Trigonometric $n = 100$	0.9983	+0.301	191	包络拟合
Convex 2 $n = 10^3$	0.3921	+0.702	13	原始拟合（轨道过短）
Convex 2 $n = 10^5$	0.2713	+0.942	12	原始拟合（轨道过短）

表 7.13　pD(S)-BB-GLL（$p = 5$）尾轨道的经验包络速率。

表 7.9 中的包络因子落在 [0.9697, 0.9771] 之间，与表 7.13 中 Trigonometric $n = 50$ 的 0.9862 在同一量级。两组独立的拟合在不同非二次问题上都给出 $\hat{\rho} \in [0.97, 1)$ 的速率，构成了定理 6.11 第 (ii)–(iii) 点局部 R-线性结论的经验证据。

第十五节　讨论

非二次实验给出五点观察。第一，纯 pD(S)-BB 实现（算法 3）在 Trigonometric 这类有较强坐标耦合的非凸问题上发散，说明仅靠秩压缩与步长截断不足以替代 GLL 全局化机制，这印证了第六章把全局收敛与局部 R-线性分别交给 GLL 检测与块映射收缩处理的双层架构。第二，GLL 全局化下 pD(S)-BB 在 Trigonometric 上随窗口 p 增大单调改善，并在两个维度上都获得整体最优。$p = 7$ 在 $n = 100$

上甚至略快于 L-BFGS-5（676 vs. 800），表明历史窗口与对称化在某些非线性问题类上的实际收益不亚于带二阶曲率信息的拟牛顿方法。第三，在病态非凸的 Chained Rosenbrock 上 pD(S)-BB-GLL 弱于 L-BFGS-5（BFGS 可提供更多的曲率信息），但在 BB 类方法中表现稳健：$n = 50$ 时 $p = 7$ 最快（1827 vs. BB-Raydan 的 2726），$n = 100$ 时 $p = 5$ 最稳健（2791 vs. BB-Raydan 的 3301）。第四，Convex 2 这类可分良态强凸问题上 BB-Raydan 仅需 8 步即收敛，反而是历史窗口方法的弱项——这印证了 $p > 1$ 的小广义特征值问题在不需要历史信息的可分问题上不仅没有增益，反而引入额外开销。第五，对 R-线性渐近行为的经验拟合实验中，表 7.13 与表 7.9 给出的两组独立的包络拟合都得到 $\hat{\rho} \in [0.97, 1)$ 的速率，与定理 6.11 局部 R-线性结论一致。

第八章　总结与展望

第一节　本文证明范式的意义

对固定记忆谱梯度法的 R-线性收敛性质而言，本文给出两层证明范式：

(i) **全局版本：**存在一个固定维数的块状态空间（锥）X，算法在理想化固定块模型下诱导出连续、一次正齐次且全局有界的块映射 $T_q : X \to X$，并且对任意初始块状态 $E \in X$ 都有 $T_q^k(E) \to 0$；

(ii) **局部版本：**算法诱导的块映射并不在全局自然定义，但归一化轨道最终停留在一个与退化边界有正距离的紧子集 K 内，并且对每个 $z \in K$ 都有 $T_q^k(z) \to 0$。

前者对应引理 2.2 与定理 2.4 的使用方式；后者则对应命题 2.5。两种版本都说明：一旦算法逐点收敛，配合局部扰动估计和最终全接受，就能推出一般非线性全局化算法的局部 R-线性收敛。这给出了独立于 Dai–Liao[7] 以及 Dai[9] 在严格凸二次情形下沿着各个特征分量建立递推关系，再借助 Property A 证明 BB、LMSD、CBB 等谱梯度法的 R-线性收敛的另一路线。

本文还给出了一个严格凸二次问题下梯度法收敛的等价判据：对任何给定的步长序列，严格凸二次梯度法是否收敛，完全可以由特征值上的乘子积来刻画。

定理 8.1　设

$$q(x) = \frac{1}{2} x^T A x, \qquad A = \operatorname{diag}(\lambda_1, \cdots, \lambda_n) \succ 0.$$

给定任意正步长序列 $\{\alpha_k\}_{k \geqslant 0}$，定义梯度迭代

$$x_{k+1} = x_k - \alpha_k A x_k.$$

则以下两条等价：

(i) 对任意初始点 x_0，都有 $x_k \to 0$（等价地 $g_k \to 0$）；

(ii) 对每个特征值 $\lambda_i \in \sigma(A)$，都成立

$$\lim_{k \to \infty} \prod_{j=0}^{k} \left(1 - \alpha_j \lambda_i\right) = 0.$$

证明　在标准对角坐标下，

$$x_{k+1}^{(i)} = (1 - \alpha_k \lambda_i) x_k^{(i)}, \qquad x_k^{(i)} = \left(\prod_{j=0}^{k-1} (1 - \alpha_j \lambda_i) \right) x_0^{(i)}.$$

因此，(ii) 成立当且仅当对任意初始点 x_0 的每个坐标分量都有 $x_k^{(i)} \to 0$，这又等价于 $x_k \to 0$。由于 $g_k = A x_k$ 且 $A \succ 0$，故 $x_k \to 0$ 与 $g_k \to 0$ 也是等价的。　∎

若回到一般形式 $q(x)=\frac{1}{2}(x-x^*)^T A(x-x^*)$，只需把上面的 x_k 理解为相对极小点的误差变量 x_k-x^*。因此，定理 8.1 说明：在严格凸二次模型上，证明收敛所真正需要的只是证明这些乘子积趋于零。对本文的固定记忆 pD-BB，本文第 4 章通过行列式恒等式与 Cauchy–Binet 展开证明了相应块乘子多项式的收缩，而一旦普通收敛得到，在一定的假设下齐次块映射框架便会自动把它提升为统一有限步收缩和 R-线性。这是一个非常普适且模块化的方法。

第二节　对其他方法的适用性

本文说明，一旦某种固定周期谱方法已经能够写成固定维连续正齐次块映射，并且其严格凸二次普通收敛已知，那么第 2 章的抽象工具就会自动把它提升为 R-线性收敛。因此，本文框架原则上可推广到一类更广的固定记忆谱方法中，包括但不限于：

- 经典 BB1/BB2 及其相关扩展算法；
- Fletcher 的 LMSD 与 harmonic-LMSD 及其周期长度固定的变体[11,13,19]；
- 周期长度固定的 cyclic BB 类方法[10]；
- 各类 delayed / retarded gradient 方法[6,14,16-17]；
- 在活动集被识别后可化为固定维数局部块映射的一些投影谱梯度方法。

第三节　动力系统视角的拓展：经典 BB 方法的紧渐近速率

第 2 章的闭锥连续正齐次映射工具，把固定记忆谱方法的收敛性问题转化为有限维赋范空间上离散动力系统的问题。本节展示这一视角定量的一面，对最经典、研究得最久的 BB1 方法，借助遍历理论的工具，可以给出比既有结果更紧且是理论上最紧的的渐近速率上界，以此解决关于 BB 方法 R-收敛速率最佳估计的公开问题。

一、问题陈述与已有结果

考虑严格凸二次问题

$$\min_{x\in\mathbb{R}^n} q(x)=\frac{1}{2}x^T Hx-b^T x,\qquad H=H^T\succ 0, \tag{68}$$

其最优解为 $x^*=H^{-1}b$，梯度为

$$g_k=\nabla q(x_k)=H(x_k-x^*).$$

本节研究的对象是经典 BB1 步长以及一次最速下降预热：

$$x_{k+1}=x_k-\alpha_k g_k,\qquad \alpha_0=\frac{g_0^T g_0}{g_0^T H g_0},\qquad \alpha_k=\frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T H g_{k-1}}\quad (k\geqslant 1). \tag{69}$$

记 H 的条件数为 $\kappa(H)=\lambda_{\max}(H)/\lambda_{\min}(H)$，并令

$$c(H):=\frac{\kappa(H)-1}{\kappa(H)+1}=\frac{\lambda_{\max}(H)-\lambda_{\min}(H)}{\lambda_{\max}(H)+\lambda_{\min}(H)}\in[0,1). \tag{70}$$

本节严格刻画

$$r_{\mathrm{asy}}(H,x_0):=\limsup_{k\to\infty}\|g_k\|^{1/k}$$

的最佳上界，并给出关于初值一致的 $c+\varepsilon$ 型 R-线性估计。

关于 BB 方法在严格凸二次问题上的收敛性，已有以下经典结果：Barzilai 与 Borwein[2] 给出了二维严格凸二次情形的分析并观察到 R-超线性现象。Raydan[3] 证明了一般维数下的全局收敛性，并把 BB 与 Grippo–Lampariello–Lucidi 非单调线搜索结合得到一般非线性的全局化算法[5,31]；Dai 与 Liao[7] 进一步证明了 R-线性收敛，速率粗略地给出为 $(1-1/\kappa)^{1/n}$。Dai[24] 在二维情形下作了精细分析：除了一类二维对称轨道外，至多三步内必出现一次超线性步。Dai 与 Fletcher[25] 以及 Fletcher[30] 讨论了 BB 及若干相关梯度法的渐近行为。Yuan[27]、Dai 与 Yuan[26]、Frassoldati–Zanni–Zanghirati[18] 以及 De Asmundis 等[32] 引入了一系列具有有限终止性质或自适应特征的步长，致力于打破二维坏轨道引发的“锯齿振荡”。近期 Li 与 Sun[28] 将一般维数下的 R-线性速率上界改进为 $1-1/\kappa$；Huang 等[29] 从谱性质角度对一组包括最速下降在内的步长族给出了统一的渐近行为分析。从二维坏轨道出发，一个自然的猜测是（70）中的 $c(H)$ 才是真正的最优常数，本节把这一猜测严格证明。

与本节相关的最速下降法的渐近行为分析，可上溯至 Akaike[34]，并由 Forsythe[35]、Nocedal–Sartenaer–Zhu[33] 与 Pronzato–Wynn–Zhigljavsky[36] 在不同形式下推广。这些工作给出了步长由当前信息完全确定时梯度范数渐近行为的精细图像，但并未涵盖 BB 这种步长依赖于上一时刻能量分布的系统。

本节的主要结果如下。

定理 8.2 (BB 最紧渐近速率)　对任意 $H\succ 0$ 与任意初值 x_0，由（69）定义的 BB1 轨道满足

$$\limsup_{k\to\infty}\|g_k\|^{1/k}\leqslant c(H). \tag{71}$$

该上界由仅激活 $\lambda_{\min}(H)$ 与 $\lambda_{\max}(H)$ 两个端点特征值的对称轨道达到。因此，对每个 $H\succ 0$，

$$\sup_{x_0\in\mathbb{R}^n} r_{\mathrm{asy}}(H,x_0)=c(H).$$

特别地，在条件数为 κ 的严格凸二次问题族上，最优渐近指数常数为 $(\kappa-1)/(\kappa+1)$。

定理 8.3 (关于初值一致的 $c+\varepsilon$ 型估计)　设 $H\succ 0$，$c(H)$ 由（70）给出。则对任意 $\varepsilon>0$，存在仅依赖于 H 与 ε 的常数 $C_\varepsilon(H)>0$，使得对每个初值 x_0，

$$\|g_k\|\leqslant C_\varepsilon(H)\,(c(H)+\varepsilon)^k\,\|g_0\|,\qquad k\geqslant 0.$$

注　常数 $(\kappa-1)/(\kappa+1)$ 也恰好等于最优固定步长梯度法

$$x_{k+1}=x_k-\frac{2}{\lambda_{\min}+\lambda_{\max}}g_k$$

及最速下降法的最坏速率，从理论上揭示了 BB 方法收敛速率的优势。下文定理 8.12 将进一步说明，在固定激活谱的终止态紧化动力系统中，唯一的极大遍历测度是支撑在两端对称不动点上的 Dirac 测度。因此，从不变测度和周期轨道的角度看，常数 $c(H)$ 由两端对称轨道给出。需要注意的是，这一测度层面的唯一性本身并不等价于说每条非坏初值轨道的渐近指数都严格小于 $c(H)$：某些非坏初值可能在有限步后落入两端对称轨道。

二、谱压缩与归一化两步动力系统

由 $g_k=H(x_k-x^*)$，梯度递推

$$g_{k+1}=(I-\alpha_k H)g_k \tag{72}$$

是（69）的等价形式。对 $k\geqslant 1$，

$$\alpha_k=\frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T H g_{k-1}},$$

即上一时刻梯度的 Rayleigh 商倒数。

设 H 的不同特征值为

$$0<\mu_1<\mu_2<\cdots<\mu_m,$$

对应正交谱投影为 $P_1,\cdots,P_m$。若 $P_i g_0\neq 0$，则称特征值 μ_i 被初始梯度激活，激活特征值的集合记为

$$I\subset\{\mu_1,\cdots,\mu_m\},$$

按升序排列。定义谱块能量

$$E_i^k:=\left\|P_i g_k\right\|^2,\qquad i=1,\cdots,m. \tag{73}$$

引理 8.4　经典 BB 在严格凸二次问题上的动力学只依赖于被激活的不同特征值及其谱块能量，而与几何重数无关：

$$\alpha_k=\frac{\sum_{j=1}^m E_j^{k-1}}{\sum_{j=1}^m \mu_j E_j^{k-1}},\qquad E_i^{k+1}=(1-\alpha_k\mu_i)^2E_i^k,\quad i=1,\cdots,m. \tag{74}$$

特别地，任意激活了恰好 m 个不同特征值的高维 BB 轨道，与一个 m 维对角二次问题上的 BB 轨道严格等价。

证明　由 $H=\sum_i \mu_i P_i$ 及（72）有 $P_i g_{k+1}=(1-\alpha_k\mu_i)P_i g_k$，两边取范数平方即得 $E_i^{k+1}=(1-\alpha_k\mu_i)^2E_i^k$。关于 α_k 的表达式可由 $g_{k-1}^T g_{k-1}=\sum_j E_j^{k-1}$，$g_{k-1}^T H g_{k-1}=\sum_j \mu_j E_j^{k-1}$ 直接得到。 ∎

当轨道尚未终止时，定义归一化权重

$$w_i^k := \frac{E_i^k}{\sum_j E_j^k}, \qquad i \in I, \tag{75}$$

则 $w^k = (w_i^k)_{i\in I}$ 属于概率单纯形

$$\Delta_I := \Big\{ w = (w_i)_{i\in I} : \ w_i \geqslant 0, \ \sum_{i\in I} w_i = 1 \Big\}.$$

再令

$$u(w) := \sum_{i\in I} \mu_i w_i, \qquad \varphi_i(z) := \Big(1 - \frac{\mu_i}{u(z)}\Big)^2. \tag{76}$$

为统一处理预热步，约定 $w^{-1} := w^0$；由于 $\alpha_0 = 1/u(w^0)$，该约定与（69）一致。则

$$w_i^{k+1} = \frac{\varphi_i(w^{k-1}) w_i^k}{\sum_{j\in I} \varphi_j(w^{k-1}) w_j^k} = \frac{\varphi_i(w^{k-1}) w_i^k}{\big\langle \varphi(w^{k-1}), w^k \big\rangle}. \tag{77}$$

自然的状态变量是两步对

$$\xi_k := (w^k, w^{k-1}) \in X_I := \Delta_I \times \Delta_I, \qquad \xi_0 = (w^0, w^0).$$

再定义

$$r_I(w,z) := \sqrt{\langle \varphi(z), w \rangle}, \qquad T_I(w,z) := \Big(\frac{\varphi(z) \odot w}{\langle \varphi(z), w \rangle}, \ w \Big), \tag{78}$$

其中 $\odot$ 表示分量乘积，T_I 仅在 $r_I(w,z) > 0$ 时有意义。对真实 BB 轨道，只要尚未终止，

$$\xi_{k+1} = T_I(\xi_k), \qquad \big\| g_{k+1} \big\| = r_I(\xi_k) \big\| g_k \big\|. \tag{79}$$

令 $f_I := \log r_I$，则

$$\log \big\| g_n \big\| - \log \big\| g_0 \big\| = \sum_{k=0}^{n-1} f_I(\xi_k), \tag{80}$$

只要在前 n 步内尚未终止。

这里需要单独指出：当 $r_I(w,z) = 0$ 时，（77）的分母为零，归一化更新失去意义。然而对真实算法而言，由（79），$r_I(\xi_k) = 0$ 正等价于 $\big\| g_{k+1} \big\| = 0$，即下一步梯度已经为零、算法已经终止。后续将基于此引入终止态以便分析。

三、二维情形与速率下界

由引理 8.4，两点谱严格等价于二维对角问题。二维情形既给出最优常数的下界，也解释了为什么二维对称轨道在射影意义下只有一个点。本节主要对 Dai[24] 在二维情形下的主要结果进行回顾与陈述。

设激活谱为 $\{a,b\}$，$0<a<b$，对应单位特征向量为 v_a, v_b。把梯度写成 $g_k = x_k v_a + y_k v_b$，并定义模态能量比

$$q_k := \frac{x_k^2}{y_k^2} \in (0,+\infty).$$

命题 8.5　在两点谱情形下，最速下降预热使得

$$q_{k+1} = \frac{q_k}{q_{k-1}^2}, \quad k \geqslant 1, \qquad q_1 = \frac{1}{q_0}. \tag{81}$$

证明　对 $k \geqslant 1$，由（69）直接计算

$$\alpha_k = \frac{x_{k-1}^2 + y_{k-1}^2}{a x_{k-1}^2 + b y_{k-1}^2},$$

$$x_{k+1} = x_k(1-a\alpha_k) = x_k \cdot \frac{(b-a)y_{k-1}^2}{a x_{k-1}^2 + b y_{k-1}^2}, \quad y_{k+1} = -y_k \cdot \frac{(b-a)x_{k-1}^2}{a x_{k-1}^2 + b y_{k-1}^2}.$$

作比并取平方即得 $q_{k+1} = q_k/q_{k-1}^2$。最速下降预热使 $\alpha_0 = (x_0^2+y_0^2)/(a x_0^2 + b y_0^2)$，代入即得 $q_1 = y_0^2/x_0^2 = 1/q_0$。 ∎

命题 8.6 (二维对称轨道)　若 $q_0 = 1$，则 $q_k \equiv 1$，$\alpha_k \equiv 2/(a+b)$，且

$$g_k = c(a,b)^k \big(x_0 v_a + (-1)^k y_0 v_b\big), \qquad \|g_k\| = c(a,b)^k \, \|g_0\|,$$

其中 $c(a,b) := (b-a)/(b+a)$。

证明　由（81），$q_1 = 1$，再归纳得 $q_k \equiv 1$。两模态等模时 $\alpha_k = 2/(a+b)$；代入（72）即得结论。 ∎

进一步可证 $q_0 = 1$ 是（81）在射影意义下的唯一不动点：令 $\nu_k = \log q_k$，递推（81）等价于

$$\nu_{k+1} = \nu_k - 2\nu_{k-1}, \qquad \nu_1 = -\nu_0,$$

其特征方程 $r^2 - r + 2 = 0$ 的两根模都为 $\sqrt{2} > 1$。因此除 $\nu_0 = 0$ 外，$|\nu_k|$ 在 $\sqrt{2}^k$ 量级上振荡发散。Dai[24] 据此证明了至多三步内必出现一次超线性步。这恰好说明了二维对称轨道在射影空间上是唯一的，且是动力学上不稳定的。

推论 8.7 (最优常数的下界)　对任意 $H \succ 0$，把初始梯度仅激活在 $\lambda_{\min}(H)$ 与 $\lambda_{\max}(H)$ 两个端点特征值上且两端等能量，便有 $\|g_k\| = c(H)^k \, \|g_0\|$。因此定理 8.2 与定理 8.3 中的 $c(H)$ 不可再缩小，本节所求的速率上界是紧的。

四、终止态紧化

固定非空激活谱 $I = \{\mu_1 < \cdots < \mu_m\}$，记

$$X_I := \Delta_I \times \Delta_I, \qquad X_I^{\mathrm{reg}} := \{(w,z) \in X_I : r_I(w,z) > 0\}, \qquad \Sigma_I := X_I \setminus X_I^{\mathrm{reg}}.$$

附加一个孤立点 ∂，并令

$$\widehat{X}_I := X_I \sqcup \{\partial\}.$$

$\widehat{X}_I$ 在自然拓扑下是紧度量空间。定义

$$\widehat{T}_I(x) = \begin{cases} T_I(x), & x \in X_I^{\mathrm{reg}}, \\ \partial, & x \in \Sigma_I \cup \{\partial\}, \end{cases} \qquad \widehat{f}_I(x) = \begin{cases} \log r_I(x), & x \in X_I^{\mathrm{reg}}, \\ -\infty, & x \in \Sigma_I \cup \{\partial\}. \end{cases} \tag{82}$$

这一定义忠实地刻画了算法的真实行为：一旦 r_I 在某一步取零，下一步梯度就为零，算法终止。

命题 8.8 $\widehat{T}_I : \widehat{X}_I \to \widehat{X}_I$ Borel 可测，并在 $\widehat{X}_I \setminus \Sigma_I$ 上连续，$\widehat{f}_I$ 上半连续且有上界。

证明 T_I 在 X_I^{reg} 上是连续函数的有理运算，故连续。∂ 孤立。$r_I^2 = \langle \varphi(z), w \rangle$ 在紧集 X_I 上连续非负，故 $\log r_I$ 在 X_I^{reg} 上连续。当 $x_j \to x \in \Sigma_I$ 时 $\log r_I(x_j) \to -\infty$，从而上半连续。由 r_I^2 的有界性可知 $\widehat{f}_I$ 有上界。 ∎

对每个非空 $J \subset I$，定义

$$\Delta_J := \{w \in \Delta_I : w_i = 0,\ i \notin J\}, \qquad X_J := \Delta_J \times \Delta_J, \qquad \widehat{X}_J := X_J \cup \{\partial\}.$$

由（78）直接可见，每个 $\widehat{X}_J$ 都在 $\widehat{T}_I$ 下前向不变。

引理 8.9 对任意 $\widehat{T}_I$-不变 Borel 概率测度 ν，$\nu(\Sigma_I) = 0$。若进一步 $\int \widehat{f}_I \, \mathrm{d}\nu > -\infty$，则 $\nu(\{\partial\}) = 0$。

证明 由（82），$\widehat{T}_I^{-1}(\{\partial\}) = \Sigma_I \cup \{\partial\}$，由不变性 $\nu(\{\partial\}) = \nu(\Sigma_I) + \nu(\{\partial\})$，故 $\nu(\Sigma_I) = 0$。若 $\nu(\{\partial\}) > 0$，由 $\widehat{f}_I(\partial) = -\infty$ 立得 $\int \widehat{f}_I \, \mathrm{d}\nu = -\infty$，与假设矛盾。 ∎

这一引理说明，所有与最优平均增长率相关的不变测度都自动支撑在正则集合内。

五、端点恒等式与最优常数

固定非空 $J \subset I$，记

$$a_J := \min J, \qquad b_J := \max J, \qquad c_J := \frac{b_J - a_J}{b_J + a_J},$$

$$P_J := \{(w, z) \in X_J : \ w_i > 0,\ z_i > 0,\ \forall i \in J\}.$$

在 P_J 上 $u(z) \in (a_J, b_J)$，下文涉及的对数均良定义。

引理 8.10 (加权端点不等式) 对任意 $0 < a < b$ 与 $u \in [a, b]$，

$$\left(\frac{u-a}{u}\right)^{\frac{b}{a+b}} \left(\frac{b-u}{u}\right)^{\frac{a}{a+b}} \leqslant \frac{b-a}{a+b}, \tag{83}$$

等号当且仅当 $u = (a+b)/2$。

证明 令 $t := (u-a)/(b-u) \in [0,\infty]$，则 $u = (a+tb)/(1+t)$，$(u-a)/u = t(b-a)/(a+tb)$，$(b-u)/u = (b-a)/(a+tb)$。（83）的左端化为

$$\frac{b-a}{a+tb}\cdot t^{\frac{b}{a+b}}.$$

对值 1 与 t 分别取权重 $a/(a+b)$ 与 $b/(a+b)$ 的加权 AM–GM，得

$$\frac{a+tb}{a+b} \geqslant t^{\frac{b}{a+b}},$$

代入即给出上界 $(b-a)/(a+b)$。等号当且仅当 $t=1$，即 $u=(a+b)/2$。 ∎

引理 8.11 (端点恒等式) 对 $x=(w,z)\in P_J$，定义

$$H_J(w,z) := -\frac{b_J}{2(a_J+b_J)}\log w_{a_J} - \frac{a_J}{2(a_J+b_J)}\log w_{b_J}, \tag{84}$$

$$\psi_J(w,z) := \log c_J - \frac{b_J}{a_J+b_J}\log\frac{u(z)-a_J}{u(z)} - \frac{a_J}{a_J+b_J}\log\frac{b_J-u(z)}{u(z)}. \tag{85}$$

则

$$\widehat{f}_I(x) = \log c_J - \psi_J(x) + H_J(\widehat{T}_I x) - H_J(x), \tag{86}$$

且 $\psi_J(x)\geqslant 0$，等号当且仅当 $u(z)=(a_J+b_J)/2$。

证明 在 P_J 上 $\widehat{T}_I = T_I$。记 $a=a_J$，$b=b_J$，$u=u(z)$。由（78），

$$w_a^+ = \frac{(u-a)^2}{u^2 r_I(x)^2}w_a, \quad w_b^+ = \frac{(b-u)^2}{u^2 r_I(x)^2}w_b.$$

直接计算

$$\begin{aligned} H_J(\widehat{T}_I x) - H_J(x) &= -\frac{b}{2(a+b)}\log\frac{(u-a)^2}{u^2 r_I(x)^2} - \frac{a}{2(a+b)}\log\frac{(b-u)^2}{u^2 r_I(x)^2} \\ &= -\frac{b}{a+b}\log\frac{u-a}{u} - \frac{a}{a+b}\log\frac{b-u}{u} + \log r_I(x). \end{aligned}$$

整理即得（86）。$\psi_J\geqslant 0$ 是引理 8.10 的对数形式。 ∎

（86）把一步对数增长率分解为三部分：临界斜率 $\log c_J$，非负缺陷 ψ_J，以及上下边界项 $H_J\circ\widehat{T}_I - H_J$。这一结构是后续证明的核心。

定义

$$\hat{\beta}_I := \sup\Big\{\int \widehat{f}_I\,\mathrm{d}\nu :\ \nu \text{ 为}\widehat{T}_I\text{-不变 Borel 概率测度}\Big\}. \tag{87}$$

定理 8.12 (最优常数) 设 $|I|\geqslant 2$，$c_I := (\max I - \min I)/(\max I + \min I)$。则

$$\hat{\beta}_I = \log c_I.$$

唯一的极大遍历测度是对称两端不动点的 Dirac 测度

$$\delta_{x^I_{\mathrm{bad}}}, \qquad x^I_{\mathrm{bad}} := \big(\tfrac{1}{2}e_{a_I} + \tfrac{1}{2}e_{b_I},\ \tfrac{1}{2}e_{a_I} + \tfrac{1}{2}e_{b_I}\big),$$

其中 e_{a_I}, e_{b_I} 为 Δ_I 的相应顶点。

证明需应用 Birkhoff 遍历定理、Poincaré 递归与上半连续函数的 Portmanteau 不等式。

引理 8.13 (i) Birkhoff 遍历定理[41-42]：对遍历测度 ν 与 $\varphi \in L^1(\nu)$，

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi(T^k x) \to \int \varphi \,\mathrm{d}\nu, \qquad \nu\text{-a.e.}$$

(ii) Poincaré 递归：设 T 为紧度量空间上的 Borel 映射，ν 为 T-不变 Borel 概率测度。则 ν-a.e. 的点 x 是递归的，即存在整数序列 $n_\ell \to \infty$ 使

$$T^{n_\ell} x \to x.$$

(iii) 上半连续 Portmanteau：紧度量空间上 $\mu_n \Rightarrow \mu$ 时，对上半连续且有上界的 $\phi : X \to [-\infty, \infty)$，

$$\limsup_n \int \phi \,\mathrm{d}\mu_n \leqslant \int \phi \,\mathrm{d}\mu.$$

定理 8.12 的证明 下界：x^I_{bad} 是 $\widehat{T}_I$ 的不动点。事实上，

$$u(z) = \frac{a_I + b_I}{2}, \qquad \varphi_{a_I}(z) = \varphi_{b_I}(z) = c_I^2,$$

于是支撑在两个端点且权重各为 1/2 的 w 经 T_I 后仍保持不变。又有 $\widehat{f}_I(x^I_{\mathrm{bad}}) = \log c_I$，故

$$\hat{\beta}_I \geqslant \log c_I.$$

上界先归约到遍历测度。对任意 $\widehat{T}_I$-不变 Borel 概率测度 ν，作遍历分解

$$\nu = \int \eta \,\mathrm{d}\Lambda_\nu(\eta).$$

若对每个遍历分量 η 都已证

$$\int \widehat{f}_I \,\mathrm{d}\eta \leqslant \log c_I,$$

则对截断函数

$$\widehat{f}_I^{(M)} := \max\{\widehat{f}_I, -M\}$$

应用遍历分解，再令 $M \to \infty$ 并用单调收敛定理，即得

$$\int \widehat{f}_I \,\mathrm{d}\nu \leqslant \log c_I.$$

因此只需考虑遍历测度。

设 ν 遍历且

$$\int \widehat{f}_I \,\mathrm{d}\nu > -\infty.$$

由引理 8.9，

$$\nu(\Sigma_I) = \nu(\{\partial\}) = 0.$$

考虑使 $\nu(X_J)=1$ 的最小非空子集 $J\subset I$。这样的 J 存在，因为 I 有限。若 $J=\{\mu\}$ 为单点，则在 X_J 上有 $r_I=0$，从而 $X_J\subset\Sigma_I$，这与 $\nu(\Sigma_I)=0$ 矛盾。因此

$$|J|\geqslant 2.$$

下面说明 ν 实际支撑在 P_J 上。由 J 的极小性，对每个 $i\in J$ 都有

$$\nu(X_{J\setminus\{i\}})=0.$$

若 $w_i=0$ 且点在正则集合内，则由 T_I 的定义可知下一步的两个边缘在第 i 个分量上都为零，因而

$$\{x\in X_J\cap X_I^{\mathrm{reg}}:w_i=0\}\subset\widehat{T}_I^{-1}(X_{J\setminus\{i\}}).$$

再结合 $\nu(\Sigma_I)=0$ 与不变性，得到

$$\nu(\{w_i=0\})=0,\qquad i\in J.$$

另一方面，由 $\widehat{T}_I(w,z)=(w^+,w)$ 在正则集合上的形式以及 ν 的不变性可知，w 与 z 的边缘测度相同：对任意连续函数 h，

$$\int h(z)\,\mathrm{d}\nu=\int h(w)\,\mathrm{d}\nu.$$

因此同样有

$$\nu(\{z_i=0\})=0,\qquad i\in J.$$

于是

$$\nu(P_J)=1.$$

记

$$P_J^\infty:=\bigcap_{n\geqslant 0}\widehat{T}_I^{-n}(P_J\cap X_I^{\mathrm{reg}}).$$

由上面的结论和不变性，

$$\nu(P_J^\infty)=1.$$

对每个 $M>0$，令

$$\psi_{J,M}:=\min\{\psi_J,M\}.$$

它是有界 Borel 函数。取同时满足 Birkhoff 遍历定理和 Poincaré 递归的全测度集合

$$E\subset P_J^\infty.$$

任取 $x\in E$。由 Poincaré 递归，存在 $n_\ell\to\infty$ 使

$$\widehat{T}_I^{n_\ell}x\to x.$$

由于这些点均属于 P_J，且 H_J 在 P_J 上连续，故

$$H_J(\widehat{T}_I^{n_\ell}x) - H_J(x) = o(n_\ell).$$

把端点恒等式（86）沿轨道求和，得到

$$\sum_{k=0}^{n-1} \widehat{f}_I(\widehat{T}_I^k x) = n\log c_J - \sum_{k=0}^{n-1} \psi_J(\widehat{T}_I^k x) + H_J(\widehat{T}_I^n x) - H_J(x).$$

令 $n = n_\ell$，两端除以 n_ℓ 并取极限。由 Birkhoff 定理与 $\psi_J \geqslant 0$ 得

$$\int \widehat{f}_I \,\mathrm{d}\nu = \log c_J - \liminf_{\ell\to\infty} \frac{1}{n_\ell} \sum_{k=0}^{n_\ell-1} \psi_J(\widehat{T}_I^k x) \leqslant \log c_J.$$

由于 $J \subset I$，有

$$c_J = \frac{b_J - a_J}{b_J + a_J} \leqslant \frac{b_I - a_I}{b_I + a_I} = c_I.$$

因此

$$\int \widehat{f}_I \,\mathrm{d}\nu \leqslant \log c_I.$$

上界得证。

最后证明唯一性。设 ν 为遍历极大测度，即

$$\int \widehat{f}_I \,\mathrm{d}\nu = \log c_I.$$

则上面所有不等式均取等。首先必须有

$$c_J = c_I.$$

由于函数 $(b-a)/(b+a)$ 关于右端点 b 单调递增、关于左端点 a 单调递减，且 $J \subset I$，故

$$a_J = a_I, \qquad b_J = b_I.$$

其次，由

$$\liminf_{\ell\to\infty} \frac{1}{n_\ell} \sum_{k=0}^{n_\ell-1} \psi_J(\widehat{T}_I^k x) = 0$$

以及 $\psi_J \geqslant 0$，对任意 $M > 0$ 有

$$\frac{1}{n_\ell} \sum_{k=0}^{n_\ell-1} \psi_{J,M}(\widehat{T}_I^k x) \leqslant \frac{1}{n_\ell} \sum_{k=0}^{n_\ell-1} \psi_J(\widehat{T}_I^k x).$$

沿使右端趋于 0 的子列并用 Birkhoff 定理，得到

$$\int \psi_{J,M} \,\mathrm{d}\nu = 0, \qquad M > 0.$$

令 $M \to \infty$，由单调收敛定理得

$$\int \psi_J \,\mathrm{d}\nu = 0.$$

因此 $\psi_J = 0$ 在 ν 几乎处处成立。由引理 8.11 中的等号条件，

$$u(z) = A_J := \frac{a_J + b_J}{2} \qquad \nu\text{-a.e.}$$

又因为 w 与 z 的边缘测度相同，故

$$u(w) = A_J \qquad \nu\text{-a.e.}$$

在该集合上，

$$r_I(x)^2 = \sum_{i \in J} \left(1 - \frac{\mu_i}{A_J}\right)^2 w_i.$$

记

$$\gamma_i := \left(1 - \frac{\mu_i}{A_J}\right)^2.$$

则

$$\gamma_{a_J} = \gamma_{b_J} = c_J^2, \qquad \gamma_i < c_J^2, \quad i \in J \setminus \{a_J, b_J\}.$$

因此

$$r_I(x)^2 \leqslant c_J^2,$$

且若中间谱权重为正，则严格小于 c_J^2。另一方面，由极大性以及上式可知

$$\widehat{f}_I(x) = \log r_I(x) \leqslant \log c_J = \log c_I$$

且其积分等于 $\log c_I$，故

$$r_I(x) = c_J \qquad \nu\text{-a.e.}$$

于是所有中间谱权重必须为零。结合前面已经证明的 $w_i > 0$、$z_i > 0$ 在 J 上几乎处处成立，只能有

$$J = \{a_J, b_J\}.$$

再由

$$u(w) = \frac{a_J + b_J}{2}$$

得到

$$w_{a_J} = w_{b_J} = \frac{1}{2}.$$

同理，

$$z_{a_J} = z_{b_J} = \frac{1}{2}.$$

因此

$$\nu = \delta_{x_{\mathrm{bad}}^I}.$$

唯一性得证。 ∎

六、从测度回到轨道：定理 8.2 与 8.3 的证明

最后一步是把测度层面的最优常数翻译成具体轨道的渐近指数和与初值一致的常数。需要再加一条关于经验测度极限不变性的引理。

引理 8.14 紧度量空间 X 上，$T : X \to X$ 为 Borel 映射，并在闭集 $D \subset X$ 之外连续。对一列轨道段 $x_{k+1}^{(j)} = T(x_k^{(j)})$，$k = 0, \cdots, n_j - 1$，记 $\mu_j := \frac{1}{n_j} \sum_{k=0}^{n_j-1} \delta_{x_k^{(j)}}$。若 $\mu_j \Rightarrow \mu$ 且 $\mu(D) = 0$，则 μ 在 T 下不变。

证明 取连续有界 $\varphi : X \to \mathbb{R}$，求和知

$$\int \varphi \circ T \,\mathrm{d}\mu_j - \int \varphi \,\mathrm{d}\mu_j = \frac{\varphi(x_{n_j}^{(j)}) - \varphi(x_0^{(j)})}{n_j} \to 0.$$

$\varphi \circ T$ 的不连续点集包含于 D，由 $\mu(D) = 0$ 及 $\mu_j \Rightarrow \mu$ 推出 $\int \varphi \circ T \,\mathrm{d}\mu_j \to \int \varphi \circ T \,\mathrm{d}\mu$，$\int \varphi \,\mathrm{d}\mu_j \to \int \varphi \,\mathrm{d}\mu$。故 $\int \varphi \circ T \,\mathrm{d}\mu = \int \varphi \,\mathrm{d}\mu$，连续有界函数能区分测度。 ∎

定理 8.15 设经典 BB 轨道激活谱为 I，$|I| \geqslant 2$。则

$$\limsup_{k \to \infty} \left\| g_k \right\|^{1/k} \leqslant c_I.$$

证明 若有限步终止，则左端为零，结论显然。下面假设轨道不有限步终止，于是

$$\xi_k \in X_I^{\mathrm{reg}}, \qquad k \geqslant 0.$$

令

$$\mu_N := \frac{1}{N} \sum_{k=0}^{N-1} \delta_{\xi_k}.$$

由于 $\widehat{X}_I$ 紧，任意子列都有弱收敛子列。

记

$$L := \limsup_{N \to \infty} \frac{1}{N} \log \left\| g_N \right\|.$$

若 $L = -\infty$，结论显然。否则取 $N_j \to \infty$ 使

$$\frac{1}{N_j} \log \left\| g_{N_j} \right\| \to L,$$

并进一步取子列使

$$\mu_{N_j} \Rightarrow \mu.$$

由于轨道不进入 ∂，有

$$\mu_{N_j}(\{\partial\}) = 0.$$

又 ∂ 是孤立点，故 $\mathbf{1}_{\{\partial\}}$ 连续，从而

$$\mu(\{\partial\}) = 0.$$

由加性公式（80），

$$\frac{1}{N_j}\log\left\|g_{N_j}\right\| = \frac{1}{N_j}\log\|g_0\| + \int \widehat{f}_I \,\mathrm{d}\mu_{N_j}.$$

因为 $\widehat{f}_I$ 上半连续且有上界，由引理 8.13(iii) 得

$$\limsup_{j\to\infty}\int \widehat{f}_I \,\mathrm{d}\mu_{N_j} \leqslant \int \widehat{f}_I \,\mathrm{d}\mu.$$

左端对应的 L 不是 $-\infty$，故右端不为 $-\infty$。再结合

$$\widehat{f}_I = -\infty \quad \text{on} \quad \Sigma_I \cup \{\partial\}$$

以及 $\mu(\{\partial\}) = 0$，推出

$$\mu(\Sigma_I) = 0.$$

由引理 8.14，取 $D = \Sigma_I$，可知 μ 是 $\widehat{T}_I$-不变测度。于是由定理 8.12，

$$\int \widehat{f}_I \,\mathrm{d}\mu \leqslant \log c_I.$$

因此

$$L \leqslant \int \widehat{f}_I \,\mathrm{d}\mu \leqslant \log c_I.$$

这等价于

$$\limsup_{k\to\infty}\left\|g_k\right\|^{1/k} \leqslant c_I.$$

∎

定理 8.2 的证明 任意激活谱 I 满足 $\min I \geqslant \lambda_{\min}(H)$，$\max I \leqslant \lambda_{\max}(H)$，故 $c_I \leqslant c(H)$。由定理 8.15 及推论 8.7（达到性）即得。 ∎

定理 8.16 设 $|I| \geqslant 2$。对任意 $\eta > 0$，存在常数 $C_{I,\eta} < \infty$ 使得对每个初值 $x \in X_I$ 与每个 $n \geqslant 1$，

$$\sum_{k=0}^{n-1} \widehat{f}_I(\widehat{T}_I^k x) \leqslant n(\log c_I + \eta) + C_{I,\eta}.$$

对真实 BB 轨道这等价于 $\left\|g_n\right\| \leqslant \exp(C_{I,\eta})(c_I \mathrm{e}^{\eta})^n \left\|g_0\right\|$。

证明 反证。若结论不成立，则存在 $x^{(j)} \in X_I$ 与 $n_j \geqslant 1$ 使

$$\sum_{k=0}^{n_j-1} \widehat{f}_I(\widehat{T}_I^k x^{(j)}) - n_j(\log c_I + \eta) \to +\infty.$$

由于 $\widehat{f}_I$ 在 $\widehat{X}_I$ 上有上界，记其上界为 M_*，则左端不大于

$$n_j(M_* - \log c_I - \eta).$$

因此必有

$$n_j \to \infty.$$

此外，上式趋于 $+\infty$ 也说明前 n_j 步中没有进入 ∂；否则和式中会出现 $-\infty$ 项，不可能趋于 $+\infty$。

定义经验测度

$$\mu_j := \frac{1}{n_j} \sum_{k=0}^{n_j-1} \delta_{\widehat{T}_I^k x^{(j)}}.$$

由 $\widehat{X}_I$ 紧，取子列仍记为 μ_j，使

$$\mu_j \Rightarrow \mu.$$

由于前 n_j 步不进入 ∂，

$$\mu_j(\{\partial\}) = 0.$$

又因为 ∂ 是孤立点，$\mathbf{1}_{\{\partial\}}$ 连续，所以

$$\mu(\{\partial\}) = 0.$$

由反设除以 n_j 得

$$\limsup_{j\to\infty} \int \widehat{f}_I \,\mathrm{d}\mu_j \geqslant \log c_I + \eta.$$

另一方面，$\widehat{f}_I$ 上半连续且有上界，故由 Portmanteau 不等式，

$$\log c_I + \eta \leqslant \limsup_{j\to\infty} \int \widehat{f}_I \,\mathrm{d}\mu_j \leqslant \int \widehat{f}_I \,\mathrm{d}\mu.$$

特别地，右端不为 $-\infty$。结合 $\mu(\{\partial\}) = 0$ 与

$$\widehat{f}_I = -\infty \quad \text{on} \quad \Sigma_I \cup \{\partial\},$$

得到

$$\mu(\Sigma_I) = 0.$$

由引理 8.14，取 $D = \Sigma_I$，可知 μ 是 $\widehat{T}_I$-不变测度。于是由定理 8.12，

$$\int \widehat{f}_I \,\mathrm{d}\mu \leqslant \log c_I.$$

这与

$$\int \widehat{f}_I \,\mathrm{d}\mu \geqslant \log c_I + \eta$$

矛盾。 ∎

定理 8.3 的证明 固定 $H \succ 0$、$\varepsilon > 0$，记 H 的不同特征值集为 $\Lambda(H)$。对每个非空 $I \subset \Lambda(H)$：若 $|I| = 1$，对应轨道有限步终止；若 $|I| \geqslant 2$，则 $c_I \leqslant c(H) < c(H) + \varepsilon$，可取 $\eta_I > 0$ 使 $c_I \mathrm{e}^{\eta_I} \leqslant c(H) + \varepsilon$，由定理 8.16 得存在 C_{I,η_I} 使 $\|g_n\| \leqslant \exp(C_{I,\eta_I})(c(H)+\varepsilon)^n \|g_0\|$。因 $\Lambda(H)$ 有限，

$$C_\varepsilon(H) := \max\bigl\{1,\, \exp(C_{I,\eta_I}) :\ \varnothing \neq I \subset \Lambda(H),\ |I| \geqslant 2\bigr\}$$

即给出关于初值一致的常数。 ∎

七、精确周期轨道的初等证明

为方便对遍历优化不熟悉的读者独立检验上述常数 $c(H)$ 的正确性，本小节给出仅用引理 8.10 与（79）的精确周期轨道初等证明。它本质上是定理 8.12 应用于支撑在有限循环上的原子遍历测度的特例，但本节论证完全脱离测度论，仅需简单的初等放缩方法。

定义 8.1 (精确 p-周期轨道)　固定 I，$|I|\geqslant 2$。映射 T_I 的精确 p-周期轨道是序列 $\xi^{(0)},\dots,\xi^{(p-1)}\in X_I^{\mathrm{reg}}$ 满足 $\xi^{(j+1)}=T_I(\xi^{(j)})$ 且 $\xi^{(j+p)}=\xi^{(j)}$（指标 $\bmod p$）。记 $\nu_j:=u(z^{(j)})$，$R_j:=r_I(\xi^{(j)})^2$，每步线性因子 $\rho:=\big(\prod_{j=0}^{p-1}R_j\big)^{1/(2p)}$。

定理 8.17 (周期轨道的速率上界)　对任意精确周期轨道，$\rho\leqslant c_I$，等号成立当且仅当该周期退化为对称两端不动点 x^I_{bad}。

证明　由 $\xi^{(j+1)}=T_I(\xi^{(j)})$ 知 $z^{(j+1)}=w^{(j)}$，故 $z^{(j)}=w^{(j-1)}\ (\mathrm{mod}\,p)$。由（78），

$$w_i^{(j+1)}=\frac{(1-\mu_i/\nu_j)^2}{R_j}\,w_i^{(j)}.$$

共同的活跃支撑 $S:=\{\mu_i\in I:\ w_i^{(j)}>0\ \text{对某个}j\}$ 良定义，且对所有 j 一致。对每个 $\mu_i\in S$，沿一周期乘起来得

$$\Big|\prod_{\ell=0}^{p-1}(1-\mu_i/\nu_\ell)\Big|=\rho^p,\qquad \mu_i\in S.$$

记 $\nu_-:=\min_j\nu_j$，$\nu_+:=\max_j\nu_j$。取达到极小、极大的两步上 S 中的某点 $\mu_-\in S\cap[a_I,\nu_-]$，$\mu_+\in S\cap[\nu_+,b_I]$。多项式 $P(\mu):=\prod_{j=0}^{p-1}(1-\mu/\nu_j)$ 在 $[a_I,\nu_-]$ 上各因子非负且关于 μ 单调不增，故

$$\rho^p=|P(\mu_-)|=P(\mu_-)\leqslant P(a_I)=\prod_{j=0}^{p-1}\frac{\nu_j-a_I}{\nu_j}=:F.$$

在 $[\nu_+,b_I]$ 上各因子非正，绝对值 $|1-\mu/\nu_j|=(\mu-\nu_j)/\nu_j$ 关于 μ 单调不减，故

$$\rho^p=|P(\mu_+)|\leqslant|P(b_I)|=\prod_{j=0}^{p-1}\frac{b_I-\nu_j}{\nu_j}=:G.$$

取权重 $\theta=b_I/(a_I+b_I)$，$1-\theta=a_I/(a_I+b_I)$，则 $\min\{F,G\}\leqslant F^{\theta}G^{1-\theta}$，故

$$\rho^p\leqslant\prod_{j=0}^{p-1}\Big(\frac{\nu_j-a_I}{\nu_j}\Big)^{\frac{b_I}{a_I+b_I}}\Big(\frac{b_I-\nu_j}{\nu_j}\Big)^{\frac{a_I}{a_I+b_I}}.$$

对每个 $\nu_j\in[a_I,b_I]$ 应用引理 8.10，得 $\rho^p\leqslant c_I^p$。

取等条件：若 $\rho=c_I$，则上述每一步都取等。由引理 8.10 的等号条件，$\nu_j\equiv A_I:=(a_I+b_I)/2$。再代回得 $|1-\mu_i/A_I|=c_I$ 对每个 $\mu_i\in S$ 成立，这一方程在 $[a_I,b_I]$ 上只有两解 a_I,b_I，故 $S\subset\{a_I,b_I\}$。又 $\nu_j=A_I$ 是支撑在 S 上的概率均值，S 不能为单点集，故 $S=\{a_I,b_I\}$。结合 $\nu_j=A_I$ 得 $z^{(j)}_{a_I}=z^{(j)}_{b_I}=1/2$，由 $z^{(j)}=w^{(j-1)}$ 同样有 $w^{(j)}_{a_I}=w^{(j)}_{b_I}=1/2$，故 $\xi^{(j)}=x^I_{\mathrm{bad}}$。反之直接验证 x^I_{bad} 给出 $\rho=c_I$。 ∎

八、与本文整体框架的关系

本节的定理证明，与第 2 章的闭锥连续正齐次映射工具属于同一思路下的不同实现。两者共同的起点是：把迭代序列在合适紧空间上的逐步演化看作离散动力系统，再利用紧性、连续性与齐次性把逐点行为升级为带常数的速率估计。第 2 章在闭锥上使用有限覆盖把“逐点收敛”自动升级为“统一有限步收缩”，本节则在终止态紧空间上用端点恒等式与遍历测度的极大化，把最坏渐近速率确定到了精确常数 $c(H)$。这两条路径相互补充，共同呈现动力系统视角在本文框架内的灵活性。

经典 BB 方法是这一视角的“最干净”实例：$p = 1$，$|I| \geqslant 2$，且 T_I 在正则集上是有理映射。对 $p > 1$ 的固定记忆方法以及 BB2、cyclic BB 等变体，状态空间维数会相应增加，但相同的方法论框架仍然适用，只是单纯形上的端点结构需要重新刻画，相应的端点恒等式也要随步长公式重新计算。这是后续工作可以进一步推进的方向。

第四节　问题与展望

从前文来看，目前有以下几点值得继续研究：

(i) 如何把带秩压缩、删列或重启的实现写成一个固定维块映射，从而把第 2 章的抽象工具真正无缝地用于实际实现；

(ii) 对 Y 型步长和其他 ω 族中的步长，如何在第 6 章中建立与 S 型完全平行的局部步长扰动估计与局部 R-线性传递；

(iii) 对更一般的 delayed / cyclic / projected fixed-memory 方法，怎样选择最自然的状态空间，使“普通收敛 $\Rightarrow$ R-线性收敛”这一机制可以继续沿用；

(iv) 紧速率定理 8.2 是否可去掉常数中的 ε；

(v) 把 §三 中终止态紧化与端点恒等式的范式推广到 BB2、cyclic BB 以及其他固定记忆 ω 加权 Ritz-sweep 方法，确定相应的最坏渐近常数。

前三点是本文齐次块映射框架能否进一步扩展到更一般实现对象的关键，后两点则是 §三 这一独立延伸自然引出的开放问题。

此外，从更广泛的角度看，将迭代过程视作离散动力系统，为我们研究各类优化方法提供了新视角。本文便从其中的两个角度体现了这一思路的效用。如何进一步引入更多的离散动力系统或遍历理论的工具来解决优化问题中其他类型的速率与稳定性问题，是后续值得继续探索的。

参考文献

[1] Grippo L, Lampariello F, Lucidi S. A nonmonotone line search technique for Newton's method[J]. SIAM Journal on Numerical Analysis, 1986, 23: 707–716.

[2] Barzilai J, Borwein J M. Two-point step size gradient methods[J]. IMA Journal of Numerical Analysis, 1988, 8: 141–148.

[3] Raydan M. On the Barzilai and Borwein choice of steplength for the gradient method[J]. IMA Journal of Numerical Analysis, 1993, 13: 321–326.

[4] Molina B, Raydan M. Preconditioned Barzilai-Borwein method for the numerical solution of partial differential equations[J]. Numerical Algorithms, 1996, 13: 45–60.

[5] Raydan M. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem[J]. SIAM Journal on Optimization, 1997, 7: 26–33.

[6] Friedlander A, Martínez J M, Molina B, et al. Gradient method with retards and generalizations[J]. SIAM Journal on Numerical Analysis, 1999, 36: 275–289.

[7] Dai Y H, Liao L Z. R-linear convergence of the Barzilai and Borwein gradient method[J]. IMA Journal of Numerical Analysis, 2002, 22: 1–10.

[8] Dai Y H. On the nonmonotone line search[J]. Journal of Optimization Theory and Applications, 2002, 112: 315–330.

[9] Dai Y H. Alternate step gradient method[J]. Optimization, 2003, 52: 395–415.

[10] Dai Y H, Hager W W, Schittkowski K, et al. The cyclic Barzilai–Borwein method for unconstrained optimization[J]. IMA Journal of Numerical Analysis, 2006, 26: 604–627.

[11] Fletcher R. A limited memory steepest descent method[J]. Mathematical Programming, 2012, 135: 413–436.

[12] Curtis F E, Guo W. Handling nonpositive curvature in a limited memory steepest descent method[J]. IMA Journal of Numerical Analysis, 2016, 36: 717–742.

[13] Curtis F E, Guo W. R-linear convergence of limited memory steepest descent[J]. IMA Journal of Numerical Analysis, 2018, 38: 720–742.

[14] Oviedo León H F. A delayed weighted gradient method for strictly convex quadratic minimization[J]. Computational Optimization and Applications, 2019, 74: 729–746.

[15] Burdakov O, Dai Y H, Huang N. Stabilized Barzilai–Borwein method[J]. Journal of Computational Mathematics, 2019, 37: 916–936.

[16] Andreani R, Raydan M. Properties of the delayed weighted gradient method[J]. Computational Optimization and Applications, 2021, 78: 167–180.

[17] Lara H, Aleixo R, Oviedo H. Delayed weighted gradient method with simultaneous step-sizes for strongly convex optimization[J]. Computational Optimization and Applications, 2024, 89: 151–182.

[18] Frassoldati G, Zanni L, Zanghirati G. New adaptive stepsize selections in gradient methods[J]. Journal of Industrial and Management Optimization, 2008, 4(2): 299–312.

[19] Ferrandi G, Hochstenbach M E. Limited memory gradient methods for unconstrained optimization[J]. Numerical Algorithms, 2025, 99: 735–764.

[20] 袁亚湘. 梯度法步长基于子空间逼近的选取 [Z]. 待刊稿, 2026.

[21] Fletcher R, Powell M J D. A rapidly convergent descent method for minimization[J]. Computer Journal, 1963, 6: 163–168.

[22] Toint Ph L. Some numerical results using a sparse matrix updating formula in unconstrained optimization[J]. Mathematics of Computation, 1978, 32: 839–852.

[23] Nocedal J. Updating quasi-Newton matrices with limited storage[J]. Mathematics of Computation, 1980, 35: 773–782.

[24] Dai Y H. A new analysis on the Barzilai–Borwein gradient method[J]. Journal of the Operations Research Society of China, 2013, 1(2): 187–198.

[25] Dai Y H, Fletcher R. On the asymptotic behaviour of some new gradient methods[J]. Mathematical Programming, 2005, 103(3): 541–559.

[26] Dai Y H, Yuan Y X. Analysis of monotone gradient methods[J]. Journal of Industrial and Management Optimization, 2005, 1(2): 181–192.

[27] Yuan Y X. Step-sizes for the gradient method[J]. AMS/IP Studies in Advanced Mathematics, 2008, 42(2): 785–796.

[28] Li D, Sun R. On a faster *R*-linear convergence rate of the Barzilai–Borwein method[J]. arXiv preprint, 2021, arXiv:2101.00205.

[29] Huang Y, Dai Y H, Liu X W, Zhang H. On the asymptotic convergence and acceleration of gradient methods[J]. Journal of Scientific Computing, 2022, 90(1): 7.

[30] Fletcher R. On the Barzilai–Borwein method[M]//Optimization and Control with Applications. Springer, Boston, 2005: 235–256.

[31] Birgin E G, Martínez J M, Raydan M. Nonmonotone spectral projected gradient

methods on convex sets[J]. SIAM Journal on Optimization, 2000, 10(4): 1196–1211.

[32] De Asmundis R, Di Serafino D, Hager W W, Toraldo G, Zhang H. An efficient gradient method using the Yuan steplength[J]. Computational Optimization and Applications, 2014, 59(3): 541–563.

[33] Nocedal J, Sartenaer A, Zhu C. On the behavior of the gradient norm in the steepest descent method[J]. Computational Optimization and Applications, 2002, 22(1): 5–35.

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

[35] Forsythe G E. On the asymptotic directions of the s-dimensional optimum gradient method[J]. Numerische Mathematik, 1968, 11(1): 57–76.

[36] Pronzato L, Wynn H P, Zhigljavsky A A. Asymptotic behaviour of a family of gradient algorithms in $\mathbb{R}^d$ and Hilbert spaces[J]. Mathematical Programming, 2006, 107(3): 409–438.

[37] Bousch T. Le poisson n'a pas d'arêtes[J]. Annales de l'Institut Henri Poincaré: Probabilités et Statistiques, 2000, 36(4): 489–508.

[38] Jenkinson O. Ergodic optimization[J]. Discrete and Continuous Dynamical Systems, 2006, 15(1): 197–224.

[39] Jenkinson O. Ergodic optimization in dynamical systems[J]. Ergodic Theory and Dynamical Systems, 2019, 39(10): 2593–2618.

[40] Garibaldi E. Ergodic Optimization in the Expanding Case[M]. SpringerBriefs in Mathematics, Springer, Cham, 2017.

[41] Walters P. An Introduction to Ergodic Theory[M]. Graduate Texts in Mathematics, Vol. 79, Springer, New York, 1982.

[42] Petersen K. Ergodic Theory[M]. Cambridge Studies in Advanced Mathematics, Vol. 2, Cambridge University Press, Cambridge, 1983.

附录 A　从最小二乘到正规方程：标量与矩阵情形

本附录补充正文第 3 章中最小二乘模型到正规方程的标准推导。虽然这一推导在回归分析中属于基础结论，但在本文中起着连接 “一维 BB” 与 “高维矩阵拟合” 的关键作用，因此在此单独写明。

第一节　标量情形：BB1 与 BB2

对 BB1 而言，考虑标量函数

$$\phi(\eta) = \frac{1}{2}\|\eta s - y\|^2 = \frac{1}{2}(\eta s - y)^T(\eta s - y).$$

把右端展开为

$$\phi(\eta) = \frac{1}{2}\eta^2 s^T s - \eta s^T y + \frac{1}{2}y^T y.$$

对 η 求导得

$$\phi'(\eta) = \eta s^T s - s^T y.$$

令导数为零即可得到正规方程

$$(s^T s)\eta = s^T y,$$

故唯一最小二乘解为

$$\eta = \frac{s^T y}{s^T s}, \qquad \alpha^{\mathrm{BB1}} = \eta^{-1} = \frac{s^T s}{s^T y}.$$

对 BB2 而言，考虑

$$\psi(\alpha) = \frac{1}{2}\|s - \alpha y\|^2 = \frac{1}{2}(s - \alpha y)^T(s - \alpha y) = \frac{1}{2}s^T s - \alpha s^T y + \frac{1}{2}\alpha^2 y^T y.$$

对 α 求导得

$$\psi'(\alpha) = -s^T y + \alpha y^T y.$$

令其为零便得到

$$(y^T y)\alpha = s^T y, \qquad \alpha^{\mathrm{BB2}} = \frac{s^T y}{y^T y}.$$

第二节　矩阵情形：原始 OLS 模型

下面推导正文中的矩阵最小二乘模型

$$M_k^{\mathrm{LS}} = \arg\min_{M\in\mathbb{R}^{p\times p}} \|Y_k - S_k M\|_F^2, \qquad N_k^{\mathrm{LS}} = \arg\min_{N\in\mathbb{R}^{p\times p}} \|S_k - Y_k N\|_F^2.$$

记

$$\Phi(M) = \frac{1}{2}\|Y - SM\|_F^2 = \frac{1}{2}\operatorname{tr}\big((Y - SM)^T(Y - SM)\big).$$

利用迹的循环不变性，把右端展开为

$$\Phi(M) = \frac{1}{2}\operatorname{tr}(Y^TY) - \operatorname{tr}(M^TS^TY) + \frac{1}{2}\operatorname{tr}(M^TS^TSM).$$

对矩阵 M 求导，得到

$$\nabla_M\Phi(M) = S^TSM - S^TY.$$

因此最小二乘驻点满足正规方程

$$(S^TS)M = S^TY.$$

当 S 满列秩时，S^TS 对称正定，从而正规方程有唯一解

$$M = (S^TS)^{-1}S^TY.$$

同理，若记

$$\Psi(N) = \frac{1}{2}\|S - YN\|_F^2 = \frac{1}{2}\operatorname{tr}\big((S - YN)^T(S - YN)\big),$$

则

$$\nabla_N\Psi(N) = Y^TYN - Y^TS,$$

故正规方程为

$$(Y^TY)N = Y^TS.$$

若 Y 满列秩，则其唯一解为

$$N = (Y^TY)^{-1}Y^TS.$$

这说明正文中的原始 OLS 模型并不是凭空给出的定义，而是由 “用一个小矩阵同时拟合最近 p 个割线关系” 这一最小二乘原理直接导出的自然结果。

附录 B　二维对称化公式的代数推导

本附录给出正文中命题 5.2 所需特征多项式的代数展开过程。

对于 2D(S) 型广义特征值问题 $\det(\boldsymbol{C}_k - \theta \boldsymbol{M}_k) = 0$，由正文记号可知

$$\boldsymbol{C}_k - \theta \boldsymbol{M}_k = \begin{pmatrix} u_k - \theta a_k & v_k - \theta b_k \\ v_k - \theta b_k & w_k - \theta c_k \end{pmatrix}.$$

计算其行列式并令其为零，有

$$\begin{aligned} \det(\boldsymbol{C}_k - \theta \boldsymbol{M}_k) &= (u_k - \theta a_k)(w_k - \theta c_k) - (v_k - \theta b_k)^2 \\ &= \left(u_k w_k - \theta(a_k w_k + c_k u_k) + \theta^2 a_k c_k\right) - \left(v_k^2 - 2\theta b_k v_k + \theta^2 b_k^2\right) \\ &= (a_k c_k - b_k^2)\theta^2 - (c_k u_k + a_k w_k - 2 b_k v_k)\theta + (u_k w_k - v_k^2). \end{aligned}$$

代入正文定义的系数：

$$A_k^{(S)} = a_k c_k - b_k^2, \qquad B_k^{(S)} = c_k u_k + a_k w_k - 2 b_k v_k, \qquad \Delta_k = u_k w_k - v_k^2,$$

即可直接得到关于 θ 的一元二次方程：

$$A_k^{(S)} \theta^2 - B_k^{(S)} \theta + \Delta_k = 0.$$

利用求根公式即可解得正文式（45）中的两个广义特征值 $\theta_{k,1}^{(S)}$ 与 $\theta_{k,2}^{(S)}$。

同理，对于 2D(Y) 型广义特征值问题 $\det(\boldsymbol{C}_k - \tau \boldsymbol{D}_k) = 0$，有

$$\boldsymbol{C}_k - \tau \boldsymbol{D}_k = \begin{pmatrix} u_k - \tau d_k & v_k - \tau e_k \\ v_k - \tau e_k & w_k - \tau f_k \end{pmatrix}.$$

计算其行列式展开得到：

$$\begin{aligned} \det(\boldsymbol{C}_k - \tau \boldsymbol{D}_k) &= (u_k - \tau d_k)(w_k - \tau f_k) - (v_k - \tau e_k)^2 \\ &= \left(u_k w_k - \tau(d_k w_k + f_k u_k) + \tau^2 d_k f_k\right) - \left(v_k^2 - 2\tau e_k v_k + \tau^2 e_k^2\right) \\ &= (d_k f_k - e_k^2)\tau^2 - (f_k u_k + d_k w_k - 2 e_k v_k)\tau + (u_k w_k - v_k^2). \end{aligned}$$

代入对应的 Y 型系数：

$$A_k^{(Y)} = d_k f_k - e_k^2, \qquad B_k^{(Y)} = f_k u_k + d_k w_k - 2 e_k v_k,$$

以及 $\Delta_k = u_k w_k - v_k^2$，即可得到关于 τ 的二次方程：

$$A_k^{(Y)} \tau^2 - B_k^{(Y)} \tau + \Delta_k = 0.$$

解之即得正文式（47）中的 $\tau_{k,1}^{(Y)}$ 与 $\tau_{k,2}^{(Y)}$。

对矩阵最小二乘模型进行对称化之后（即将交叉项强制取平均 $v_k = \frac{p_k + q_k}{2}$），我们在计算广义特征值时便规避了非对称矩阵可能带来的复数根风险，并且所有的推导都可以通过简单的二维行列式二次型彻底显式化。

附录 C　三维与四维显式公式：原始 OLS 与对称化模型的统一写法

正文中已经指出，对于 $p \geqslant 3$，从算法实现和理论分析的角度都没有必要把步长彻底写成根式解。不过，若希望给出纯代数意义上的显式公式，则原始 OLS 模型与对称化模型都可以统一地通过特征多项式处理。本附录把“特征多项式系数从何而来”这一中间步骤也写出来。

第一节　系数的来源：迹、主子式与 Newton 恒等式

设 $\mathcal{M} \in \mathbb{R}^{p\times p}$ 的特征值为 $\mu_1, \cdots, \mu_p$。则其特征多项式可写为

$$\chi_p(\lambda) = \det(\lambda I - \mathcal{M}) = \lambda^p - e_1\lambda^{p-1} + e_2\lambda^{p-2} - \cdots + (-1)^p e_p,$$

其中 $e_j = e_j(\mu_1, \cdots, \mu_p)$ 为第 j 个初等对称多项式。另一方面，幂和

$$s_\ell := \mu_1^\ell + \cdots + \mu_p^\ell = \operatorname{tr}(\mathcal{M}^\ell)$$

可由矩阵幂的迹直接计算。Newton 恒等式给出 e_j 与 s_ℓ 之间的关系：

$$e_1 = s_1, \qquad 2e_2 = e_1 s_1 - s_2, \qquad 3e_3 = e_2 s_1 - e_1 s_2 + s_3,$$

以及更高阶的递推关系。因此，只要知道 $\operatorname{tr}(\mathcal{M})$、$\operatorname{tr}(\mathcal{M}^2)$、$\operatorname{tr}(\mathcal{M}^3)$ 等量，就能系统地写出三维、四维特征多项式的系数。

在本文中，$\mathcal{M}$ 可以统一表示为以下四类小矩阵之一：

$$\mathcal{M} = M_k^{-1} B_k \quad (p\mathrm{D}(S)\text{-OLS}), \qquad \mathcal{M} = D_k^{-1} B_k^T \quad (p\mathrm{D}(Y)\text{-OLS}),$$

或

$$\mathcal{M} = M_k^{-1/2} C_k M_k^{-1/2} \quad (p\mathrm{D}(S)\text{-sym}), \qquad \mathcal{M} = D_k^{-1/2} C_k D_k^{-1/2} \quad (p\mathrm{D}(Y)\text{-sym}).$$

一旦 $\mathcal{M}$ 的定义固定，后续特征多项式与显式根式的代数步骤完全相同。

第二节　三维情形

当 $p = 3$ 时，特征多项式为

$$\chi_3(\lambda) = \lambda^3 - c_1\lambda^2 + c_2\lambda - c_3. \tag{88}$$

这里

$$c_1 = e_1 = s_1 = \operatorname{tr}(\mathcal{M}),$$

$$c_2 = e_2 = \frac{1}{2}(s_1^2 - s_2) = \frac{1}{2}\Big((\operatorname{tr}\mathcal{M})^2 - \operatorname{tr}(\mathcal{M}^2)\Big),$$

$$c_3 = e_3 = \frac{1}{6}(s_1^3 - 3s_1s_2 + 2s_3) = \det(\mathcal{M}).$$

令

$$\lambda = z + \frac{c_1}{3}, \qquad p = c_2 - \frac{c_1^2}{3}, \qquad q = -\frac{2c_1^3}{27} + \frac{c_1c_2}{3} - c_3,$$

则（88）化为

$$z^3 + pz + q = 0.$$

若采用 Cardano 公式，记

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3, \qquad u = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}}, \qquad v = \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}},$$

则一个根为 $z_1 = u + v$，另外两个根由乘以复三次单位根得到。在对称化模型中，$\mathcal{M}$ 为实对称矩阵，因此三个根必全为实数。此时更适合使用三角形式

$$z_j = 2\sqrt{-\frac{p}{3}}\cos\left(\frac{1}{3}\arccos\left(\frac{3q}{2p}\sqrt{-\frac{3}{p}}\right) + \frac{2(j-1)\pi}{3}\right), \qquad j = 1, 2, 3.$$

最终特征值为

$$\lambda_j = z_j + \frac{c_1}{3}, \qquad j = 1, 2, 3.$$

对 S 型步长取 $\alpha_j = 1/\lambda_j$，对 Y 型步长则直接取 $\alpha_j = \lambda_j$，即得 3D-BB 步长的显式公式。

第三节　四维情形

当 $p = 4$ 时，特征多项式为

$$\chi_4(\lambda) = \lambda^4 + a_3\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0. \tag{89}$$

若记特征值的初等对称多项式为 e_1, e_2, e_3, e_4，则

$$a_3 = -e_1, \qquad a_2 = e_2, \qquad a_1 = -e_3, \qquad a_0 = e_4.$$

由 Newton 恒等式可得

$$e_1 = s_1 = \operatorname{tr}(\mathcal{M}),$$

$$e_2 = \frac{1}{2}(s_1^2 - s_2) = \frac{1}{2}\Big((\operatorname{tr}\mathcal{M})^2 - \operatorname{tr}(\mathcal{M}^2)\Big),$$

$$e_3 = \frac{1}{6}(s_1^3 - 3s_1s_2 + 2s_3) = \frac{1}{6}\Big((\operatorname{tr}\mathcal{M})^3 - 3\operatorname{tr}(\mathcal{M})\operatorname{tr}(\mathcal{M}^2) + 2\operatorname{tr}(\mathcal{M}^3)\Big),$$

$$e_4 = \det(\mathcal{M}).$$

故（89）中的系数可以写成

$$a_3 = -\operatorname{tr}(\mathcal{M}),$$

$$a_2 = \frac{1}{2}\Big((\operatorname{tr}\mathcal{M})^2 - \operatorname{tr}(\mathcal{M}^2)\Big),$$

$$a_1 = -\frac{1}{6}\Big((\operatorname{tr}\mathcal{M})^3 - 3\operatorname{tr}(\mathcal{M})\operatorname{tr}(\mathcal{M}^2) + 2\operatorname{tr}(\mathcal{M}^3)\Big), \qquad a_0 = \det(\mathcal{M}).$$

令

$$\lambda = z - \frac{a_3}{4},$$

则（89）化为

$$z^4 + pz^2 + qz + r = 0,$$

其中

$$p = a_2 - \frac{3a_3^2}{8}, \qquad q = a_1 - \frac{a_2 a_3}{2} + \frac{a_3^3}{8}, \qquad r = a_0 - \frac{a_1 a_3}{4} + \frac{a_2 a_3^2}{16} - \frac{3a_3^4}{256}.$$

Ferrari 方法表明，可先求解

$$m^3 - \frac{p}{2}m^2 - rm + \frac{4rp - q^2}{8} = 0. \tag{90}$$

设 m 是（90）的任一根，并取

$$R = \sqrt{2m - p},$$

则四个根可写为

$$z_{1,2} = \frac{1}{2}\left(R \pm \sqrt{-2m - p + \frac{2q}{R}}\right), \qquad z_{3,4} = \frac{1}{2}\left(-R \pm \sqrt{-2m - p - \frac{2q}{R}}\right).$$

最后令

$$\lambda_j = z_j - \frac{a_3}{4}, \qquad j = 1, 2, 3, 4.$$

对 S 型与 Y 型步长的处理方式与三维情形完全相同。

需要再次强调的是：附录给出这些显式公式仅仅是为了代数上的完整性。在实际实现中，3D/4D 乃至更高维的步长都应直接通过小规模特征值求解器得到，对原始 OLS 模型尤其如此，因为它在一般非二次情形下可能出现复特征值，而对称化模型则天然避免了这一困难。

附录 D　核心代码片段

本附录摘录论文中各核心算法的 MATLAB 实现片段，便于读者复现实验。

第一节　对称化 S 型步长生成器

下面的函数对应正文中的对称广义特征值问题 $C_k z = \theta M_k z$（算法 2 步骤 3–4）。S, Y 为最近 p 个割线对组成的 $n \times p$ 矩阵，通过对 S 做 rank-revealing QR 自动处理近秩亏，在出现非正广义特征值时直接舍弃，并按算法 2 的步骤 8–9 用参考步长替补。

Listing D.1　对称化 S 型步长生成器

```
function alphas = symmetric_steps_s(S, Y)
    [n, p] = size(S);
    if p == 0, alphas = []; return; end

    % rank-revealing QR on S
    [Q, R] = qr(S, 0);
    diagR = abs(diag(R));
    if isempty(diagR) || diagR(1) == 0, alphas = []; return; end
    keep = diagR >= 1e-8 * diagR(1);
    rk = sum(keep);
    if rk == 0, alphas = []; return; end

    Sb = S(:, 1:rk); Yb = Y(:, 1:rk);
    M = Sb.' * Sb; M = 0.5 * (M + M.'); % S' S
    C = 0.5 * (Sb.' * Yb + Yb.' * Sb); % (S' Y + Y' S)/2
    C = 0.5 * (C + C.');

    % generalized eigenvalues of (C, M)
    ev = real(eig(C, M));
    ev = ev(ev > 0);
    if isempty(ev), alphas = []; return; end
    alphas = sort(1 ./ ev, 'descend');
end
```

第二节　精确 Hessian 投影基线的实现

第七章 §十 中作为强对照基线的精确 Hessian 投影实现：在最近 p 个位移张成的子空间上直接投影 $\nabla^2 f(x_k)$ 并取其特征值倒数。该实现要求对每一块求解一次 Hessian 算子，因而代价显著高于对称化公式。

Listing D.2　精确 Hessian 投影步长生成

```
function alphas = exact_hessian_steps(S, H)
    [Q, R] = qr(S, 0);
    diagR = abs(diag(R));
    if isempty(diagR) || diagR(1) == 0, alphas = []; return; end
    keep = diagR >= 1e-8 * diagR(1);
    rk = sum(keep);
    if rk == 0, alphas = []; return; end
    Qb = Q(:, 1:rk);

    HQ = H * Qb;
    P = Qb.' * HQ;
    P = 0.5 * (P + P.'); % symmetrize

    ev = real(eig(P));
    pos = ev > 0;
    if ~any(pos), alphas = []; return; end
    alphas = sort(1 ./ ev(pos), 'descend');
end
```

在严格凸二次问题上，本片段与 symmetric_steps_s 给出相同步长（第四章 §四 末段的等价性论证）。非二次时二者出现可观测差异：表 7.9 即为对应的实验结果。

第三节　ω-加权 Ritz-sweep 步长

下面的片段对应算法 1 的核心步骤——在 thin QR 后构造两个 $r_k \times r_k$ 对称矩阵并求解广义特征值问题。

Listing D.3　ω-加权 Ritz-sweep 步长（对应算法 1）。

```
function alphas = omega_ritz_steps(G, Hop, omega_kind, varargin)
    % G : n x p history-gradient bundle
    % Hop : function handle x -> H*x (Hessian-vector product)
    % omega_kind : 'one' (LMSD), 'lin' (harmonic-LMSD), or
    % 'pow' with extra arg a for omega(t)=t^a
    [Q, R] = qr(G, 0);
    diagR = abs(diag(R));
    if isempty(diagR) || diagR(1) == 0, alphas = []; return; end
    rk = sum(diagR >= 1e-8 * diagR(1));
    Qb = Q(:, 1:rk);

    HQ = zeros(size(Qb));
    for j = 1:rk, HQ(:,j) = Hop(Qb(:,j)); end

    switch omega_kind
        case 'one' % omega(t) = 1
            A = Qb.' * HQ; B = eye(rk);
        case 'lin' % omega(t) = t
            A = HQ.' * HQ; B = Qb.' * HQ;
        case 'pow' % omega(t) = t^a, a in varargin{1}
            error('Generic omega(H) requires Hop''s spectral access.');
```

```
    end
    A = 0.5 * (A + A.'); B = 0.5 * (B + B.');

    ev = real(eig(A, B));
    ev = ev(ev > 0);
    alphas = sort(1 ./ ev, 'descend');
end
```

第四节　块队列执行器

算法 3（带压缩纯实现）与算法 4（GLL 全局化版本）的共同骨架是“块首生成 r_k 个步长，块内依次执行”，下面的精简实现即此机制。

Listing D.4　块队列执行器

```
function res = run_block_method(obj, grad, hess, x0, p, formula, ...
                                alpha_init, tol, maxiter)
    n = length(x0);
    x = x0; g = grad(x);
    sec_S = zeros(n, 0); sec_Y = zeros(n, 0);

    % --- p warmup steps (constant alpha_init) ---
    for j = 1:p
        x_new = x - alpha_init * g;
        g_new = grad(x_new);
        sec_S = [sec_S, x_new - x];
        sec_Y = [sec_Y, g_new - g];
        x = x_new; g = g_new;
    end

    % --- main loop: block-queue ---
    while obj(x) > tol && size(sec_S, 2) >= 1
        % generate block of step sizes
        if strcmp(formula, 'sym')
            alphas = subspace_bb.symmetric_steps_s(sec_S, sec_Y);
        elseif strcmp(formula, 'exact_hessian')
            alphas = subspace_bb.exact_hessian_steps(sec_S, hess(x));
        end
        if isempty(alphas), alphas = alpha_init; end

        % execute the entire block before recomputing
        for r = 1:length(alphas)
            x_try = x - alphas(r) * g;
            g_try = grad(x_try);
            sec_S = [sec_S, x_try - x];
            sec_Y = [sec_Y, g_try - g];
            if size(sec_S, 2) > p % roll the window
                sec_S = sec_S(:, end-p+1:end);
                sec_Y = sec_Y(:, end-p+1:end);
            end
            x = x_try; g = g_try;
        end
```

```
    end
    res.x = x; res.f = obj(x);
end
```

第五节　GLL 非单调线搜索内核

算法 4 的核心保护机制是 Grippo–Lampariello–Lucidi[1] 非单调接受准则：试探步长按 $\beta^h\tilde{\alpha}$（$h = 0, 1, 2, \cdots$）回溯，直到函数值不超过最近 M 步的最大值减 Armijo 项。

Listing D.5　GLL 非单调线搜索

```
function [xn, gn, fn, ng] = gll_linesearch(f, gradf, xk, gk, fk, ...
                                           fhist, alpha, sigma, beta, M)
    fmax = max(fhist(max(1, end-M+1):end));
    gnorm2 = gk' * gk;
    h = 0; ng = 0;
    while h < 50
        a = alpha * beta^h;
        xt = xk - a * gk;
        ft = f(xt);
        if ft <= fmax - sigma * a * gnorm2
            gt = gradf(xt); ng = ng + 1;
            xn = xt; gn = gt; fn = ft; return;
        end
        h = h + 1;
    end
    xn = []; gn = []; fn = NaN; % failure: caller breaks
end
```

第六节　包络拟合与稳定 $\log\cosh$

以下是对第七章 §十四 的包络拟合与 §十 的 $\log\cosh$ 稳定求值的具体实现。前者先在尾部识别局部峰值，再对峰值的对数做最小二乘线性回归；后者用恒等式

$$\log\cosh(z) = |z| - \log 2 + \log(1 + \mathrm{e}^{-2|z|})$$

避免 $\cosh(z)$ 的上溢。

Listing D.6　包络拟合与稳定 $\log\cosh$ 计算

```
function fit = fit_envelope(loss, peaks, peak_threshold, min_peaks)
    if nargin < 3, peak_threshold = []; end
    if nargin < 4, min_peaks = 2; end

    L = length(loss);
    if isempty(peaks) || length(peaks) < min_peaks
```

```
        idx = (1:L)'; % fall back to raw fit
    else
        idx = peaks(:);
    end
    if ~isempty(peak_threshold) && peak_threshold > 0
        cutoff = floor(peak_threshold * L);
        idx = idx(idx > cutoff);
    end
    k = idx - 1; % zero-based index
    y = log(max(loss(idx), realmin));
    p = polyfit(k, y, 1);
    fit.coeff = p(:);
    fit.r = exp(p(1)); % envelope factor
end

function v = safe_log_cosh(z)
    a = abs(z);
    v = a - log(2) + log1p(exp(-2 * a));
end
```

致谢

春天又来了。

写到这里，我的毕业设计也走到了尾声。在真正落笔感谢之前，我想为这一年的旅途留下一段回望——它太崎岖，也太美好，仅仅写一句“感谢”是不够的。

本科四年里，我的方向一直是概率统计。直到接近保研那段时间，才转向计算数学。去年五月，我通过中国科学院数学与系统科学研究院计算数学方向的推免选拔　袁亚湘院士便给了我一个关于常微分方程与最优化关系的报告，内容大致是陈龙、罗浩等人在强 Lyapunov 框架方面的工作。那时我对计算数学几乎一无所知，对着 PPT 上的每一个符号都要琢磨许久。即便如此，我还是被这个方向迷住了，于是顺着这个题目报了大学生创新项目，题目叫做《强 Lyapunov 框架下优化算法的全局收敛性研究与应用》。九月毕设开题时，最初的题目也是沿着这个方向定下的——《基于常微分方程数值解法的加速一阶优化算法研究》。

大四上学期是松弛的。三年的成绩压力一朝卸下，我把时间慷慨地交给了同学、乐团、和一些意义不大的课程。对科研的态度变得疏淡：兴致来了就读两篇文章，没兴致就搁置。Su 的那篇关于 Nesterov 加速算法的微分方程文章，Wilson 关于加速方法的 Lyapunov 分析，我都泛泛地读过——许多细节没有真正搞清楚，但能从中感受到数学的某种美感，也就够了。

寒假很快过去。下学期我打算在北京做科研，便发邮件问袁老师有什么具体安排。老师让我好好度假并做好毕业设计即可。这让我的二月与三月异常自由：本应去中关村的数学与系统科学研究院，但因没有具体任务，我或者去国科大玉泉路的图书馆自习，或者干脆躺在家里。那段时间，我第一次真切地感受到科研态度从兴趣到困难的转折。毕设题目远比我想象的难——它需要太多优化与微分方程的前置知识，独自摸索的效率很低。我读了不少文章，做了不少 MATLAB 实验，写出了一份一百多页的 LaTeX 文档，但越写越觉得空洞。那些工作或许已足以应付本科毕业的标准，可我希望做出真正属于自己的、哪怕原理朴素的、前人未曾做过的东西。

转折发生在三月中下旬。袁老师两会结束后，给我发来一封邮件，附上他近期一篇关于梯度法步长的工作。文章把 BB 步长基于子空间投影的视角推广到了二维与三维，给出显式公式，证明了一些性质。他让我做些数值实验，并思考可能的推广。

那是一个很普通的傍晚。我对着邮件里的文章看了一下午，等到晚上才隐约觉察到些什么——这些公式可以放进一个广义最小二乘的框架里，把一维 BB 与

二维、三维公式统一起来，进一步推广到任意维数。更妙的是，推广后的公式不再依赖任何显式 Hessian 信息，这意味着它可以平滑地推广到一般非线性问题。那个晚上我在书桌前坐到很晚，第一次有了“也许这件事真的可以做”的念头。

从那之后，研究像一条暗河，悄悄在脑子里冲刷着河道。我读了大量 BB 方法的文献，了解 GLL 非单调线搜索之类的全局化技术，也读出了袁老师证明里的一些细节问题——这是后来我读 Fletcher 论文时也产生类似怀疑的源头。与此同时，我做了一件后来看起来朴素却带来全部好运的事：我不再单步分析，而是把每 p 步状态打包成一个整体看待。打包之后，整个块到下一个块的递推映射，立刻显出了两条性质——连续，正齐次。这两条放在一起，几乎就是一个离散动力系统的天然形状。我把此前从微分方程那里学到的“全局渐近稳定蕴含指数稳定”这一类经典结果迁移到这个离散场景，“逐点收敛蕴含 R-线性收敛”的雏形便在草稿纸上慢慢成形了。也正是同样的视角，让方法从二次问题向一般非线性问题的过渡变得自然——只需要一些基础的分析与线性代数工具，就足够把整条路走通。

我用了大约十天时间，把所有想法整理成一篇几十页的小文章——一些证明、一些猜想、一些数值实验、一些向袁老师提出的问题。

就在我准备发邮件的那天，我无意间搜到了 Fletcher 关于 LMSD 的经典论文。

那一刻是惊愕的。原来类似的思想，早在十多年前就已被人系统地提出过。后来又陆续读到 Curtis–Guo 关于 LMSD 的 R-收敛性证明，读到 Ferrandi 与 Hochstenbach 在 2025 年的最新工作——那些文章不仅做过我做过的事，还发表在我当时仰望的顶级期刊上。一瞬间所有的兴奋都掉了下来。我把整理好的稿子连同这些文献一并发给袁老师，那天的邮件，是带着一些失落的。

但读完 Fletcher 的原始证明，我又生出新的疑问。他论证中关于单个 Ritz 值连续性的步骤，与我此前看出的问题非常相似。我反复琢磨，依然不认为那个证明是正确的，可我也举不出反例否定算法本身的收敛性。一边是发表在顶级期刊上的经典结果，一边是自己反复推敲都过不去的几行论证——这样的处境让人不安。

三月底我回了趟合肥。和科大的同学玩了几天，参加了“溯风”文化月的开幕式，与乐团的同学一起排练。那几天我几乎不去想数学，只把自己交给音乐。可一回到北京，那些悬而未决的问题又重新压回来。我当时几乎认为自己唯一剩下的有价值的工作，就是从逐点收敛到 R-线性收敛的这条提升机制——而如果 Fletcher 的证明确实是错的，那么这条机制连前提都不复存在，我也只是“重新”证明了一遍 BB 方法的 R-线性收敛而已。

清明节假期的某一天，我在书桌前翻着 Raydan 1993 年那篇 BB 收敛性证明

的旧文。那是一个稀疏而温暖的下午，窗外有风，一切看起来再普通不过。可就在那一刻——我忽然意识到，只要把块乘子写成行列式之比，再用 Cauchy–Binet 展开一次，那些项立刻就组合出 Raydan 证明里那种 “加权平均” 的结构。

我把笔停下，盯着草稿纸看了很久。

那一瞬间的心情是科研里独有的——那种确认自己可能真的触到一个东西的颤抖。我反复检查推导的每一步，确认它是对的。那一晚没怎么睡，眼前一直浮现出整篇论文的结构。

接下来的日子像被按下了快进键。我把 “块映射” 视角进一步打磨，意识到它本质上就是一种离散动力系统结构，而这恰好与我那段时间正在自学的遍历理论相通。在阅读文献的过程中，我注意到 BB 方法 R-收敛速率的最紧上界，至今仍是一个开放问题：Dai–Liao 的证明过程大致给出 $(1-1/\kappa)^{1/n}$ 这一粗略的界，2021 年又被改进到 $1-1/\kappa$。而 Dai 在 2013 年关于 BB 方法的一篇分析文章中提到过一个二维的、唯一不超 R-线性收敛的轨道——其速率恰是 $(\kappa-1)/(\kappa+1)$。我先在有限周期轨道上看出这就是确切的下界，再借助块映射视角与遍历理论的一些基础工具，把它推广到了一般情形，似乎完整地证明了这个最紧上界。整个论证里只有一处技术性的不等式构造稍稍费了点心思——其余都顺得不可思议，甚至让我自己都不太敢相信。这个工作让我相信，“块映射” 这一离散动力系统的视角，必然有更广的应用前景。

四月的国家大剧院上演 “国乐之春” 系列。我几乎每天去看，每个周日去中关村与国科大的乐团合排，沉浸在民族音乐里。那时我确信自己已经把 LMSD 的 R-收敛证完了，整理论文虽然繁琐，但有一种沾沾自喜的愉悦。我也偷空逛了北京动物园、国家图书馆——那些一直想去却被无尽的论文拖住的地方。北京的春天就在这样一种愉悦里慢慢铺展开。

然后我得了一场严重的流感。

那两天几乎无法工作。可就在我稍稍好转、迫不及待想要继续整理稿子的时候，我意识到——块状态空间太大了，许多状态是真实算法根本不可达的。“连续性” 这一假设也比我以为的要脆弱：替代步长稍一引入，连续性立刻破碎。这意味着我的框架可能连最简单的 BB 方法的 R-收敛都证不出来。

那一刻，所有的兴奋与得意从头顶倾泻而下。

我尝试用 “闭锥” 这一更小的状态空间替代全空间，至少把 BB 方法的证明 gap 修补完整。但对行为更复杂的 LMSD，连续性在历史束秩退化时的崩坏，与 Fletcher 证明里的那处问题，竟在某种深层意义上同根同源——而 Curtis–Guo 那种历史束非渐近退化的强假设，几乎就是为绕开这一点而存在的。

四月中下旬，我不得不暂时妥协。我也加上了一些非退化条件，把算法 “勉强” 塞进自己的框架，对一般非线性的部分则做了相当多的额外假设，把整理好

的稿子发给袁老师，并决心把毕设方向正式转向这一年的研究。我又回了一趟合肥，找同学走了走。那一阵子，无条件的 LMSD 收敛性已经成了一朵搁浅在心上的云。

四月底，我打算把 “修复 Fletcher 证明” 这一小段单独整理一篇短文投给学院的学生刊物《蛙鸣》。就是在重读自己过去的草稿时，我忽然撞见一个再朴素不过的事实：历史束秩退化只可能在活跃谱的数量小于历史窗口长度时发生，而那种情形下算法是必然有限终止的。

——退化的情形是平凡的。

这个再简单不过的观察，把整片乌云一推就推开了。我只需要在退化的边界上证一下连续性，就可以把无条件的 R-收敛性收回来。而修复 Fletcher 证明所用到的工具，又能进一步抽象为一条更一般的 “乘子权重引理”，由此衍生出一族 ω 加权 Ritz-sweep 方法。本文中的 ω 加权族，便是这样从一篇短文里意外结出的果实。

有心栽花花不开，无心插柳柳成荫——我从未对这句话有过这样具体的体会。

五一来了，所有的间隙都被填上。北京的天气从将熄的春走入将燃的夏。剩下的便是论文的整理与排版。这段时间我也没让自己太累——和同学逛朝阳公园，去新中关，看汤尤杯的赛程在激动与失落里一场一场推进，回来再补几页文字。最后这一切汇聚成现在这本论文时，我回望这一年的反复与挣扎，竟觉得它最终的样子真是太好了。

在本文成稿的同时，我也以本文相关工作为基础，整理出了四篇英文文章，均已基本成稿，后续将继续打磨后陆续投出。它们的题目与各自的内容简述如下：

(1) *A determinantal multiplier framework for limited-memory steepest descent and weighted Ritz-sweep gradient methods.* 对应本文第 3 章与第 4 章前半。把每个 Ritz-sweep 的块乘子写成行列式之比，借助 Cauchy–Binet 展开把它表示为非负加权平均，从而修正 Fletcher 2012 的原始证明，并把 LMSD、harmonic-LMSD 统一推广为一族 ω-加权 Ritz 方法。

(2) *A closed-cone homogeneous mapping framework and unconditional R-linear convergence of fixed-memory spectral gradient methods.* 对应本文第 2 章与第 4 章后半。在闭锥状态空间上证明 “连续正齐次 + 逐点收敛” 自动给出统一有限步收缩，从而对一族固定记忆谱方法（包含 BB1/BB2、LMSD、harmonic-LMSD 以及 ω-加权 Ritz-sweep）给出无附加历史束非退化假设的 R-线性收敛。

(3) *Symmetrized fixed-memory spectral gradient methods with rank compression and GLL globalization.* 对应本文第 5–6 章与第 7 章后半。系统讨论对称化步长生

成器、秩压缩纯实现与 GLL 全局化 pD(S)-BB 算法，并给出一般非线性问题局部 R-线性收敛的主定理及对应数值实验。

(4) *A termination-state framework for the classical Barzilai–Borwein method: sharp asymptotic rate on strictly convex quadratics.* 对应本文 §三。借助终止态紧化与端点恒等式，证明经典 BB1 方法在严格凸二次问题上的紧渐近速率为 $(\kappa-1)/(\kappa+1)$，并给出与初值一致的 $c+\varepsilon$ 型 R-线性估计，在数量上改进 Li–Sun 2021 的 $1-1/\kappa$ 上界。

此外，一篇综述 *A unified spectral state-space view of limited-memory steepest descent and the Barzilai–Borwein method* 整理了三条可迁移的原理：加权端点不等式、闭锥状态空间与齐次提升。

回望整个过程，最要感谢的是我的研究生导师袁亚湘院士。是他在合适的时机递来那篇关于 BB 步长子空间推广的工作，让我真正叩开了这片研究领域的门。我们之间的交流并不算多，但每一次邮件往来里他寥寥几句的反馈，都足以让我反复琢磨许久——那是一种在远处亮着的灯光，让我知道方向不致走偏。这一年的所有奇遇与转折，都源自最初他点的那一盏灯。

也要感谢校内导师陈士祥老师。从开题、中期审核到论文的写作意见，他一直在我身后默默把关。有他在，我才能放心地走得这样远。

要感谢我的爸爸妈妈。在北京的最后这段日子里，科研有时像一座始终翻不过去的山，几天几天卡在同一个地方，我会变得焦躁、脾气不好。可他们从未怪过我，反而带我去吃好吃的，去看电影，督促我作息规律，守着家里那种朴素的节奏。我后来才慢慢明白，我所有突如其来的灵感，都不是凭空降下的，而是被这种最日常的、稳稳托住的爱缓缓托上来的。

要感谢一路上陪伴的同学。有人与我讨论数学，有人与我闲聊度日，有人和我一起在春天的北京到处乱走。这些看起来与论文无关的时刻，其实是支撑论文得以写完的真正底色。

要感谢中国科学技术大学数学科学学院的老师们。班主任许小卫老师在学习和生活上都对我百般照顾；做乐珏老师《线性代数 B1/B2》助教的那一年，让我对线性代数这门看似熟悉的工具有了真正深入的理解，而这份理解几乎渗透进了本论文的每一行推导；陈小伍老师在每次期中座谈会上都极其认真地回应同学的问题；行政老师们在我每一个需要帮助的瞬间都给予我无条件的支持；其他每一位曾教过我的老师身上都体现着科大那种朴素的、对教学的郑重。这些点点滴滴汇成了我数学的底子，也汇成了我看待世界的方式。

最后，我要把这一段单独留给中国科学技术大学学生民族乐团——NMOU。

本科四年，我在 NMOU 也整整四年。声部长、团支书、教学负责人、团长——一个一个角色这样走过来，一周一次的排练，一学期一次的演出，太多的曲

子，太多的故事，太多的起起落落，欢笑与泪水。在那些迷茫与焦灼的夜晚，是和团员们一起奏响的弦歌把我从焦虑里一点一点拉回来；在那些自我怀疑的时刻，是合奏里某一个齐落的瞬间让我重新相信美与善是真实的。如今我也即将卸任团长，把这把火传递给下一届。四年的大学生活本身就像一首不断变奏的乐曲，激昂处激昂，沉静处沉静——而 NMOU，便是这首乐曲里贯穿始终的那条主线。

谢谢你们每一个人。

最后，本文部分文字曾借助生成式人工智能工具进行润色与表述上的调整，论文的全部研究内容、数学论证、数值实验与思想观点均由本人独立完成。当 AI 的便利已经悄然融入这个时代每一个写作者的日常，我也在此向这种新的工具诚实致谢——它让我得以把更多的时间留给真正需要思考的部分。

春天还在继续。这本论文便在此搁笔。

愿一切热爱都能被珍重，愿所有的暗河终将汇入光中。

2026 年 5 月