Global and unconditional *R*-linear convergence of rank-compressed weighted LMSD sweeps for strictly convex quadratics

SHUTAI YANG[†]
State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
School of Mathematical Sciences, University of Science and Technology of China, Hefei 230026, China

AND

YA-XIANG YUAN[‡]
State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[†]Corresponding author: yangshutai26@mails.ucas.ac.cn
[‡]Email: yyx@lsec.cc.ac.cn

We study limited memory steepest descent (LMSD) with exact algebraic rank compression for strictly convex quadratic optimization. At each cycle, the method restricts the gradient history to its column space and applies the reciprocals of all Ritz values of the compressed projection in the next sweep. If the block-start gradient has at most p active distinct eigenvalues, the delayed sweep terminates finitely. For nonterminating trajectories, a determinant and Cauchy–Binet formula for the complete-sweep polynomial yields global convergence without a full-rank history or a run-wise normalized-conditioning bound. Consecutive sweep endpoints define a continuous positively homogeneous map on a closed cone of compatible states. Compactness then gives an *R*-linear endpoint estimate, uniform over compatible initial states, with constants depending only on H and p; the estimate extends to inner gradients, iterate errors, and objective gaps. Every fixed positive spectral weight $W = \omega(H)$ is linearly conjugate to the standard method. The weighted methods therefore share its decay factor, while the Euclidean-norm prefactor is increased by at most $\sqrt{\kappa_2(W)}$. This includes harmonic-Ritz LMSD, fixed power weights, and the delayed BB1 and BB2 recurrences.

Keywords: limited memory steepest descent; Ritz values; Krylov rank loss; algebraic rank compression; spectral conjugacy; finite termination; positively homogeneous dynamics; *R*-linear convergence.

Mathematics Subject Classification: 65K05, 90C20, 65F10, 65F15.

1. Introduction

Consider

$$\min_{x\in\mathbb{R}^n} f(x), \qquad f(x) = \tfrac{1}{2}x^\top Hx - b^\top x, \qquad H = H^\top \succ 0. \tag{1.1}$$

For a gradient step $x^+ = x - \alpha g$, the gradient satisfies $g^+ = (I - \alpha H)g$. Spectral gradient methods exploit this polynomial action by selecting reciprocal stepsizes from spectral information about H. The two classical Barzilai–Borwein (BB) stepsizes were introduced in Barzilai & Borwein (1988). For strictly convex quadratics, Raydan proved global convergence for the first BB choice in arbitrary dimension and observed that the same argument applies to the second (Raydan, 1993). Dai and Liao subsequently established an *R*-linear rate for the first choice and noted the analogous result for the second

(Dai & Liao, 2002). Friedlander *et al.* proved global convergence for a finite-delay family of retarded Rayleigh-quotient steps (Friedlander *et al.*, 1999). For a systematic spectral interpretation and numerical comparison of BB-type, LMSD and related gradient steplengths, see di Serafino *et al.* (2018); a broader survey is given in Zou & Magoulès (2022).

Fletcher's limited memory steepest descent method (LMSD) stores the pre-step gradients generated in a sweep, extracts Ritz values from their Krylov span and applies the reciprocal values in the following sweep, with the Ritz values ordered from largest to smallest (Fletcher, 2012, Sections 2–3). In the appendix of that paper, Fletcher stated that the basic fixed-length standard-Ritz sweep on a strictly convex quadratic either terminates or has gradients converging to zero; Section 7 introduced the corresponding harmonic-Ritz variant.

Curtis and Guo subsequently established an R-linear rate for any fixed history length for the standard method and, in their appendix, for the harmonic variant (Curtis & Guo, 2018). Their analysis is carried out under Assumption 3.4. In particular, if $G_k = Q_k R_k$, then G_k is required to have full column rank and there must exist a constant $\rho \geq 1$, uniform in the cycle index k along the run, such that

$$\left\|R_k^{-1}\right\| \leq \rho \left\|g_{k,1}\right\|^{-1}. \tag{1.2}$$

Accordingly, their rate constants depend on a run-wise normalized-conditioning bound.

Fletcher (2012, Section 3) and, later, Curtis & Guo (2018, Remark 2.2) noted that loss of rank or severe history ill-conditioning may require fewer stored gradients in a finite-precision implementation. Di Serafino *et al.* form the projected matrix through a Cholesky factorization of $G^\top G$ and discard the oldest stored gradient when that factorization loses numerical definiteness (di Serafino *et al.*, 2018, Section 2.2). The analyses of Fletcher and of Curtis and Guo use fixed full-rank histories. Gu and Du's modified LMSD (MLMSD) instead selects one safeguarded Ritz candidate from the largest full-rank suffix of the available history (Gu & Du, 2021). Ferrandi and Hochstenbach study Cholesky, pivoted QR factorization and singular value decomposition (SVD) realizations, numerical rank truncation, harmonic variants and non-linear extensions (Ferrandi & Hochstenbach, 2025).

Alignment-based and cyclic-reuse extensions of LMSD were explored by Zou and Magoulès (Zou & Magoulès, 2019). Ferrandi, Hochstenbach and Krejić developed a target-dependent harmonic framework for selecting BB-type spectral steplengths in gradient methods (Ferrandi *et al.*, 2023). Zhou, Gu and Yuan studied the relationships among determinant-pencil, Krylov–Ritz and finite-moment constructions under explicit structural assumptions and proposed a rebound-aware spectral gradient method (RASG) that selects and reuses one admissible candidate before refreshing the reduced model (Zhou *et al.*, 2026, Sections 2 and 4). The present paper considers a different update rule: every value of the algebraically compressed pencil is applied once in the following sweep, including across changes in the algebraic history rank.

We analyse the exact-arithmetic standard-Ritz complete-sweep dynamics on the algebraic column space of each history. A determinant representation of the complete-sweep polynomial gives pointwise global convergence without the full-rank and normalized-conditioning assumptions in (1.2). A compatible two-endpoint formulation then yields an R-linear estimate uniform over compatible initial states. Fixed positive spectral weights are handled by the change of variables $W^{1/2}$, which preserves the Hessian and the extracted stepsizes.

For the one-step BB recurrences, a closed-cone convergence analysis and the fixed-weight conjugacy are given in Yang & Yuan (2026). The present paper treats complete LMSD sweeps and the continuity issues caused by Krylov-rank loss.

The main results are as follows.

(i) We formulate standard-Ritz LMSD with exact algebraic rank compression. If a completed history starts from a gradient with at most p active distinct eigenvalues, the compressed projection recovers the active spectrum and the following delayed sweep terminates finitely. This treats the nominal full-rank and rank-deficient termination strata within the same algorithmic rule.

(ii) A three-dimensional example directly disproves the fixed-length Ritz lower-bound implication used to obtain equation (36) in Fletcher's appendix. Thus the published proof is incomplete at that step as written unless an additional orbit-specific non-degeneracy argument is supplied. This motivates our use of complete-sweep multipliers rather than individual Ritz roots.

(iii) For non-terminating standard sweeps, a determinant ratio and its Cauchy–Binet expansion represent the complete-sweep multiplier through fixed spectral templates with non-negative state-dependent weights. This representation yields pointwise global convergence across changes in Krylov rank.

(iv) Consecutive sweep endpoints define a continuous positively homogeneous self-map on a closed cone of compatible states, with zero continuation on the finite-termination strata. For fixed H and p, compactness upgrades pointwise convergence to an R-linear endpoint estimate uniform over compatible initial states. The estimate extends to inner gradients, iterate errors and objective gaps.

(v) Every fixed positive spectral weight $W = \omega(H)$ is linearly conjugate to the standard endpoint map. Hence all such weights admit the standard decay factor, while the Euclidean-norm prefactor is bounded by the standard prefactor times $\sqrt{\kappa_2(W)}$. This covers harmonic-Ritz LMSD, fixed power weights and the delayed BB1 and BB2 recurrences.

Numerical diagnostics compare fixed power weights on a repeated-spectrum quadratic and record the normalized history-conditioning quantity along standard and harmonic-Ritz trajectories.

Proposition 2.5 identifies the finite-termination strata, while Proposition 4.2 supplies the determinant identity used in Theorem 5.1 to prove pointwise global convergence. Continuity of the compatible endpoint map in Theorem 8.4, combined with the homogeneous compactness principle in Theorem 7.1, yields the state-uniform endpoint rate in Theorem 9.2. For an actual run, the rate is initialized at $z_0 = (g_{1,0}, g_{2,0})$, the start/end pair of the first completed sweep after the warm-up; arbitrary positive finite warm-up steps enter through this initial state, whereas a bound relative to the original gradient would also depend on the warm-up stepsizes. Proposition 9.3 and Corollaries 9.4 and 9.5 transfer the same decay factor to every fixed positive spectral weight and give the corresponding bounds for inner gradients, iterate errors and objective gaps.

2. Rank-compressed LMSD sweeps and spectral conjugacy

2.1. *Active spectral reduction*

Translate the minimizer to the origin, so that

$$f(x) = \frac{1}{2} x^\top H x, \qquad g = Hx, \qquad H = H^\top \succ 0. \tag{2.1}$$

A gradient step along the negative-gradient direction satisfies

$$g^+ = (I - \alpha H) g. \tag{2.2}$$

Consequently every generated gradient is a polynomial in H applied to the initial gradient. Write the spectral decomposition by distinct eigenvalues as

$$H = \sum_{\ell=1}^{s} \mu_\ell P_\ell, \qquad 0 < \mu_1 < \cdots < \mu_s. \tag{2.3}$$

LEMMA 2.1 *Set $u_\ell = P_\ell g_0$ in (2.3). For every generated gradient g_j there is a scalar $c_{j,\ell}$ such that*

$$P_\ell g_j = c_{j,\ell} u_\ell. \tag{2.4}$$

Hence the trajectory lies in

$$\mathscr{A}_0 = \operatorname{span}\{u_\ell : u_\ell \neq 0\}.$$

Using the normalized non-zero vectors u_ℓ as a basis of $\mathscr{A}_0$, the restriction of H is diagonal and the eigenvalues of $H|_{\mathscr{A}_0}$ are distinct.

Proof If $g_j = p_j(H)g_0$, then

$$P_\ell g_j = P_\ell p_j(H) g_0 = p_j(\mu_\ell) P_\ell g_0.$$

This proves (2.4) and the asserted active-coordinate reduction. □

We may therefore work in active spectral coordinates and assume

$$H = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \qquad 0 < \lambda_1 < \cdots < \lambda_n, \tag{2.5}$$

where n is the number of active distinct eigenvalues. This is the standard active-coordinate reduction used in quadratic spectral-gradient and LMSD analyses; see Fletcher (2012, Section 2).

2.2. *History blocks and the delayed LMSD sweep*

A block of ℓ gradient steps with positive finite stepsizes is written

$$g_{j+1} = (I - \alpha_j H) g_j, \qquad j = 0, \ldots, \ell - 1, \tag{2.6}$$

and its pre-step history is

$$G = [g_0, \ldots, g_{\ell-1}]. \tag{2.7}$$

PROPOSITION 2.2 *For every block (2.6),*

$$\operatorname{span}(G) = \mathscr{K}_\ell(H, g_0) := \operatorname{span}\{g_0, H g_0, \ldots, H^{\ell-1} g_0\}. \tag{2.8}$$

Thus the span depends only on the block-start gradient and the block length.

Proof For each j,

$$g_j = p_j(H) g_0, \qquad p_j(t) = \prod_{h=0}^{j-1} (1 - \alpha_h t).$$

The polynomial p_j has degree j and non-zero leading coefficient. Hence the change-of-basis matrix between $[g_0, H g_0, \ldots, H^{\ell-1} g_0]$ and G is triangular with non-zero diagonal. □

For the algorithmic indexing, write $x_{k,j}$ and $g_{k,j} = Hx_{k,j}$ for the jth pre-step state of block k. If block k has length ℓ_k, then

$$G_k = [g_{k,0}, \dots, g_{k,\ell_k-1}], \qquad x_{k,\ell_k} = x_{k+1,0}, \qquad g_{k,\ell_k} = g_{k+1,0}. \tag{2.9}$$

The warm-up block has $\ell_0 = p$ unless it terminates early; block $k+1$ has length $\ell_{k+1} = d_k$. Given a history G_k, let Q_k have orthonormal columns spanning $\operatorname{span}(G_k)$ and set

$$d_k = \operatorname{rank}(G_k) \le p. \tag{2.10}$$

The standard method extracts the Ritz values

$$Q_k^\top H Q_k v = \theta v. \tag{2.11}$$

They are positive and are used one block later as reciprocal stepsizes. We list them in non-increasing order,

$$\theta_{k,1} \ge \cdots \ge \theta_{k,d_k} > 0. \tag{2.12}$$

This is the conventional LMSD ordering used by Fletcher (2012, Section 3). The endpoint of a completed sweep is independent of this ordering, although the intermediate iterates are not.

Algorithm 2.1 Standard LMSD sweep with algebraic rank compression

Require: $H \succ 0$, an initial point and nominal history length p.

Generate p warm-up gradient steps with positive finite stepsizes, storing their pre-step gradients in G_0, unless $g = 0$ is reached earlier.
if $g = 0$ **then**
  return the minimizer.
end if
for $k = 0, 1, 2, \dots$ **do**
  Compute an orthonormal basis Q_k of $\operatorname{span}(G_k)$ and set $d_k = \operatorname{rank}(G_k)$.
  Solve (2.11) and order the values as in (2.12).
  Starting from the current gradient, apply the stepsizes $\theta_{k,1}^{-1}, \dots, \theta_{k,d_k}^{-1}$ until either termination or completion of the sweep, storing the pre-step gradients as G_{k+1}.
  if $g = 0$ **then**
    return the minimizer.
  end if
end for

Let $g_{k,0}$ denote the first gradient in history block G_k. On any non-terminating tail for which $d_k = p$, the block-start recurrence is

$$g_{k+2,0} = \psi_k(H) g_{k+1,0}, \qquad \psi_k(t) = \prod_{j=1}^{p} \left(1 - \frac{t}{\theta_{k,j}}\right). \tag{2.13}$$

2.3. *Fixed spectral weights and conjugacy*

A *positive spectral weight* is a matrix of the form

$$W = \sum_{\ell=1}^{s} w_\ell P_\ell = \omega(H), \qquad w_\ell > 0. \tag{2.14}$$

Thus $W \succ 0$, W commutes with H and $WH \succ 0$. We write

$$\kappa_2(W) := \|W\|_2 \left\|W^{-1}\right\|_2$$

for its spectral condition number. In the active coordinates of (2.5), the weight is diagonal. The fixed-weight variant uses

$$Q_k^\top W H Q_k v = \theta\, Q_k^\top W Q_k v. \tag{2.15}$$

PROPOSITION 2.3 *Let $W = \omega(H) \succ 0$, set $S = W^{1/2}$ and define*

$$\widetilde{x}_j = S x_j, \qquad \widetilde{g}_j = S g_j, \qquad \widetilde{G}_k = S G_k.$$

Then $\widetilde{g}_j = H\widetilde{x}_j$, and every gradient step is conjugated with the same stepsize:

$$\widetilde{g}_{j+1} = (I - \alpha_j H)\widetilde{g}_j. \tag{2.16}$$

For every full-column-rank basis B of a history space,

$$(SB)^\top H (SB) = B^\top W H B, \qquad (SB)^\top (SB) = B^\top W B. \tag{2.17}$$

Consequently the generalized Ritz values of (2.15) on $\operatorname{span}(G_k)$ *are exactly the standard Ritz values of H on* $\operatorname{span}(\widetilde{G}_k)$. *Moreover,*

$$\operatorname{rank}(\widetilde{G}_k) = \operatorname{rank}(G_k),$$

and g_j and $\widetilde{g}_j$ have the same active spectral support. Hence the complete fixed-weight rank-compressed LMSD trajectory is the pullback under S^{-1} of the standard rank-compressed trajectory starting from $\widetilde{x}_0 = S x_0$, with identical warm-up stepsizes, sweep lengths, extracted values and finite-termination indices.

Proof Since S and H are functions of the same symmetric matrix, they commute. Thus $\widetilde{g}_j = SHx_j = HSx_j = H\widetilde{x}_j$, and multiplying the gradient recurrence by S gives (2.16). The same commutation relation and $S^2 = W$ give (2.17). Generalized eigenvalues are invariant under a change of basis within a fixed subspace, so the two projected problems have the same values. The matrix S is invertible, which preserves history ranks; its restriction to every eigenspace of H is multiplication by the positive scalar $\sqrt{w_\ell}$, which preserves active spectral support. The asserted trajectory correspondence now follows block by block. □

Thus it is enough to prove the convergence theory for the standard method; Section 9 transfers the resulting bounds back to every fixed positive spectral weight with an explicit norm-equivalence factor.

The weighted pencil (2.15) includes the two principal LMSD variants:

$$\begin{aligned} W = I \quad &: \quad Q_k^\top H Q_k v = \theta v, && \text{standard Ritz LMSD,} \\ W = H \quad &: \quad Q_k^\top H^2 Q_k v = \theta Q_k^\top H Q_k v, && \text{harmonic-Ritz LMSD.} \end{aligned} \tag{2.18}$$

The algorithm applies the reciprocal $\alpha = \theta^{-1}$ as the stepsize. For a one-dimensional history, the reciprocal values in the two cases are the quadratic BB1 and BB2 stepsizes, respectively. For $p > 1$, these are the standard and harmonic LMSD pencils studied in Fletcher (2012); Curtis & Guo (2018); Ferrandi & Hochstenbach (2025).

More generally, because $H \succ 0$,

$$W = H^a, \qquad a \in \mathbb{R}, \tag{2.19}$$

is positive definite for every real exponent. The projected problem becomes

$$Q_k^\top H^{a+1} Q_k v = \theta\, Q_k^\top H^a Q_k v. \tag{2.20}$$

Thus $a = 0$ is standard LMSD and $a = 1$ is harmonic-Ritz LMSD, while other values of a define a power-weighted family.

2.4. *Support, Krylov rank and finite termination*

For $u = (u_1, \ldots, u_n)^\top$, define

$$S(u) := \{i : u_i \neq 0\}, \qquad s(u) := |S(u)|. \tag{2.21}$$

LEMMA 2.4 *Under (2.5),*

$$\operatorname{rank}[u, Hu, \ldots, H^{p-1}u] = \min\{p, s(u)\}. \tag{2.22}$$

Proof Delete the zero rows and factor out the non-zero components of u. The remaining matrix is a rectangular Vandermonde matrix with distinct nodes λ_i, and its rank is $\min\{p, s(u)\}$. □

Lemma 2.4 separates three cases:

$$s(u) > p, \qquad s(u) = p, \qquad s(u) < p.$$

The first is the regular branch. In the second case the history has rank p but already spans the active invariant subspace; in the third its rank is $s(u) < p$. Both latter cases lead to finite termination by the next proposition.

PROPOSITION 2.5 *Let*

$$G = [g_0, \ldots, g_{p-1}]$$

be a completed p-step history block with non-zero block-start gradient $g := g_0$ and block endpoint $h := g_p$. Suppose that $d = s(g) \leq p$ (hence $1 \leq d \leq p$). Then

$$\operatorname{span}(G) = \mathscr{K}_p(H, g) = \operatorname{span}\{e_i : i \in S(g)\}. \tag{2.23}$$

The eigenvalues of the standard projected matrix on $\operatorname{span}(G)$ *are precisely $\{\lambda_i : i \in S(g)\}$. The effective standard sweep generated from this history and applied to h annihilates h in at most d steps.*

Proof Proposition 2.2 gives $\operatorname{span}(G) = \mathscr{K}_p(H,g)$. By Lemma 2.4, this space has dimension d. The rightmost space in (2.23) is a d-dimensional H-invariant subspace containing $\mathscr{K}_p(H,g)$, so (2.23) holds.

On this active invariant subspace, the compressed standard projection is the restriction of H. Therefore its eigenvalues are exactly $\{\lambda_i : i \in S(g)\}$. The endpoint of the history block has the form

$$h = p_p(H)g, \qquad p_p(t) = \prod_{j=0}^{p-1}(1-\alpha_j t),$$

so

$$S(h) \subseteq S(g). \tag{2.24}$$

Define

$$\zeta_g(t) := \prod_{i\in S(g)} \left(1 - \frac{t}{\lambda_i}\right).$$

For each spectral coordinate r, either $h_r = 0$ or $r \in S(g)$, in which case $\zeta_g(\lambda_r) = 0$. Consequently

$$\zeta_g(H)h = \prod_{i\in S(g)} \left(I - \lambda_i^{-1} H\right) h = 0.$$

Thus the values extracted from the history starting at g annihilate the endpoint h to which the delayed sweep is applied. The order of the d factors is immaterial for the endpoint, and termination may occur before all d factors have been used. □

COROLLARY 2.6 *Proposition 2.5 holds unchanged for every fixed positive spectral weight $W = \omega(H)$: the weighted compressed pencil recovers the same active eigenvalues and the effective delayed sweep terminates in at most d steps.*

Proof Apply Proposition 2.3. The transformation by $W^{1/2}$ preserves active spectral support and history rank, while the weighted pencil has exactly the standard Ritz values of the transformed history. □

COROLLARY 2.7 *On every non-terminating standard run, each completed history has rank p and each block-start gradient satisfies*

$$s(g_{k,0}) > p.$$

Consequently all blocks after the warm-up have length p, and (2.13) holds throughout the non-terminating tail.

Proof Suppose, to the contrary, that k is the first index with $d_k < p$. Since the run is non-terminating, G_0 is a completed p-step warm-up history. If $k > 0$, minimality gives $d_{k-1} = p$, so G_k was also generated by a completed p-step sweep. Hence Proposition 2.2 and Lemma 2.4 give

$$d_k = \min\{p, s(g_{k,0})\} < p.$$

Proposition 2.5 then terminates the following sweep, a contradiction. Thus $d_k = p$ for every k.

If some block-start gradient satisfied $s(g_{k,0}) = p$, the same proposition would again terminate the following sweep. Therefore $s(g_{k,0}) > p$ on every non-terminating run. □

REMARK 2.8 If $p \geq n$ in the reduced problem, every non-zero block-start gradient has at most p active eigenvalues, so Proposition 2.5 implies finite termination. In particular, the total number of gradient steps is at most $p+n$.

Although every finite history on a non-terminating run has rank p, normalized histories may approach lower-support boundary states. This issue is examined in Section 3.

3. Fixed-length Ritz continuity at Krylov-rank loss

At a rank-changing limit, convergence of normalized block-start gradients does not by itself control all Ritz values of a projected problem whose prescribed dimension exceeds the limiting Krylov rank. The following example isolates this difficulty for the lower bound used in Fletcher's equation (36).

PROPOSITION 3.1 *Let*

$$H = \operatorname{diag}(1,2,3), \qquad p = 2, \qquad q_\varepsilon := \frac{(\varepsilon, \varepsilon^2, 1)^\top}{\sqrt{1+\varepsilon^2+\varepsilon^4}}, \qquad \varepsilon > 0. \tag{3.1}$$

Let $\theta_{\varepsilon,1} \geq \theta_{\varepsilon,2}$ be the two standard Ritz values of H on $\mathscr{K}_2(H, q_\varepsilon)$. Then

$$q_\varepsilon \to e_3, \qquad \mathscr{K}_2(H, q_\varepsilon) \longrightarrow \operatorname{span}\{e_1, e_3\}, \qquad (\theta_{\varepsilon,1}, \theta_{\varepsilon,2}) \to (3,1). \tag{3.2}$$

Here the subspace convergence is in the operator norm of the corresponding orthogonal projectors. Consequently, for all sufficiently small ε,

$$\theta_{\varepsilon,2} < 2 = \frac{2}{3}\lambda_3. \tag{3.3}$$

The same pair limit $(3,1)$ holds for the generalized Ritz values associated with every fixed positive spectral weight $W = \omega(H)$.

Proof Clearly $q_\varepsilon \to e_3$. Set

$$r_\varepsilon := \frac{(H-3I)q_\varepsilon}{\|(H-3I)q_\varepsilon\|} = \frac{(-2, -\varepsilon, 0)^\top}{\sqrt{4+\varepsilon^2}}.$$

Then $r_\varepsilon \in \mathscr{K}_2(H, q_\varepsilon)$ and $r_\varepsilon \to -e_1$. Moreover $q_\varepsilon^\top r_\varepsilon \to 0$. Orthogonalizing r_ε against q_ε therefore gives a unit vector $\widehat{r}_\varepsilon \in \mathscr{K}_2(H, q_\varepsilon)$ with $\widehat{r}_\varepsilon \to -e_1$. Hence

$$Q_\varepsilon := [q_\varepsilon, \widehat{r}_\varepsilon] \longrightarrow [e_3, -e_1]$$

is a convergent orthonormal basis, which proves the subspace limit in (3.2). It follows that

$$Q_\varepsilon^\top H Q_\varepsilon \longrightarrow \operatorname{diag}(3,1),$$

so the ordered Ritz values tend to $(3,1)$ and (3.3) follows.

For a fixed positive spectral weight $W = \operatorname{diag}(w_1, w_2, w_3)$,

$$Q_\varepsilon^\top W H Q_\varepsilon \longrightarrow \operatorname{diag}(3w_3, w_1), \qquad Q_\varepsilon^\top W Q_\varepsilon \longrightarrow \operatorname{diag}(w_3, w_1).$$

The limiting generalized eigenvalues are again 3 and 1. Equivalently, the orthogonal projectors onto $\mathscr{K}_2(H, q_\varepsilon)$ converge in operator norm to the projector onto $\operatorname{span}\{e_1, e_3\}$. □

Relation to Fletcher's equation (36) In Fletcher's appendix, let r denote the spectral-component induction index (denoted p there) and let m denote the sweep length. Along an infinite subsequence, the first $r-1$ components of the normalized sweep-start gradients tend to zero, so every accumulation point is supported on $\{r,\dots,n\}$. The proof then invokes continuity of eigenvalues to conclude that the m prelimit Ritz values eventually satisfy

$$\theta_{k,\ell} \in (2\lambda_r/3, \lambda_n), \qquad \ell = 1,\dots,m,$$

which is equation (36) of Fletcher (2012, Appendix).

The difficulty is that the limiting m-step Krylov problem may have rank below m. Proposition 3.1, with $r=3$ and $m=2$, gives a path for which the normalized block-start gradient tends to e_3, whereas the smaller prelimit Ritz value tends to $1 < 2 = (2/3)\lambda_3$. Hence the normalized-gradient convergence and finite-iteration full-rank information used at this point in the appendix do not by themselves imply equation (36). Recovering that conclusion would require an additional orbit-specific constant-rank or quantitative non-degeneracy argument that excludes such rank-changing limiting paths.

Our convergence proof avoids the issue by working with the complete-sweep multiplier rather than an ordered Ritz tuple. The next section derives determinant and Cauchy–Binet representations that remain suitable at rank-changing limits.

4. Determinant representation of a complete sweep

For each completed block,

$$g_{k+1,0} = q_k(H) g_{k,0}$$

for some polynomial q_k, and hence

$$S(g_{k+1,0}) \subseteq S(g_{k,0}).$$

The support therefore stabilizes after finitely many blocks. On a non-terminating run its cardinality exceeds p by Corollary 2.7. Restricting to the resulting invariant subspace and relabeling the eigenvalues, we may analyse the tail under

$$n > p, \qquad g_{k,0}^{(i)} \neq 0 \quad (i = 1,\dots,n), \qquad k \geq k_0. \tag{4.1}$$

4.1. *Spectral enclosure*

LEMMA 4.1 *Let $Q \in \mathbb{R}^{n\times d}$ have orthonormal columns. Then every eigenvalue of the standard projected matrix $Q^\top H Q$ lies in*

$$[\lambda_1, \lambda_n]. \tag{4.2}$$

Proof This follows directly from the min–max principle applied to the restriction of H to $\operatorname{span}(Q)$. □

4.2. *The complete-sweep multiplier*

On a regular block, write $g_k := g_{k,0}$ and define

$$K_k = [g_k, Hg_k, \dots, H^{p-1}g_k] = D_k V, \tag{4.3}$$

where

$$D_k = \operatorname{diag}(g_k^{(1)}, \dots, g_k^{(n)}), \qquad V_{ij} = \lambda_i^{j-1}. \tag{4.4}$$

Let $\mathscr{I}_p$ denote the family of all p-element subsets $I \subset \{1, \dots, n\}$. For $I \in \mathscr{I}_p$, let V_I be the corresponding $p \times p$ row submatrix of V and set $\Delta_I = \det(V_I)$; under (2.5) every Δ_I is a Vandermonde determinant with distinct nodes and is therefore non-zero.

PROPOSITION 4.2 *For every $t \in \mathbb{R}$, the complete-sweep multiplier satisfies*

$$\psi_k(t) = \frac{\det\big(K_k^\top (H - tI) K_k\big)}{\det\big(K_k^\top H K_k\big)}. \tag{4.5}$$

For a spectral index r, define

$$\beta_I^{(k)} = \Delta_I^2 \prod_{i \in I} \lambda_i \big(g_k^{(i)}\big)^2 > 0, \qquad \phi_I^{(r)} = \prod_{i \in I} \left(1 - \frac{\lambda_r}{\lambda_i}\right), \tag{4.6}$$

and

$$\pi_I^{(k)} = \frac{\beta_I^{(k)}}{\sum_{J \in \mathscr{I}_p} \beta_J^{(k)}}. \tag{4.7}$$

Then

$$\psi_k(\lambda_r) = \sum_{I \in \mathscr{I}_p} \pi_I^{(k)} \phi_I^{(r)}, \qquad \pi_I^{(k)} > 0, \qquad \sum_{I \in \mathscr{I}_p} \pi_I^{(k)} = 1. \tag{4.8}$$

Thus $\psi_k(\lambda_r)$ is a convex combination of the fixed real numbers $\phi_I^{(r)}$.

Proof In the regular regime, K_k has full column rank. Since $H \succ 0$, the matrix $K_k^\top H K_k$ is positive definite, so the denominator in (4.5) is non-zero. The generalized eigenvalues are unchanged when the orthonormal basis Q_k is replaced by the full-rank basis K_k of the same subspace. Hence

$$\det\Big(K_k^\top H K_k - \lambda K_k^\top K_k\Big) = \det(K_k^\top K_k) \prod_{j=1}^{p} (\theta_{k,j} - \lambda).$$

Evaluating at $\lambda = t$ and at $\lambda = 0$ and dividing gives (4.5), because $\prod_j (\theta_{k,j} - t)/\prod_j \theta_{k,j} = \psi_k(t)$. Specializing to $t = \lambda_r$, substituting $K_k = D_k V$ and applying Cauchy–Binet to the denominator gives

$$\det(K_k^\top H K_k) = \sum_{I \in \mathscr{I}_p} \beta_I^{(k)}. \tag{4.9}$$

The same algebraic Cauchy–Binet expansion for the numerator gives

$$\det\Big(K_k^\top (H - \lambda_r I) K_k\Big) = \sum_{I \in \mathscr{I}_p} \beta_I^{(k)} \phi_I^{(r)}. \tag{4.10}$$

Dividing (4.10) by (4.9) yields (4.8). □

For completeness, we record the tail form of the scalar lemma used in Fletcher's appendix (Fletcher, 2012).

LEMMA 4.3 *Let $k_0 \geq 0$ and let $\{a_k\}_{k\geq k_0}$ be a non-negative sequence. Suppose there are constants $M \geq 1$ and $c \in (0,1)$ such that*

$$a_{k+2} \leq M a_{k+1} \qquad (k \geq k_0). \tag{4.11}$$

Assume further that for every $\varepsilon > 0$ there is $k_\varepsilon \geq k_0$ such that

$$k \geq k_\varepsilon, \quad a_k \geq \varepsilon \quad \Longrightarrow \quad a_{k+2} \leq c a_{k+1}. \tag{4.12}$$

Then $a_k \to 0$.

Proof Fix $\delta > 0$. We first show that there exists $N \geq k_\delta + 1$ such that $a_N < \delta$. Otherwise $a_k \geq \delta$ for every $k \geq k_\delta + 1$, and (4.12) would give

$$a_{k+2} \leq c a_{k+1} \qquad (k \geq k_\delta + 1).$$

Equivalently, $a_{j+1} \leq c a_j$ for every sufficiently large j, which forces $a_j \to 0$ and contradicts $a_j \geq \delta$.

Fix such an N. We claim that

$$a_j < M^2 \delta \qquad \text{for every } j \geq N.$$

For $j = N$ this follows from $a_N < \delta$ and $M \geq 1$. Moreover, since $N - 1 \geq k_0$, (4.11) gives

$$a_{N+1} \leq M a_N < M\delta \leq M^2 \delta.$$

Assume inductively that the claim holds through index m, where $m \geq N + 1$. If $a_{m-1} < \delta$, then two applications of (4.11) give

$$a_{m+1} \leq M a_m \leq M^2 a_{m-1} < M^2 \delta.$$

If $a_{m-1} \geq \delta$, then $m - 1 \geq k_\delta$, and (4.12) gives

$$a_{m+1} \leq c a_m < a_m < M^2 \delta.$$

The induction is complete. Hence

$$\limsup_{k\to\infty} a_k \leq M^2 \delta.$$

Letting $\delta \downarrow 0$ proves that $a_k \to 0$. □

We now combine Proposition 4.2 and Lemma 4.3 in a spectral induction.

5. Global convergence

THEOREM 5.1 *Consider Algorithm 2.1 for (2.1) with algebraic rank compression. The method either reaches $g = 0$ in finitely many steps or, if it does not terminate,*

$$g_{k,0} \longrightarrow 0, \qquad \max_{0\le j\le p} \|g_{k,j}\| \longrightarrow 0, \qquad \max_{0\le j\le p} \|x_{k,j}\| \longrightarrow 0. \tag{5.1}$$

In the original variables of (1.1), before the translation and active reduction of Section 2, the conclusion states that every iterate converges to the unique minimizer of f.

Proof Assume that the method does not terminate. By Corollary 2.7, every post-warm-up block has length p and every block-start gradient has more than p active components. We therefore work on the regular tail (4.1).

Set

$$B := \max\left\{1, \frac{\lambda_n}{\lambda_1} - 1\right\}.$$

By Lemma 4.1, every extracted value lies in $[\lambda_1, \lambda_n]$. Therefore, for every r and k,

$$|\psi_k(\lambda_r)| \le B^p. \tag{5.2}$$

We prove componentwise convergence by induction on r. For $r = 1$, every set I containing 1 has $\phi_I^{(1)} = 0$. If $1 \notin I$, then

$$0 < \phi_I^{(1)} \le \left(1 - \frac{\lambda_1}{\lambda_n}\right)^p =: \rho_1 < 1.$$

The convex-weight identity (4.8) gives

$$|\psi_k(\lambda_1)| \le \rho_1, \tag{5.3}$$

so (2.13) implies $g_{k,0}^{(1)} \to 0$.

Let $r \ge 2$ and assume $g_{k,0}^{(i)} \to 0$ for every $i < r$. We verify the hypotheses of Lemma 4.3 for the sequence

$$a_k = \left|g_{k,0}^{(r)}\right|, \qquad k \ge k_0,$$

with k_0 the start of the regular tail (4.1). The recurrence (2.13) gives

$$a_{k+2} = |\psi_k(\lambda_r)|\, a_{k+1}.$$

Thus (5.2) yields the two-step bound (4.11) with $M = B^p \ge 1$.

It remains to verify the trigger implication (4.12). Fix an arbitrary $\varepsilon > 0$ and put

$$S_\varepsilon = \{k \ge k_0 : \left|g_{k,0}^{(r)}\right| \ge \varepsilon\}. \tag{5.4}$$

If S_ε is finite, the trigger condition is immediate after its last element. Assume that S_ε is infinite. Partition $\mathscr{I}_p$ into

$$A_r = \{I : r \in I\}, \tag{5.5}$$

$$B_r = \{I : I \subset \{r+1, \dots, n\}\}, \tag{5.6}$$

$$C_r = \mathscr{I}_p \setminus (A_r \cup B_r). \tag{5.7}$$

For $I \in A_r$, $\phi_I^{(r)} = 0$. For $I \in B_r$,

$$0 < \phi_I^{(r)} \le \left(1 - \frac{\lambda_r}{\lambda_n}\right)^p =: \rho_r < 1, \tag{5.8}$$

with $\rho_r = 0$ if B_r is empty.

Every $I \in C_r$ omits r and contains at least one index $i(I) < r$. Define

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

By (4.6),

$$\frac{\beta_I^{(k)}}{\beta_{J(I)}^{(k)}} = c_I \frac{(g_{k,0}^{(i(I))})^2}{(g_{k,0}^{(r)})^2}, \tag{5.9}$$

where

$$c_I = \frac{\Delta_I^2}{\Delta_{J(I)}^2} \frac{\lambda_{i(I)}}{\lambda_r} > 0$$

is independent of k. Since the denominator in (4.7) is at least $\beta_{J(I)}^{(k)}$, for $k \in S_\varepsilon$,

$$\pi_I^{(k)} \le \frac{\beta_I^{(k)}}{\beta_{J(I)}^{(k)}} \le \frac{c_I}{\varepsilon^2} (g_{k,0}^{(i(I))})^2 \longrightarrow 0 \qquad (k \in S_\varepsilon,\ k \to \infty), \tag{5.10}$$

by the induction hypothesis. There are only finitely many sets in C_r, and every $\phi_I^{(r)}$ is a fixed finite scalar. For $k \in S_\varepsilon$,

$$|\psi_k(\lambda_r)| \le \rho_r \sum_{I \in B_r} \pi_I^{(k)} + \sum_{I \in C_r} \pi_I^{(k)} \left|\phi_I^{(r)}\right| \le \rho_r + \sum_{I \in C_r} \pi_I^{(k)} \left|\phi_I^{(r)}\right|.$$

Equations (5.10) and (4.8) therefore imply

$$\limsup_{\substack{k \to \infty \\ k \in S_\varepsilon}} |\psi_k(\lambda_r)| \le \rho_r < 1. \tag{5.11}$$

Choose the fixed contraction factor

$$c_r := \frac{1 + \rho_r}{2} \in (0, 1).$$

For every $\varepsilon > 0$, (5.11) yields an index k_ε such that

$$\left|g_{k+2,0}^{(r)}\right| \le c_r \left|g_{k+1,0}^{(r)}\right| \qquad \text{for every } k \in S_\varepsilon \text{ with } k \ge k_\varepsilon, \tag{5.12}$$

while for $k \notin S_\varepsilon$ the premise of (4.12) is not met. The factor $c = c_r$ therefore verifies (4.12) for every $\varepsilon > 0$, with k_ε depending on ε. All hypotheses of Lemma 4.3 are satisfied, and the lemma yields $a_k \to 0$, that is, $g_{k,0}^{(r)} \to 0$. The induction is complete, so $g_{k,0} \to 0$.

Finally, each individual factor $I - H/\theta$ has spectral norm at most B, and a non-terminating block has p steps. Therefore

$$\max_{0\le j\le p} \|g_{k,j}\| \le B^p \|g_{k,0}\| \longrightarrow 0.$$

Since $x_{k,j} = H^{-1} g_{k,j}$, the iterate convergence in (5.1) follows. Undoing the translation and the active reduction of Section 2 restores the statement in the original variables. □

COROLLARY 5.2 *For every fixed positive spectral weight $W = \omega(H)$, the rank-compressed weighted LMSD method defined by (2.15) has the same finite-termination alternative and pointwise convergence conclusions as Theorem 5.1.*

Proof Proposition 2.3 identifies the transformed weighted trajectory with a standard trajectory for the same Hessian. Apply Theorem 5.1 and then use the invertibility of $W^{1/2}$. □

We now formulate the standard sweep dynamics as a first-order map on consecutive endpoints.

6. Endpoint-state formulation

To write the delayed recurrence as a first-order system, we define the compatible endpoint pairs before spectral reduction. For a symmetric positive definite matrix H, let

$$\mathscr{Q}_p(H) := \left\{ q(t) = \prod_{j=1}^{p} (1 - \alpha_j t) : \alpha_j \in [\lambda_{\max}(H)^{-1}, \lambda_{\min}(H)^{-1}] \right\}, \tag{6.1}$$

and define

$$\Omega_p(H) := \{(u,v) : v = q(H)u \text{ for some } q \in \mathscr{Q}_p(H)\}. \tag{6.2}$$

We use the product norm

$$\|(u,v)\|_\times := \left(\|u\|^2 + \|v\|^2\right)^{1/2}. \tag{6.3}$$

This cone contains every consecutive endpoint pair of a non-terminating sweep and serves as the state space for the rate analysis.

For $(u,v) \in \Omega_p(H)$, compatibility gives $v = q(H)u$ for some $q \in \mathscr{Q}_p(H)$. Hence both endpoints and all subsequent polynomial updates lie in

$$E(u) = \operatorname{span}\{P_\ell u : P_\ell u \neq 0\}.$$

Identifying the normalized non-zero vectors $P_\ell u$ with coordinate vectors reduces H isometrically to a diagonal matrix with distinct eigenvalues, while preserving projected matrices, endpoint updates and product norms. We therefore work in the coordinates (2.5), set

$$K_p(u) = [u, Hu, \ldots, H^{p-1}u],$$

retain the support notation $S(u)$ and $s(u)$ from Section 2, and abbreviate

$$\mathscr{Q}_p = \mathscr{Q}_p(H), \qquad \Omega_p = \Omega_p(H).$$

The cases $s(u) > p$, $s(u) = p$ and $s(u) < p$ correspond respectively to the regular branch, the full-rank termination boundary and the rank-deficient termination boundary.

If $s(u) \geq p$, Lemma 2.4 gives $\operatorname{rank} K_p(u) = p$. The generalized eigenvalues of

$$K_p(u)^\top H K_p(u) z = \theta K_p(u)^\top K_p(u) z \tag{6.4}$$

are therefore well defined; denote them by $\theta_1(u), \ldots, \theta_p(u)$ and set

$$\Psi_u(t) := \prod_{j=1}^{p} \left(1 - \frac{t}{\theta_j(u)}\right). \tag{6.5}$$

Because H is symmetric positive definite and $K_p(u)$ has full column rank,

$$K_p(u)^\top H K_p(u) \succ 0.$$

Consequently,

$$\Psi_u(t) = \frac{\det\big(K_p(u)^\top (H - tI) K_p(u)\big)}{\det(K_p(u)^\top H K_p(u))} \tag{6.6}$$

has a strictly positive denominator. If $s(u) = p$, the columns of $K_p(u)$ span the active invariant subspace, so the values in (6.4) are exactly $\{\lambda_i : i \in S(u)\}$ and

$$\Psi_u(H) v = 0 \qquad \text{for every } (u, v) \in \Omega_p \text{ with } s(u) = p. \tag{6.7}$$

Here compatibility gives $S(v) \subseteq S(u)$. When $s(u) < p$, the p-column Krylov matrix is rank deficient; exact compression returns the $s(u)$ active eigenvalues and the following effective sweep terminates.

Define the endpoint output by

$$F_p : \Omega_p \longrightarrow \mathbb{R}^n, \qquad F_p(u, v) := \begin{cases} \Psi_u(H) v, & s(u) > p, \\ 0, & s(u) \leq p. \end{cases} \tag{6.8}$$

When $s(u) = p$, the projected values are the active eigenvalues, so $\Psi_u(H) v = 0$. In particular, $F_p(0, 0) = 0$. Define the transition

$$T_p : \Omega_p \longrightarrow \mathbb{R}^n \times \mathbb{R}^n, \qquad T_p(u, v) := \big(v, F_p(u, v)\big). \tag{6.9}$$

Section 8 shows that T_p is a continuous self-map of $\Omega_p(H)$.

7. A compactness principle for homogeneous dynamics

Let X be a finite-dimensional normed space. A non-empty closed set $\Omega \subset X$ is a closed cone if $0 \in \Omega$ and $t\Omega \subseteq \Omega$ for every $t \geq 0$. Related stability results for homogeneous discrete-time systems are discussed in, for example (Shen & Hu, 2012). The following statement is the implication (i)⇒(ii) of Yang & Yuan (2026, Theorem 2.2). We include its proof to keep the LMSD analysis self-contained.

THEOREM 7.1 *Let $\Omega \subset X$ be a closed cone and let $T : \Omega \to \Omega$ be continuous and positively homogeneous of degree one:*

$$T(tx) = tT(x), \qquad t \geq 0, \quad x \in \Omega. \tag{7.1}$$

If $T^k(x) \to 0$ for every $x \in \Omega$, then there exist $C \geq 1$ and $\rho \in (0, 1)$ such that

$$\left\| T^k(x) \right\| \leq C \rho^k \|x\|, \qquad x \in \Omega, \quad k \geq 0. \tag{7.2}$$

Proof If $\Omega = \{0\}$, the conclusion is immediate. Otherwise let

$$\Sigma := \{x \in \Omega : \|x\| = 1\}.$$

The set Σ is non-empty and compact. For each $x \in \Sigma$, pointwise convergence gives an integer $n_x \geq 1$ such that $\|T^{n_x}(x)\| < 1/2$. By continuity of T^{n_x}, there is a relative neighborhood U_x of x in Σ on which

$$\|T^{n_x}(y)\| \leq \tfrac{1}{2}.$$

Choose a finite subcover $U_{x_1}, \ldots, U_{x_r}$ and set

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

Thus every $y \in \Sigma$ has an integer $m(y) \in \{1, \ldots, N_0\}$ for which $\left\|T^{m(y)}(y)\right\| \leq 1/2$. Continuity and compactness also give the finite transient bound

$$L := \max_{0 \leq j \leq N_0} \max_{y \in \Sigma} \left\|T^j(y)\right\| < \infty. \tag{7.3}$$

Fix $x \in \Omega \setminus \{0\}$. Set $z_0 = x$ and $s_0 = 0$. Whenever $z_q \neq 0$, choose

$$m_q := m(z_q / \left\|z_q\right\|) \in \{1, \ldots, N_0\}, \qquad s_{q+1} := s_q + m_q, \qquad z_{q+1} := T^{m_q}(z_q).$$

Every iterate of T is positively homogeneous, so

$$\left\|z_{q+1}\right\| \leq \tfrac{1}{2} \left\|z_q\right\| \quad \text{and hence} \quad \left\|z_q\right\| \leq 2^{-q} \|x\|$$

for every constructed stage. If $z_{q_*} = 0$ for the first time, then $T(0) = 0$ by (7.1), and $T^k(x) = 0$ for every $k \geq s_{q_*}$. For $k < s_{q_*}$ there is a unique $q < q_*$ with $s_q \leq k < s_{q+1}$. If no zero occurs, then $s_q \geq q \to \infty$, so the same unique q exists for every $k \geq 0$. In either case, whenever $T^k(x)$ has not already vanished, positive homogeneity and (7.3) give

$$\left\|T^k(x)\right\| = \left\|T^{k-s_q}(z_q)\right\| \leq L \left\|z_q\right\| \leq L 2^{-q} \|x\|.$$

Since $k < s_{q+1} \leq (q+1)N_0$, one has $q > k/N_0 - 1$. Therefore, also at the zero iterates,

$$\left\|T^k(x)\right\| \leq 2L\left(2^{-1/N_0}\right)^k \|x\|.$$

Equation (7.2) follows with $C = \max\{1, 2L\}$ and $\rho = 2^{-1/N_0}$. □

8. Compatible endpoint cone and boundary continuity

8.1. *Cone properties and the endpoint self-map*

LEMMA 8.1 *The set Ω_p is a closed cone. If $(u, v) \in \Omega_p$, then*

$$S(v) \subseteq S(u). \tag{8.1}$$

Proof Cone invariance follows from $v = q(H)u$. Let $(u^\ell, v^\ell) \to (u, v)$ with $(u^\ell, v^\ell) \in \Omega_p$. Choose

$$v^\ell = q_\ell(H)u^\ell, \qquad q_\ell(t) = \prod_{j=1}^{p}(1 - \alpha_j^\ell t),$$

where the parameter vectors belong to the compact box $[\lambda_n^{-1}, \lambda_1^{-1}]^p$. A subsequence satisfies $\alpha_j^\ell \to \alpha_j$ for every j. The corresponding polynomials converge at every spectral point to a polynomial $q \in \mathscr{Q}_p$, and passage to the limit gives $v = q(H)u$. Thus $(u, v) \in \Omega_p$. Finally, $v_i = q(\lambda_i)u_i$ gives (8.1). □

PROPOSITION 8.2 *The transition T_p defined in (6.9) maps Ω_p into itself. For every $(u, v) \in \Omega_p$ and every $t \geq 0$,*

$$F_p(tu, tv) = tF_p(u, v), \qquad T_p(tu, tv) = tT_p(u, v). \tag{8.2}$$

Proof Let $(u, v) \in \Omega_p$. If $s(u) > p$, Lemma 4.1 places every $\theta_j(u)$ in $[\lambda_1, \lambda_n]$. Hence the reciprocal roots in (6.5) lie in $[\lambda_n^{-1}, \lambda_1^{-1}]$, so $\Psi_u \in \mathscr{Q}_p$ and

$$\big(v, F_p(u, v)\big) = \big(v, \Psi_u(H)v\big) \in \Omega_p.$$

If $s(u) \leq p$, then (8.1) gives $s(v) \leq p$. If $v = 0$, any polynomial in $\mathscr{Q}_p$ annihilates it. If $v \neq 0$, choose factors $1 - t/\lambda_i$ for all $i \in S(v)$ and fill the remaining positions, if any, by repeating an arbitrary legal factor. This produces a polynomial in $\mathscr{Q}_p$ that annihilates v. Since $F_p(u, v) = 0$ on both finite-termination strata, $(v, 0) \in \Omega_p$. Thus T_p is a self-map.

For $t > 0$, one has $S(tu) = S(u)$ and $K_p(tu) = tK_p(u)$. On the regular stratum the two matrices in (6.4) are both multiplied by t^2, so the generalized eigenvalues and Ψ_u are unchanged. On the two finite-termination strata the output is zero. Hence $F_p(tu, tv) = tF_p(u, v)$ for $t > 0$. At $t = 0$, the state is $(0, 0)$ and the same identity follows from the definition. The identity for T_p follows componentwise. □

8.2. *Cauchy–Binet coordinates*

If $p \geq n$, then $s(u) \leq n \leq p$ for every state, so $F_p \equiv 0$ and continuity is immediate. The formulas below are needed for possible regular states, which necessarily satisfy $n > p$.

Let $V \in \mathbb{R}^{n \times p}$ be the Vandermonde matrix

$$V_{ij} = \lambda_i^{j-1}, \qquad 1 \leq i \leq n, \quad 1 \leq j \leq p. \tag{8.3}$$

For a p-element set $I \subset \{1, \ldots, n\}$, let V_I be the corresponding square row submatrix and set

$$\Delta_I := \det(V_I) \neq 0. \tag{8.4}$$

Write $\mathscr{I}_p$ for all p-element index sets. For $r \in \{1, \ldots, n\}$, define

$$\beta_I(u) := \Delta_I^2 \prod_{i \in I} \lambda_i u_i^2, \qquad \phi_I^{(r)} := \prod_{i \in I}\left(1 - \frac{\lambda_r}{\lambda_i}\right). \tag{8.5}$$

LEMMA 8.3 *If $s(u) > p$, then for every spectral index r,*

$$\Psi_u(\lambda_r) = \frac{\sum_{I \in \mathscr{I}_p} \beta_I(u)\phi_I^{(r)}}{\sum_{I \in \mathscr{I}_p} \beta_I(u)}. \tag{8.6}$$

The normalized coefficients are non-negative and sum to one.

Proof Let $S = S(u)$. Since $|S| > p$, the active reduction indexed by S is in the regular regime of Proposition 4.2: the reduced Krylov matrix has full column rank. Applying (4.8) on that active subspace gives (8.6) with both sums restricted to the p-element subsets of S. If $I \not\subseteq S$, then $\beta_I(u) = 0$, so extending the sums to all of $\mathscr{I}_p$ changes neither numerator nor denominator. At least one p-element subset of S exists, and every factor in its $\beta_I(u)$ is positive. The denominator is therefore strictly positive. □

8.3. *Continuity across the finite-termination strata*

THEOREM 8.4 *The maps $F_p : \Omega_p \to \mathbb{R}^n$ and $T_p : \Omega_p \to \Omega_p$ are continuous. In particular, the transition is continuous at the full-rank finite-termination boundary $s(u) = p$ and at every rank-changing boundary $s(u) < p$.*

Proof Let $(u^\ell, v^\ell) \to (u, v)$ in Ω_p, and set $y^\ell := F_p(u^\ell, v^\ell)$ and $y := F_p(u, v)$. Since the first component of T_p is v, it is enough to prove $y^\ell \to y$. Partition the indices into

$$I_+ := \{\ell : s(u^\ell) > p\}, \qquad I_- := \{\ell : s(u^\ell) \le p\}. \tag{8.7}$$

For $\ell \in I_-$, the definition gives $y^\ell = 0$. To prove convergence of the full sequence, it is enough in each case to prove that every infinite restricted subsequence indexed by I_+ or I_- converges to the same limit y; a finite index class is irrelevant.

Case 1: $s(u) > p$. Choose $p+1$ non-zero coordinates of u. They remain non-zero for all sufficiently large ℓ, so eventually $\ell \in I_+$. The matrix $K_p(u)$ has full column rank and

$$K_p(u)^\top H K_p(u) \succ 0.$$

The determinants in (6.6) depend continuously on u, and the limiting denominator is positive. Hence, for every spectral coordinate i,

$$\Psi_{u^\ell}(\lambda_i) \longrightarrow \Psi_u(\lambda_i). \tag{8.8}$$

It follows coordinatewise that $y^\ell \to \Psi_u(H)v = y$.

Case 2: $s(u) = p$. This is a full-column-rank finite-termination boundary, and $y = 0$ by (6.8). The subsequence indexed by I_- is identically zero. On an infinite I_+ subsequence, the determinant argument from Case 1 remains valid because $K_p(u)$ still has rank p and its limiting denominator is positive. For

$r \in S(u)$, (6.7) gives

$$\Psi_u(\lambda_r) = 0,$$

so (8.8) and $v_r^\ell \to v_r$ imply $y_r^\ell \to 0$. If $i \notin S(u)$, then $v_i = 0$ by Lemma 8.1. Lemma 4.1 gives the uniform bound

$$|\Psi_{u^\ell}(\lambda_i)| \le B_p, \qquad B_p := \max\left\{1, \frac{\lambda_n}{\lambda_1} - 1\right\}^p, \tag{8.9}$$

for $\ell \in I_+$, and therefore $y_i^\ell \to 0$. Both restricted subsequences converge to zero, hence so does the original sequence.

Case 3: $1 \le s(u) < p$. Again $y = 0$ and the I_- subsequence is identically zero. Consider an infinite I_+ subsequence. For $i \notin S(u)$, (8.9) and $v_i^\ell \to 0$ give $y_i^\ell \to 0$. Fix $r \in S(u)$. We prove

$$\Psi_{u^\ell}(\lambda_r) \longrightarrow 0 \qquad (\ell \in I_+). \tag{8.10}$$

Every set $I \in \mathscr{I}_p$ containing r has $\phi_I^{(r)} = 0$. Let $I \in \mathscr{I}_p$ omit r. Since $|S(u)| < p$, the set I contains an index $h \notin S(u)$. Put

$$J := (I \setminus \{h\}) \cup \{r\}. \tag{8.11}$$

For sufficiently large ℓ, $u_r^\ell \neq 0$. If $\beta_I(u^\ell) = 0$, its normalized contribution in (8.6) is zero. Otherwise every coordinate indexed by I is non-zero, so $\beta_J(u^\ell) > 0$, and

$$\frac{\beta_I(u^\ell)}{\beta_J(u^\ell)} = \frac{\Delta_I^2}{\Delta_J^2}\frac{\lambda_h}{\lambda_r}\frac{(u_h^\ell)^2}{(u_r^\ell)^2} \longrightarrow 0. \tag{8.12}$$

If π_I^ℓ denotes the normalized coefficient in (8.6), then

$$0 \le \pi_I^\ell \le \frac{\beta_I(u^\ell)}{\beta_J(u^\ell)} \longrightarrow 0. \tag{8.13}$$

There are finitely many sets omitting r, and their templates are fixed finite real numbers. The sets containing r contribute zero, so (8.6) proves (8.10). Since $v_r^\ell \to v_r$, one has $y_r^\ell \to 0$. Thus the I_+ subsequence and the I_- subsequence both converge to $y = 0$.

Case 4: $u = 0$. Compatibility gives $v = 0$. The I_- subsequence again has zero output. On I_+, (8.9) gives

$$\left\|y^\ell\right\| \le B_p \left\|v^\ell\right\| \longrightarrow 0.$$

Therefore $y^\ell \to 0 = y$.

The four cases cover every possible limit state. Sequential continuity is equivalent to continuity in these finite-dimensional metric spaces. The continuity of T_p follows from that of F_p and the continuous first component $(u, v) \mapsto v$. □

9. Global *R*-linear convergence

Rate convention For fixed H and p, let $z_k = T_p^k(z_0)$. We say that the standard endpoint dynamics is state-uniformly R-linearly convergent if there exist $C \ge 1$ and $\rho \in (0,1)$, independent of $z_0 \in \Omega_p$, such that

$$\|z_k\|_\times \le C\rho^k \|z_0\|_\times, \qquad k \ge 0.$$

Here k counts completed sweeps. Finitely terminating orbits are extended by zeros.

9.1. *Pointwise convergence input*

LEMMA 9.1 *Let $(u_0,u_1)\in\Omega_p$. There are stepsizes*

$$\alpha_1,\dots,\alpha_p\in[\lambda_n^{-1},\lambda_1^{-1}]$$

such that

$$u_1=\prod_{j=1}^{p}(I-\alpha_jH)u_0. \tag{9.1}$$

Use these steps as an admissible warm-up block for Algorithm 2.1. Until the algorithm terminates, let G_j denote the history block starting at u_j and ending at u_{j+1}. Then

$$\operatorname{span}(G_0)=\mathscr{K}_p(H,u_0), \tag{9.2}$$

and the projected values extracted from G_0 agree exactly with those used in the definition of $F_p(u_0,u_1)$. More generally, if

$$z_j:=(u_j,u_{j+1}),$$

then the algorithmic endpoints and the abstract endpoints satisfy

$$z_{j+1}=T_p(z_j) \tag{9.3}$$

until finite termination, after which the abstract orbit reaches $(0,0)$ and remains there. Consequently,

$$T_p^k(u_0,u_1)\longrightarrow(0,0)\qquad\text{for every }(u_0,u_1)\in\Omega_p. \tag{9.4}$$

Proof Compatibility gives the representation (9.1); use these factors as an admissible warm-up block. If an intermediate gradient vanishes, then $u_1=0$ and one endpoint transition reaches $(0,0)$.

Otherwise the history is completed and, by Proposition 2.2, its span is $\mathscr{K}_p(H,u_0)$. If $s(u_0)>p$, the next endpoint is $\Psi_{u_0}(H)u_1$. If $s(u_0)\le p$, Proposition 2.5 gives the zero output prescribed by F_p. Repeating this argument shows that the algorithmic and abstract endpoints agree until termination. The conclusion then follows from Theorem 5.1. □

For an actual run with arbitrary positive finite warm-up steps, we initialize the endpoint estimate at

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

the start/end pair of the first completed standard sweep. If the warm-up steps themselves lie in $[\lambda_n^{-1},\lambda_1^{-1}]$, the warm-up start/end pair may be used as the initial state.

9.2. *The standard rank-compressed method*

THEOREM 9.2 *Fix the reduced matrix H in (2.5) and an integer $p\ge1$. Then there exist constants*

$$C_*=C_*(H,p)\ge1,\qquad\rho_*=\rho_*(H,p)\in(0,1),$$

such that for every initial state $z_0=(u_0,u_1)\in\Omega_p$, if

$$z_k=T_p^k(z_0)=(u_k,u_{k+1}),$$

then

$$\|(u_k,u_{k+1})\|_\times\le C_*\rho_*^k\|(u_0,u_1)\|_\times,\qquad k\ge0. \tag{9.5}$$

Finite termination is included by extending the endpoint sequence with zeros.

Proof Lemma 8.1 shows that Ω_p is a finite-dimensional closed cone. Proposition 8.2 and Theorem 8.4 show that T_p is a continuous positively homogeneous self-map. Lemma 9.1 gives pointwise convergence for every state in Ω_p. Theorem 7.1 yields (9.5) with constants depending only on (H,p). □

9.3. *Fixed spectral weights by conjugacy*

For a fixed positive spectral weight $W=\omega(H)$, let $T_{p,W}$ denote the endpoint transition obtained from (6.9) by using the weighted pencil (2.15) on the regular stratum and the same zero output on the two finite-termination strata. Set $S=W^{1/2}$ and define the product-space isomorphism

$$\mathscr{S}_W:\Omega_p\to\Omega_p,\qquad \mathscr{S}_W(u,v):=(Su,Sv). \tag{9.6}$$

PROPOSITION 9.3 *The map $\mathscr{S}_W$ is a linear homeomorphism of Ω_p onto itself, and*

$$\mathscr{S}_W T_{p,W}=T_p\mathscr{S}_W,\qquad \textit{equivalently}\qquad T_{p,W}=\mathscr{S}_W^{-1}T_p\mathscr{S}_W. \tag{9.7}$$

Thus the two maps have identical complete-sweep polynomials at corresponding conjugate states, conjugate orbit structure, identical Krylov ranks and identical finite-termination indices.

Proof If $v=q(H)u$ for some $q\in\mathscr{Q}_p$, then $Sv=q(H)Su$ because S commutes with H; the same argument with S^{-1} shows that $\mathscr{S}_W$ maps Ω_p onto itself. Moreover,

$$K_p(Su)=SK_p(u),$$

and hence

$$K_p(Su)^\top HK_p(Su)=K_p(u)^\top WHK_p(u),\qquad K_p(Su)^\top K_p(Su)=K_p(u)^\top WK_p(u).$$

Therefore the standard pencil at Su is exactly the weighted pencil at u. Since S is invertible and acts by a positive scalar on each eigenspace of H, it preserves active support, Krylov rank and both finite-termination strata. On the regular stratum the two maps consequently use the same complete-sweep polynomial, while on the termination strata both outputs are zero. This proves (9.7). □

COROLLARY 9.4 *Let C_* and ρ_* be the standard constants in Theorem 9.2. For every fixed positive spectral weight $W=\omega(H)$, every $z\in\Omega_p$ and every $k\ge 0$,*

$$\left\|T_{p,W}^k(z)\right\|_\times\le C_*\sqrt{\kappa_2(W)}\,\rho_*^k\,\|z\|_\times. \tag{9.8}$$

In particular, every fixed weight has the same admissible decay factor $\rho_*=\rho_*(H,p)$. For $W=H^a$ with arbitrary $a\in\mathbb{R}$,

$$\sqrt{\kappa_2(W)}=\kappa_2(H)^{|a|/2}. \tag{9.9}$$

Proof Iterating (9.7) and applying Theorem 9.2 give

$$\left\|T_{p,W}^k(z)\right\|_\times\le\left\|S^{-1}\right\|_2 C_*\rho_*^k\,\|\mathscr{S}_W z\|_\times\le C_*\left\|S^{-1}\right\|_2\|S\|_2\,\rho_*^k\,\|z\|_\times.$$

Since $S=W^{1/2}$, one has $\left\|S^{-1}\right\|_2\|S\|_2=\sqrt{\kappa_2(W)}$. Equation (9.9) follows from the spectral condition number of H^a. □

9.4. *Inner iterates and optimization errors*

Let

$$M_H := \max_{\lambda,\theta\in[\lambda_1,\lambda_n]} |1-\lambda/\theta| \le \max\left\{1, \frac{\lambda_n}{\lambda_1}-1\right\}, \tag{9.10}$$

and define

$$M_{H,p} := \max_{0\le j\le p} M_H^j = \max\{1, M_H^p\}. \tag{9.11}$$

For a non-terminating actual run, reindex the gradients from the start of the first sweep after the warm-up by

$$\widehat{g}_{kp+j} := g_{k+1,j}, \qquad k\ge 0, \quad 0\le j\le p-1. \tag{9.12}$$

At a block boundary, (2.9) gives

$$\widehat{g}_{(k+1)p} = g_{k+1,p} = g_{k+2,0},$$

so the endpoint relation is

$$u_k = \widehat{g}_{kp} = g_{k+1,0}, \qquad k\ge 0. \tag{9.13}$$

The within-sweep bound, including the endpoint $j=p$, is

$$\left\|\widehat{g}_{kp+j}\right\| \le M_H^j \left\|u_k\right\|, \qquad 0\le j\le p. \tag{9.14}$$

Any exact shorter block belongs to the finite-termination branch and does not occur on a non-terminating run.

COROLLARY 9.5 *Fix H and p, let $W=\omega(H)\succ 0$ and consider the corresponding rank-compressed method with arbitrary positive finite warm-up stepsizes. Define*

$$\widehat{\rho}_* := \rho_*^{1/p}, \qquad \widehat{C}_W := \max\left\{1, M_{H,p} C_* \sqrt{\kappa_2(W)}\,\widehat{\rho}_*^{-(p-1)}\right\}, \tag{9.15}$$

where C_* and ρ_* are the standard endpoint constants in Theorem 9.2. Either the method terminates finitely or the reindexed sequence (9.12) satisfies

$$\|\widehat{g}_\ell\| \le \widehat{C}_W \widehat{\rho}_*^{\ell} \|(u_0,u_1)\|_\times, \qquad \ell\ge 0, \tag{9.16}$$

where $(u_0,u_1)=(g_{1,0},g_{2,0})$ is the start/end pair of the first completed sweep after the warm-up. If $\widehat{x}_\ell$ denotes the correspondingly reindexed iterate, then

$$\|\widehat{x}_\ell - x^\star\| \le \lambda_1^{-1}\widehat{C}_W \widehat{\rho}_*^{\ell} \|(u_0,u_1)\|_\times. \tag{9.17}$$

The objective gap admits the bound

$$0\le f(\widehat{x}_\ell) - f(x^\star) \le \frac{\widehat{C}_W^2}{2\lambda_1} (\widehat{\rho}_*^2)^\ell \|(u_0,u_1)\|_\times^2. \tag{9.18}$$

Thus every fixed spectral weight admits the same gradient and iterate decay factor $\widehat{\rho}_*$ and the same objective-gap factor $\widehat{\rho}_*^2$.

Proof By Proposition 2.3, a non-terminating fixed-weight run corresponds to a non-terminating standard run, so every sweep has length p. Write $\ell = kp + j$ with $0 \le j \le p-1$. Equations (9.14) and (9.8) give

$$\|\widehat{g}_\ell\| \le M_{H,p} C_* \sqrt{\kappa_2(W)}\, \rho_*^k \, \|(u_0, u_1)\|_\times .$$

Since $k = (\ell - j)/p$,

$$\rho_*^k = \widehat{\rho}_*^{\ell - j} \le \widehat{\rho}_*^{-(p-1)} \widehat{\rho}_*^{\ell}.$$

Equation (9.16) follows from (9.15). The gradient-error equivalence

$$\lambda_1 \|x - x^\star\| \le \|g(x)\| \le \lambda_n \|x - x^\star\| \tag{9.19}$$

gives (9.17). Finally,

$$0 \le f(x) - f(x^\star) = \tfrac{1}{2} g(x)^\top H^{-1} g(x) \le \frac{1}{2\lambda_1} \|g(x)\|^2 , \tag{9.20}$$

which gives (9.18). □

COROLLARY 9.6 *All standard and fixed-weight endpoint, inner-gradient, iterate-error and objective-gap estimates in this section remain valid when H has repeated eigenvalues. The common decay factors ρ_* and $\widehat{\rho}_*$ may be chosen to depend only on p and the distinct eigenvalues of H.*

Proof Apply the isometric spectral-projector reduction described in Sections 2 and 6. It preserves compatible endpoint pairs, projected pencils, all polynomial endpoint and inner-sweep updates and their Euclidean norms. The reduced endpoint cone contains every active coordinate pattern, so the reduced constants apply uniformly to all initial active subsets. The iterate-error and objective-gap estimates then follow from the same spectral bounds. □

10. Special cases and relation to earlier analyses

COROLLARY 10.1 *The global-convergence and R-linear results apply to*

(i) *standard Ritz LMSD, $W = I$;*
(ii) *harmonic-Ritz LMSD, $W = H$;*
(iii) *every fixed power weight $W = H^a$, $a \in \mathbb{R}$;*
(iv) *the delayed BB1 and BB2 recurrences obtained for $p = 1$.*

All four cases admit the standard endpoint decay factor $\rho_*(H, p)$ and the standard per-step factor $\widehat{\rho}_* = \rho_*^{1/p}$. The endpoint prefactor is C_* for $W = I$, at most $C_* \sqrt{\kappa_2(H)}$ for $W = H$ and at most $C_* \kappa_2(H)^{|a|/2}$ for $W = H^a$.

Proof These choices are positive spectral weights, and for $p = 1$ the corresponding reciprocal generalized Rayleigh quotients are the quadratic BB1 and BB2 stepsizes. Apply Proposition 9.3, Corollary 9.4 and (9.9). □

For $p = 1$, the closed-cone conjugacy and the state-uniform R-linear estimate are also established in Yang & Yuan (2026, Proposition 3.2 and Theorem 3.4).

Curtis & Guo (2018) obtain finite-cycle contractions for standard and harmonic-Ritz LMSD under Assumption 3.4. Their constants inherit the normalized-conditioning parameter through equation (3.12c) and Lemmas 3.10–3.11. Theorem 9.2 instead uses the compatible-state formulation and algebraic rank compression, and Corollary 9.4 transfers the resulting rate to fixed positive spectral weights by conjugacy.

For $W = I$, Theorem 5.1 also recovers Fletcher's global-convergence result (Fletcher, 2012) through the complete-sweep determinant representation.

11. Numerical diagnostics

All computations were carried out in double precision with MATLAB R2023a. Ritz values were applied in non-increasing order. Power weights were formed in the logarithmic domain and normalized by a common positive scale. All experiments use the known diagonal spectral representation.

A history matrix G was assigned numerical rank equal to the number of singular values satisfying

$$\sigma_j > \max\{10^{-12}\,\sigma_1,\, \varepsilon_{\mathrm{mach}}\max(\mathrm{size}(G))\sigma_1\}. \tag{11.1}$$

The retained left singular vectors formed the basis of the compressed projected pencil. Iterative runs stopped when

$$\frac{\|g\|}{\|g_0\|} \le 10^{-11},$$

and Cauchy steps were used for the warm-up block.

For reproducibility, define the deterministic coefficients

$$d_i(s) := 1 + \sin\big(0.37(i+s)\big) + 0.3\cos\big(1.17(i+s)\big) + 0.5\sin\big(0.11(i+s)\big). \tag{11.2}$$

The initial gradients used below are specified in terms of these coefficients.

The experiment combines the power-weight comparison, repeated eigenvalues and actual-history conditioning. We take $p = 5$ and 20 distinct eigenvalues, consisting of four equally spaced points in each of

$$[1,2], \quad [10,15], \quad [50,60], \quad [200,250], \quad [800,1000].$$

Every distinct eigenvalue has multiplicity three, so the unreduced dimension is $N = 60$. In the diagonal spectral coordinates, the initial gradient is

$$(g_0)_i = d_i(7), \qquad i = 1,\dots,60,$$

and therefore has a non-zero projection onto every eigenspace. We compare

$$a \in \{-1, -\tfrac{1}{2}, 0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2\}, \qquad W = H^a.$$

Since the minimizer has been translated to the origin, the objective gap is evaluated in spectral coordinates as

$$f(x_j) - f_\star = \frac{1}{2} g_j^\top H^{-1} g_j.$$

The quantity displayed in the left panel is the relative gap

$$r_j := \frac{f(x_j) - f_\star}{f(x_0) - f_\star}.$$

The left panel of Fig. 1 shows the raw relative objective gap after every gradient step. The right panel shows

$$\chi_k = \frac{\|g_{k,0}\|}{\sigma_{\min}(G_k)}$$

for the standard and harmonic choices, which are the two variants considered in the earlier conditioning-based analysis.

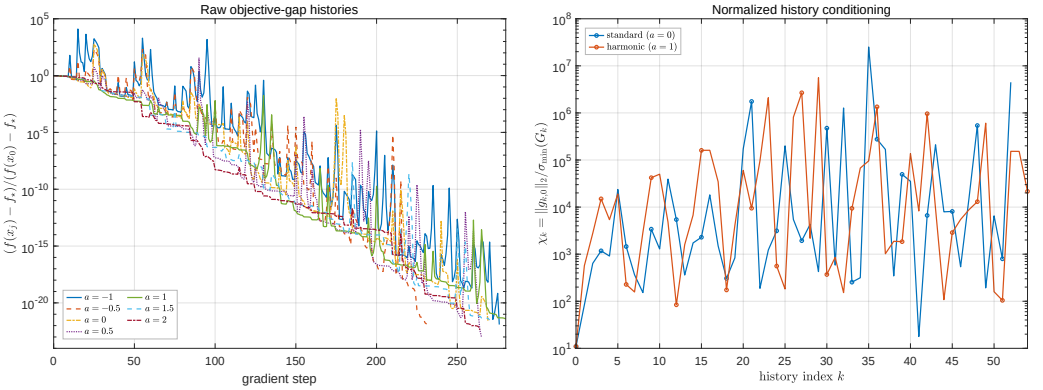


FIG. 1. A repeated-spectrum quadratic with $N = 60$, 20 distinct eigenvalues, and $p = 5$.

Alt text: Two line plots for the repeated-spectrum quadratic. The left plot shows the relative objective gap on a logarithmic scale after each gradient step for seven fixed power weights. The curves have transient spikes and an overall downward trend. The right plot shows the normalized history-conditioning quantity on a logarithmic scale for standard and harmonic-Ritz LMSD. Both curves fluctuate strongly across histories.

TABLE 1. *Diagnostics for the repeated-spectrum quadratic.* $A_k = \max_{0\le j\le p} \|g_{k,j}\| / \|g_{k,0}\|$.

a	Gradient steps	Weighted sweeps	$\max_k \chi_k$	$\max_k A_k$
-1	276	55	1.30×10^8	1.33×10^3
-0.5	231	46	2.35×10^8	1.55×10^5
0	269	53	2.51×10^7	2.26×10^5
0.5	265	52	1.59×10^7	6.93×10^4
1	280	55	5.59×10^6	3.30×10^4
1.5	270	53	1.99×10^7	2.21×10^4
2	264	52	1.48×10^7	6.99×10^1

As summarized in Table 1, all seven weights reach the prescribed gradient tolerance in 231–280 gradient steps. The raw objective histories are strongly non-monotone and can undergo substantial transient growth, but the overall decay on the logarithmic scale is consistent with the global R-linear result.

All detected histories have rank five, while the normalized quantities attain

$$\max_k \chi_k = 2.51 \times 10^7 \quad \text{and} \quad 5.59 \times 10^6,$$

for the standard and harmonic runs, respectively. Similar, and in some cases substantially larger, values were already reported by Curtis & Guo (2018, Table 1). The observations illustrate that full rank alone provides neither a moderate nor an *a priori* bound on the normalized conditioning constant.

12. Conclusions and limitations

Exact algebraic rank compression gives a common formulation of the regular and finite-termination strata of the standard complete-sweep method. If a block-start gradient has at most p active distinct eigenvalues, the following delayed sweep terminates finitely. On non-terminating trajectories, the determinant and Cauchy–Binet representation of the complete-sweep multiplier yields pointwise global convergence across changes in Krylov rank.

Consecutive endpoints define a continuous positively homogeneous self-map on the compatible-state cone. Pointwise convergence and compactness give an endpoint R-linear estimate with constants $C_*(H,p)$ and $\rho_*(H,p)$ uniform over compatible initial states. Endpoint conjugacy transfers the same admissible decay factor to every fixed positive spectral weight, with Euclidean-norm prefactor at most $C_*\sqrt{\kappa_2(W)}$. The estimates extend to inner gradients, iterate errors, objective gaps and matrices with repeated eigenvalues.

The analysis concerns strictly convex quadratics in exact arithmetic with algebraic rank compression and fixed positive spectral weights. Stability under numerical rank thresholds, trajectory-dependent weights and non-quadratic extensions remains open.

Acknowledgements

SY thanks Shixiang Chen for supervising his undergraduate thesis at the University of Science and Technology of China, and Xiaowei Xu for supervising his project under the National College Students Innovation and Entrepreneurship Training Program at the same university. The present article grew out of those two undergraduate projects. He also thanks Gexuan Zhu and Bowen Deng for carefully reading an earlier Chinese-language manuscript presenting the global convergence part of this work and for offering helpful comments and suggestions.

Use of generative AI During the preparation of this manuscript, Shutai Yang used GPT-5.5 Thinking (OpenAI) for language and LaTeX editing, the presentation of mathematical notation, assistance in checking mathematical arguments and references and revision of the MATLAB scripts used for the numerical experiments. He directed the use of the tool, independently checked and revised all model-assisted material and made all scholarly judgments and final decisions. Both authors reviewed and approved the final manuscript. The authors assume responsibility for all content.

Conflict of interest

Ya-xiang Yuan is an Associate Editor of the IMA Journal of Numerical Analysis. The authors request independent editorial handling of this manuscript.

Funding

National College Students Innovation and Entrepreneurship Training Program (202510358091 to S.Y.); National Natural Science Foundation of China (12288201 to Y.-X.Y.).

References

Barzilai, J. & Borwein, J. M. (1988) Two-point step size gradient methods. *IMA J. Numer. Anal.*, **8**, 141–148. https://doi.org/10.1093/imanum/8.1.141.

Curtis, F. E. & Guo, W. (2018) *R*-linear convergence of limited memory steepest descent. *IMA J. Numer. Anal.*, **38**, 720–742. https://doi.org/10.1093/imanum/drx016.

Dai, Y.-H. & Liao, L.-Z. (2002) *R*-linear convergence of the Barzilai and Borwein gradient method. *IMA J. Numer. Anal.*, **22**, 1–10. https://doi.org/10.1093/imanum/22.1.1.

di Serafino, D., Ruggiero, V., Toraldo, G. & Zanni, L. (2018) On the steplength selection in gradient methods for unconstrained optimization. *Appl. Math. Comput.*, **318**, 176–195. https://doi.org/10.1016/j.amc.2017.07.037.

Ferrandi, G. & Hochstenbach, M. E. (2025) Limited memory gradient methods for unconstrained optimization. *Numer. Algorithms*, **99**, 735–764. https://doi.org/10.1007/s11075-024-01895-9.

Ferrandi, G., Hochstenbach, M. E. & Krejić, N. (2023) A harmonic framework for stepsize selection in gradient methods. *Comput. Optim. Appl.*, **85**, 75–106. https://doi.org/10.1007/s10589-023-00455-6.

Fletcher, R. (2012) A limited memory steepest descent method. *Math. Program.*, **135**, 413–436. https://doi.org/10.1007/s10107-011-0479-6.

Friedlander, A., Martínez, J. M., Molina, B. & Raydan, M. (1999) Gradient method with retards and generalizations. *SIAM J. Numer. Anal.*, **36**, 275–289. https://doi.org/10.1137/S003614299427315X.

Gu, R. & Du, Q. (2021) A modified limited memory steepest descent method motivated by an inexact super-linear convergence rate analysis. *IMA J. Numer. Anal.*, **41**, 247–270. https://doi.org/10.1093/imanum/drz059.

Raydan, M. (1993) On the Barzilai and Borwein choice of steplength for the gradient method. *IMA J. Numer. Anal.*, **13**, 321–326. https://doi.org/10.1093/imanum/13.3.321.

Shen, J. & Hu, J. (2012) Stability of discrete-time switched homogeneous systems on cones and conewise homogeneous inclusions. *SIAM J. Control Optim.*, **50**, 2216–2253. https://doi.org/10.1137/110845215.

Yang, S. & Yuan, Y.-X. (2026) *R*-linear convergence of Barzilai–Borwein methods on strictly convex quadratics via closed-cone homogeneous dynamics. arXiv:2609.12100. https://arxiv.org/abs/2609.12100.

Zhou, Y., Gu, R. & Yuan, Y.-X. (2026) Structured spectral step-sizes and rebound-aware ordering for gradient methods. arXiv:2606.25311v3. https://arxiv.org/abs/2606.25311v3.

Zou, Q. & Magoulès, F. (2019) On extensions of limited memory steepest descent method. *2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. IEEE, pp. 172–175. https://doi.org/10.1109/DCABES48411.2019.00050.

Zou, Q. & Magoulès, F. (2022) Delayed gradient methods for symmetric and positive definite linear systems. *SIAM Rev.*, **64**, 517–553. https://doi.org/10.1137/20M1321140.