Convergence of Normalized Residuals for Restarted Conjugate Gradients of Length Two

July 29, 2026

Abstract

Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite, let $f(x) = \frac{1}{2}x^T Ax - b^T x$, and set $r_k = b - Ax_k$. In exact arithmetic, define each restart of length two by minimizing f over $x_k + \operatorname{span}\{r_k, Ar_k\}$. If the initial residual has nonzero spectral projection on at least three distinct eigenspaces of A, then every residual remains nonzero with spectral grade at least three, and the even and odd subsequences of Euclidean-normalized residuals converge to unit vectors y_{e} and y_{o}, respectively. Both limits have spectral grade three or four and form a two-cycle under the normalized monic-quadratic residual map. Moreover, if $E_k = \|r_k\|^2_{A^{-1}}$, then

$$0 < \frac{E_{k+1}}{E_k} \le \frac{E_{k+2}}{E_{k+1}} \le \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 < 1,$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the extreme eigenvalues in the initial cyclic subspace. No simple-spectrum or genericity assumption is imposed. Each restart uses two matrix–vector products and $O(n)$ additional arithmetic operations.

Note. This paper was fully AI generated by a harness wrapped around GPT 5.6 Sol Ultra from a one-shot prompt by Richard Peng.

Contents

1 Introduction

Conjugate gradients (CG) solves a symmetric positive definite linear system $Ax = b$ by minimizing

$$f(x) = \frac{1}{2}x^T Ax - b^T x$$

over expanding Krylov spaces. In exact arithmetic, the unrestarted method terminates after at most the number of distinct eigenvalues active in the initial residual [HS52]. A fixed restart length changes the question: after s inner CG steps, the accumulated conjugacy relations are discarded and the same construction begins again from the new residual. Thus, if $r_k = b - Ax_k$, one outer step minimizes f over

$$x_k + \mathcal{K}_s(A, r_k), \qquad \mathcal{K}_s(A, r_k) = \operatorname{span}\{r_k, Ar_k, \ldots, A^{s-1}r_k\}.$$

Forsythe called this the optimum s-gradient method; in modern terms it is CG restarted every s steps [For68, FLT23].

Even as the iterates approach the solution and the residuals tend to zero, their normalized directions may oscillate. For restart length $s = 1$, the iteration is exact-line-search steepest descent. Forsythe and Motzkin formulated the corresponding direction problem, and Akaike proved that the even and odd normalized residuals converge separately to two alternating directions supported on the extreme active eigenspaces [FM51, Aka59]. Forsythe's general conjecture asks for the same parity convergence at every fixed restart length: each of $\{r_{2k}/\|r_{2k}\|\}$ and $\{r_{2k+1}/\|r_{2k+1}\|\}$ should have a single limit [For68, FLT23].

We prove this statement for restart length two. For $v \neq 0$, let $d(A, v)$ denote its spectral grade, namely the number of distinct eigenvalues of A on whose eigenspaces v has nonzero projection. For a subspace V, write $u \perp_A V$ when $\langle Au, v\rangle = 0$ for every $v \in V$.

Theorem 1.1 (Forsythe conjecture for restart length two)**.** *In exact arithmetic, let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite, let $b, x_0 \in \mathbb{R}^n$, put $x_\star = A^{-1}b$ and $r_0 = b - Ax_0$, and assume $d(A, r_0) \geq 3$. For $k \geq 0$, define*

$$\mathcal{K}_2(A, r_k) = \operatorname{span}\{r_k, Ar_k\}, \qquad r_k = b - Ax_k,$$

and let x_{k+1} be the unique vector satisfying

$$x_{k+1} \in x_k + \mathcal{K}_2(A, r_k), \qquad x_\star - x_{k+1} \perp_A \mathcal{K}_2(A, r_k).$$

Equivalently, these are the iterates generated by RESTARTED-CG(2)*. Then $d(A, r_k) \geq 3$ for every k. If $E_k = \|r_k\|^2_{A^{-1}}$, and if $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues on which r_0 has nonzero spectral projection, then*

$$0 < \frac{E_{k+1}}{E_k} \leq \frac{E_{k+2}}{E_{k+1}} \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 < 1.$$

Furthermore, there are unit vectors $y_{\mathrm{e}}, y_{\mathrm{o}}$ such that

$$\frac{r_{2k}}{\|r_{2k}\|} \longrightarrow y_{\mathrm{e}}, \qquad \frac{r_{2k+1}}{\|r_{2k+1}\|} \longrightarrow y_{\mathrm{o}}.$$

Both limit vectors have grade three or four. For the monic-quadratic map T defined in Equation (1), they satisfy $y_{\mathrm{o}} = T(y_{\mathrm{e}})$ and $y_{\mathrm{e}} = T(y_{\mathrm{o}})$. The deterministic implementation produces the iterates, residuals, and normalized residuals. If one multiplication by A costs T_A, then each restart costs $2T_A + O(n)$ arithmetic operations, and hence $O(n^2)$ operations for an explicit dense matrix.

The grade hypothesis is the natural nontermination threshold. A residual of grade at most two is annihilated by a CG residual polynomial of degree at most two, after which its direction is undefined. The theorem shows that grade at least three instead persists at every restart. It also aggregates repeated eigenvalues through their spectral projections, so no simple-spectrum assumption is needed, and it imposes no genericity condition on the starting residual. The two limiting directions are companion points under T, and their grades are rigidly restricted to three or four.

Results for the same iteration. The classical $s = 1$ result and the case $d(A, r_0) = s + 1$, in which Forsythe showed that the normalized process is already an exact two-cycle, are restricted versions of the same problem [Aka59, For68]. For $s = 2$, Zhuk and Bondarenko subsequently presented a confirmation of Forsythe's conjecture and derived convergence of the relevant polynomial coefficients together with rate and asymptotic-spectrum results [ZB84]. Faber, Liesen, and Tichý later observed that the passage from coefficient convergence to a single residual direction relies on an implication attributed to Zabolotskaya; the accessible English translation describes the coefficient limit as indicating, rather than establishing, a unique direction [Zab79, FLT23]. The present proof supplies the direction-convergence argument directly, without treating coefficient convergence alone as sufficient.

Pronzato, Wynn, and Zhigljavsky studied exactly the same optimum s-gradient iteration as a dynamical system on spectral probability measures [PWZ09]. Their $s = 2$ analysis establishes a singleton in the three-point regime, while fixed limiting invariants can leave a one-parameter family of four-point candidates. Faber, Liesen, and Tichý obtained a modern projection formulation, two-step asymptotic regularity, connected parity limit sets, and the bound $s < d \leq 2s$ for every limiting direction [FLT23]. At $s = 2$, these results isolate grades three and four but do not turn a possible four-point continuum into a single signed direction. Theorem 1.1 covers both grades for every finite-dimensional instance satisfying the stated nontermination hypothesis.

Nearby but different iterations. The nonsymmetric Arnoldi cross iteration of [FLT23] alternates projection steps for A and A^T; only its symmetric specialization has the direction dynamics studied here. Restarted GMRES instead minimizes the Euclidean residual norm, and its cycle-convergence results concern a different process [SS86, VL10]. Likewise, the communication-avoiding "s-step CG" method retains conjugacy across blocks and is not CG restarted after every s inner steps [CG89]. These methods provide neighboring Krylov context, but they are not same-problem partial results.

Proof perspective and organization. The proof represents one restart by a monic quadratic that is orthogonal for the current spectral measure. Two monotone scalar quantities then force a limiting two-restart quartic and a connected omega-limit set supported on three or four spectral coordinates. Three coordinates reduce to a finite algebraic classification. Four coordinates admit a continuum of exact two-cycles, so polynomial convergence alone does not rule out motion along that continuum. A corrected barycentric coordinate retains the effect of vanishing outside coordinates, while a Lyapunov-tail estimate and a logarithmic cocycle exclude the remaining resonant drift.

Section 2 fixes the linear-algebra, spectral, and orthogonal-polynomial notation. Section 3 gives the proof architecture and the main identities. Section 4 proves the theorem, beginning with the exact restarted-CG implementation and ending with the collapse of the three- and four-coordinate omega-limit sets.

2 Preliminaries

Linear-algebra conventions. All arithmetic is exact. Inner products and norms without a subscript are Euclidean. For a symmetric positive definite matrix A, write

$$\langle u, v\rangle_A = \langle Au, v\rangle, \qquad \langle u, v\rangle_{A^{-1}} = \langle A^{-1}u, v\rangle$$

for the energy and inverse-energy inner products, with corresponding norms $\|\cdot\|_A$ and $\|\cdot\|_{A^{-1}}$. The matrix norm $\|A\|$ is the induced Euclidean operator norm. For a subspace V, the notation $u \perp_A V$ means $\langle u, v\rangle_A = 0$ for every $v \in V$.

For fixed A and b, put

$$x_\star = A^{-1}b, \qquad f(x) = \frac{1}{2}x^T Ax - b^T x, \qquad r(x) = b - Ax = A(x_\star - x),$$

and set $\mathcal{K}_2(A, r) = \operatorname{span}\{r, Ar\}$. The standard quadratic identities

$$\nabla f(x) = -r(x), \qquad f(x) - f(x_\star) = \frac{1}{2}\|x - x_\star\|_A^2 = \frac{1}{2}\|r(x)\|_{A^{-1}}^2$$

show that, for any subspace V and any $x^+ \in x + V$,

$$x^+ = \operatorname*{argmin}_{z\in x+V} f(z) \iff r(x^+) \perp V \iff x_\star - x^+ \perp_A V.$$

For a nonzero residual $r_k = r(x_k)$, we write $y_k = r_k/\|r_k\|$.

Spectral data.

Definition 2.1 (Spectral grade and support)**.** *Let Π_λ be the orthogonal spectral projector of a symmetric matrix A associated with λ. For $v \in \mathbb{R}^n$, define*

$$\Lambda(A, v) = \{\lambda : \Pi_\lambda v \neq 0\}, \qquad d(A, v) = |\Lambda(A, v)|.$$

For a coordinate vector z, write $\operatorname{supp}(z) = \{i : z_i \neq 0\}$.

For $v \neq 0$, the cyclic subspace generated by v is

$$\operatorname{span}\{p(A)v : p \in \mathbb{R}[t]\} = \operatorname{span}\{\Pi_\lambda v : \lambda \in \Lambda(A, v)\}.$$

The spectral theorem gives

$$p(A)v = \sum_{\lambda\in\Lambda(A,v)} p(\lambda)\Pi_\lambda v.$$

Thus $p(A)v = 0$ exactly when p vanishes on $\Lambda(A, v)$. When A and v are clear, the extreme elements of this active spectrum are denoted by $\lambda_{\min}$ and $\lambda_{\max}$.

For a unit vector y, its spectral measure and moments are

$$\nu_y = \sum_{\lambda\in\Lambda(A,y)} \|\Pi_\lambda y\|^2 \delta_\lambda, \qquad \mu_j(y) = \int t^j \, d\nu_y(t) = \langle A^j y, y\rangle \quad (j \geq 0).$$

Here δ_λ is the unit point mass at λ.

Definition 2.2 (Normalized monic-quadratic map)**.** *For a unit vector y with $d(A, y) \geq 3$, let*

$$\pi_y(t) = t^2 + c_1(y)t + c_0(y)$$

be the monic quadratic satisfying

$$\langle \pi_y(A)y, y\rangle = 0, \qquad \langle \pi_y(A)y, Ay\rangle = 0.$$

Its coefficients are characterized by

$$\begin{pmatrix} \mu_1(y) & \mu_0(y) \\ \mu_2(y) & \mu_1(y) \end{pmatrix} \begin{pmatrix} c_1(y) \\ c_0(y) \end{pmatrix} = - \begin{pmatrix} \mu_2(y) \\ \mu_3(y) \end{pmatrix}.$$

Strict Cauchy–Schwarz gives

$$c_0(y) = \frac{\mu_1(y)\mu_3(y) - \mu_2(y)^2}{\mu_0(y)\mu_2(y) - \mu_1(y)^2} > 0.$$

Define

$$\theta(y) = \|\pi_y(A)y\|, \qquad T(y) = \frac{\pi_y(A)y}{\theta(y)}, \qquad P_y(t) = \frac{\pi_y(t)}{c_0(y)}. \tag{1}$$

A quadratic cannot vanish on three distinct active eigenvalues, so $\theta(y) > 0$. Hence the objects in Definition 2.2 are well defined, $P_y(0) = 1$, and c_1, c_0, θ, T, and $y \mapsto P_y$ are continuous on the grade-at-least-three domain. For every monic quadratic q, orthogonality and the fact that $q - \pi_y$ is linear give

$$\|q(A)y\|^2 = \|\pi_y(A)y\|^2 + \|(q - \pi_y)(A)y\|^2. \tag{2}$$

Thus $\theta(y)$ is the minimum of $\|q(A)y\|$ over monic quadratics q.

Limit sets. For a precompact sequence $\{z_k\}$, let $\omega(z_k)$ denote its set of subsequential limits. This set is nonempty and compact. If $\|z_{k+1} - z_k\| \to 0$, then $\omega(z_k)$ is connected; if $\omega(z_k)$ is a singleton, then z_k converges.

3 Overview

The proof converts restarted conjugate gradients into a dynamical system on the unit sphere and then shows that the even omega-limit set of this system is a singleton. The common part of the argument produces a rigid limiting quartic whose equality set contains only three or four active eigenvalues. The grade-three case then collapses by finite-dimensional algebra. The main technical challenge is the four-eigenvalue case: the exact dynamics on a four-node spectral face has a continuum of two-cycles, so convergence of the quartic does not by itself prevent drift along that continuum. Moreover, spectral coordinates outside the face may vanish without geometric decay. We control this drift by a corrected barycentric coordinate, charge the exceptional transverse mass to a Lyapunov tail, and exclude the remaining multiplier--1 resonance with a logarithmic cocycle.

Polynomial and spectral reduction. In exact arithmetic, the algorithm RESTARTED-CG(2) takes A, b, x_0, with A symmetric positive definite, $r_0 = b - Ax_0$, and $d(A, r_0) \geq 3$, and returns the iterates x_k, residuals $r_k = b - Ax_k$, and directions $y_k = r_k/\|r_k\|$. At each restart it minimizes the quadratic objective over $x_k + \operatorname{span}\{r_k, Ar_k\}$, or equivalently makes r_{k+1} orthogonal to that Krylov space. The two optimality equations give the exact residual-polynomial form

$$r_{k+1} = P_{y_k}(A)r_k, \qquad P_y(t) = \frac{\pi_y(t)}{c_0(y)}, \qquad y_{k+1} = T(y_k) = \frac{\pi_{y_k}(A)y_k}{\theta(y_k)}.$$

Here π_y, $c_0(y) = \pi_y(0) > 0$, and $\theta(y)$ are as in Definition 2.2. The actual residual polynomial is normalized to have constant term one, while its monic multiple π_y is useful because it is the minimum-norm monic quadratic in Equation (2); its norm will be the first Lyapunov quantity. A fixed spectral isometry then replaces each active eigenspace by one coordinate, reducing the problem to $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ with distinct eigenvalues. This loses no information and is why repeated eigenvalues require no genericity assumption. Computationally, one restart forms Ar_k, $A^2 r_k$, solves a 2×2 system, and performs vector updates, for a cost of $2T_A + O(n)$ when one multiplication by A costs T_A, or $O(n^2)$ for a dense matrix; see Lemmas 4.1 and 4.2.

Two scalar limits. Put $\theta_k = \theta(y_k)$. The monic-projection identity gives

$$\theta_k = \theta_{k+1}\langle y_k, y_{k+2}\rangle, \qquad \|y_{k+2} - y_k\|^2 = 2\left(1 - \frac{\theta_k}{\theta_{k+1}}\right).$$

Thus θ_k is nondecreasing and converges to a number $\tau > 0$, and the squared two-step displacements are summable. Independently, if $E_k = \|r_k\|_{A^{-1}}^2$ and $\rho_k = E_{k+1}/E_k$, the cross identity $\langle r_k, r_{k+2}\rangle_{A^{-1}} = E_{k+1}$ yields

$$0 < \rho_k \le \rho_{k+1} \le \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 < 1.$$

Consequently $\rho_k \to \rho_\infty \in (0,1)$. The same argument proves that no residual vanishes and that grade at least three persists, so the normalized orbit is defined for all time. The second scalar limit is also needed below to fix the constant coefficient of the limiting quartic when only three interpolation nodes are present; see Lemmas 4.3 and 4.4.

Omega-limit rigidity and the grade-three case. Let $\Omega = \omega(y_{2k})$. Compactness of the unit sphere and vanishing two-step displacement make Ω nonempty, compact, and connected. Continuity of the monic map shows that every $y \in \Omega$ satisfies

$$T^2 y = y, \qquad q_y(t) := \pi_{Ty}(t)\pi_y(t), \qquad q_y(A)y = \beta y, \qquad \beta := \tau^2.$$

Thus every active eigenvalue of y is a root of the monic quartic $q_y - \beta$, while $\tau > 0$ rules out grade below three. Every omega-limit point therefore has grade three or four.

The energy-ratio limit removes the remaining ambiguity among these quartics. In fact,

$$q_{2k} := \pi_{y_{2k+1}}\pi_{y_{2k}} \longrightarrow q_\star, \qquad q_\star = \pi_\star^{\mathrm{o}}\pi_\star^{\mathrm{e}}, \qquad q_\star(0) = \frac{\beta}{\rho_\infty},$$

and the even and odd quadratic factors converge separately. The reason is that $q_y(\lambda) = \beta$ on the support of y, and the displayed constant coefficient supplies the fourth condition on a three-point support. There are only finitely many possible supports, so connectedness forces one quartic throughout Ω. With

$$I_\star = \{i : q_\star(\lambda_i) = \beta\}, \qquad 3 \le |I_\star| \le 4, \qquad \sum_{k=0}^{\infty}\sum_{j \notin I_\star} y_{2k,j}^2 < \infty,$$

the even-parity mass away from the quartic equality set is summable. If every point of Ω has grade three, a support and a quadratic divisor of $q_\star$ determine the squared coordinates through the two orthogonality equations and normalization. Only finitely many supports, divisors, and coordinate-sign choices remain. Hence the connected set Ω is finite and therefore a singleton. This proves the grade-three regime, including grade-three boundary accumulation when $|I_\star| = 4$; see Lemmas 4.5 to 4.8.

Corrected four-node transport. It remains to consider an omega-limit point positive on four equality nodes $\xi_1 < \xi_2 < \xi_3 < \xi_4$. Put

$$L(t) = \prod_{i=1}^{4}(t - \xi_i).$$

Then $q_\star = L+\beta$. On the exact four-node face, Lagrange interpolation supplies a barycentric parameter ζ; when the cubic coefficient defect vanishes, two normalized steps return the corresponding state. The resulting family of exact two-cycles is the source of the possible drift. Coordinates outside $I_\star$ cannot simply be deleted, because they may remain nonzero at finite times and enter the orthogonality equations.

For the full orbit, write $p_n = \pi_{y_n}$, $H_n = \theta_n^2$, $q_n = p_n p_{n+1}$, let $v_{n,i}$ be the squared mass at ξ_i, and let $w_{n,j} = y_{n,j}^2$ outside $I_\star$. With

$$\ell_i(x) = \frac{L(x)}{L'(\xi_i)(x - \xi_i)},$$

the corrected coordinate ζ_n is defined by the four identities

$$v_{n,i} p_n(\xi_i) + \sum_{j \notin I_\star} w_{n,j} p_n(\lambda_j) \ell_i(\lambda_j) = H_n \frac{\xi_i - \zeta_n}{L'(\xi_i)} \qquad (1 \le i \le 4).$$

The Lagrange correction is chosen precisely so that all outside coordinates are retained and the resulting transport law is exact. Set

$$\delta_n = [t^3]q_n + \sum_{i=1}^{4} \xi_i, \qquad \epsilon_n = \sum_{j \notin I_\star} w_{n,j}.$$

Here $[t^3]q_n$ denotes the cubic coefficient of q_n. The outside-coordinate terms collect into an explicit polynomial F_n whose coefficients are $O(\epsilon_n)$, and the main transport identity becomes

$$\zeta_{n+1} - \zeta_n = \frac{\delta_n L(\zeta_n) - F_n(\zeta_n)/H_n}{H_{n+1}}.$$

An accompanying exact defect-transport identity relates δ_{n+1} to δ_n. These two identities are the mechanism that turns decay of the transverse mass into convergence along the four-node family; see Lemmas 4.9 and 4.15.

Tail control, resonance exclusion, and collapse. For a surviving outside coordinate λ_j, the limiting two-step multiplier is $q_\star(\lambda_j)/\beta$. A modulus greater than one is incompatible with the already proved decay, +1 would put j in $I_\star$, and a strict modulus below one gives geometric decay. The only noncontracting possibility is therefore the signed multiplier -1. For $\Delta_n = \|y_{n+2} - y_n\|^2$, the exact Lyapunov tail

$$\tau - \theta_n = \frac{1}{2}\sum_{\ell=n}^{\infty} \theta_{\ell+1}\Delta_\ell$$

charges the mass of every such resonant coordinate. Together with the strictly contracting coordinates and the forced transport identities, it gives constants $C > 0$ and $0 < \gamma < 1$ such that

$$\sum_{\ell=n}^{\infty}(\Delta_\ell + \epsilon_\ell) \le C(\tau - \theta_n + \gamma^n), \qquad |\delta_n| \le C(\tau - \theta_n + \gamma^n).$$

Endpoint estimates are essential for the defect bound when ζ_n approaches ξ_2 or ξ_3 and an active squared coordinate tends to zero.

For a surviving resonant node $\lambda = \lambda_j$, the proof introduces the exact logarithmic cocycle

$$\Psi_{n,\lambda} = \log w_{n,j} + \sum_{i=1}^{4} \ell_i(\lambda) \log v_{n,i}.$$

Writing $e_n = \beta/(\theta_n \theta_{n+1}) - 1$ and $b_{n,i} = q_n(\xi_i)/(\theta_n \theta_{n+1}) - 1$, cubic interpolation cancels every term linear in the active errors and yields

$$\frac{\Psi_{n+2,\lambda} - \Psi_{n,\lambda}}{2} = 2e_n + O\left(e_n^2 + \sum_{i=1}^{4} b_{n,i}^2\right).$$

The Lyapunov and endpoint bounds make the negative parts of these increments summable. Thus the cocycle is bounded below on each parity, contradicting $w_{n,j} \to 0$ along a subsequence approaching a four-positive omega-limit point. No multiplier--1 coordinate survives.

All remaining outside mass now decays geometrically, and the exact defect transport reduces to the parabolic estimate

$$|\delta_{n+1} - \delta_n| \le C\delta_n^2 + C\gamma^n.$$

The sign-or-summability dichotomy says that δ_n is either summable or eventually one-signed. In the first case the barycentric transport gives finite total variation of ζ_n; in the second, $L \ge 0$ on $[\xi_2, \xi_3]$ makes the adverse increments summable. Hence ζ_n converges in either case. The corrected barycentric identities then recover convergence of the four active squared coordinates, and the positive limiting two-step multipliers stabilize their signs. Thus y_{2k} converges. Continuity of T gives convergence of y_{2k+1} to its companion point in the T-two-cycle, both limits have grade three or four, and the spectral isometry transfers the conclusion back to the original residual directions. This completes Theorem 1.1; see Lemmas 4.19 and 4.21 to 4.25.

4 Proof of the Main Theorem

4.1 Polynomial form of one restart

The normalized monic-quadratic map from Definition 2.2 is the spectral form of one restarted CG step. The following algorithm gives its exact-arithmetic implementation.

Restarted-CG(2): Perform conjugate-gradient restarts of length two.
Input: A symmetric positive definite matrix $A \in \mathbb{R}^{n\times n}$, vectors $b, x_0 \in \mathbb{R}^n$, and $r_0 = b - Ax_0$ with $d(A, r_0) \geq 3$.
Output: Sequences $\{x_k\}_{k\geq 0}$, $\{r_k\}_{k\geq 0}$, and $\{y_k\}_{k\geq 0}$, where $r_k = b - Ax_k$ and $y_k = r_k/\|r_k\|$.
Guarantee: At restart k, x_{k+1} is the unique minimizer of f over $x_k + \mathcal{K}_2(A, r_k)$, and $r_{k+1} \perp \mathcal{K}_2(A, r_k)$.
$y_0 \leftarrow r_0/\|r_0\|$. For $k = 0, 1, 2, \ldots$, perform:

1. At restart k, compute the named vectors Ar_k and A^2r_k.
2. Solve for $\gamma_{k,0}, \gamma_{k,1} \in \mathbb{R}$ so that
$$x_{k+1} = x_k + \gamma_{k,0}r_k + \gamma_{k,1}Ar_k$$
minimizes f on $x_k + \mathcal{K}_2(A, r_k)$.
3. Set
$$r_{k+1} = r_k - \gamma_{k,0}Ar_k - \gamma_{k,1}A^2r_k, \qquad y_{k+1} = \frac{r_{k+1}}{\|r_{k+1}\|}.$$

Lemma 4.1 (Residual-polynomial form and implementation cost)**.** *For every nonzero residual r_k of* Restarted-CG(2) *having grade at least three, there is a unique degree-two polynomial P_k with $P_k(0) = 1$ such that*
$$r_{k+1} = P_k(A)r_k, \qquad r_{k+1} \perp \operatorname{span}\{r_k, Ar_k\}.$$
Writing $y_k = r_k/\|r_k\|$, one has $P_k = P_{y_k}$ and $y_{k+1} = T(y_k)$. If a multiplication by A costs T_A, then a restart uses $2T_A + O(n)$ arithmetic operations; for an explicit dense matrix the cost is $O(n^2)$.

Proof. Since $x_{k+1} - x_k \in \operatorname{span}\{r_k, Ar_k\}$, the residual has the form
$$r_{k+1} = r_k - A(x_{k+1} - x_k) = (I + a_kA + b_kA^2)r_k.$$
First-order optimality for the strictly convex function f says exactly that $r_{k+1} \perp \operatorname{span}\{r_k, Ar_k\}$; conversely, these two orthogonality equations characterize the minimizer. Their moment matrix is nonsingular when the grade is at least three, so the constant-normalized polynomial $P_k(t) = 1 + a_kt + b_kt^2$ is unique. Also $b_k \neq 0$: otherwise the two equations for $1 + a_kt$ would imply $\mu_0(y_k)\mu_2(y_k) = \mu_1(y_k)^2$, contrary to strict Cauchy–Schwarz. Hence P_k has degree two.

Multiplying P_k by a nonzero scalar gives the unique monic quadratic orthogonal to $1, t$ for ν_{y_k}. It is therefore π_{y_k}. Its constant term is positive, so $P_k = \pi_{y_k}/c_0(y_k)$. Division by this positive scalar does not change the normalized direction, which proves $y_{k+1} = T(y_k)$ and the algorithm's stated output guarantee.

Once Ar_k and A^2r_k have been computed, the two moment equations for $\gamma_{k,0}, \gamma_{k,1}$, the constant-size solve, and the vector updates take $O(n)$ scalar operations. This proves the resource bound. ☐

The polynomial representation permits a lossless reduction to the distinct eigenvalues that occur in the initial residual.

Lemma 4.2 (Reduction to distinct spectral coordinates)**.** *Let $\Lambda(A, r_0) = \{\lambda_1 < \cdots < \lambda_m\}$, where $m = d(A, r_0)$. There are fixed orthonormal eigenvectors u_i and scalars $z_{k,i}$ such that*
$$Au_i = \lambda_iu_i, \qquad r_k = \sum_{i=1}^{m} z_{k,i}u_i$$

for every k. Under the isometry $\sum_i z_i u_i \mapsto z$, the normalized residual iteration is exactly Equation (1) with $A = \operatorname{diag}(\lambda_1, \dots, \lambda_m)$. Thus convergence in reduced coordinates is equivalent to convergence in the original space, including when the original matrix has repeated eigenvalues.

Proof. Let $\Pi_i = \Pi_{\lambda_i}$ and set $u_i = \Pi_i r_0 / \|\Pi_i r_0\|$. Every residual is a polynomial in A applied to r_0 by Lemma 4.1; hence $\Pi_i r_k$ is always a scalar multiple of u_i. The displayed expansion follows. The stated map is an isometry intertwining A with the diagonal matrix of distinct active eigenvalues, and so it preserves all moment equations, projection polynomials, norms, and normalized directions. $\square$

From now on, we work in these reduced coordinates. Thus

$$A = \operatorname{diag}(\lambda_1, \dots, \lambda_m), \qquad 0 < \lambda_1 < \dots < \lambda_m.$$

Coordinates that become zero remain zero.

4.2 Two monotone quantities

The first Lyapunov quantity is the norm of the monic quadratic projection. Its exact two-step identity also controls the displacement of a fixed parity of normalized residuals.

Lemma 4.3 (Monic-projection Lyapunov identity)**.** *Let $y_{k+1} = T(y_k)$ and $\theta_k = \theta(y_k)$. As long as all iterates have grade at least three,*

$$\theta_k = \theta_{k+1} \langle y_k, y_{k+2} \rangle. \tag{3}$$

Consequently, $\{\theta_k\}$ is nondecreasing and converges to a finite $\tau > 0$, and

$$\|y_{k+2} - y_k\|^2 = 2\left(1 - \frac{\theta_k}{\theta_{k+1}}\right) \longrightarrow 0. \tag{4}$$

In fact, $\sum_{k \ge 0} \|y_{k+2} - y_k\|^2 < \infty$.

Proof. Put $\pi_k = \pi_{y_k}$. Since $\pi_k(A) y_k \perp \operatorname{span}\{y_k, A y_k\}$, every monic quadratic q satisfies

$$\begin{aligned} \theta_k^2 &= \langle \pi_k(A) y_k, A^2 y_k \rangle \\ &= \langle \pi_k(A) y_k, q(A) y_k \rangle = \theta_k \langle y_{k+1}, q(A) y_k \rangle. \end{aligned}$$

Symmetry of A and division by $\theta_k > 0$ give $\theta_k = \langle y_k, q(A) y_{k+1} \rangle$. Taking $q = \pi_{k+1}$ proves Equation (3). The inner product in that identity is positive and at most one, so $0 < \theta_k \le \theta_{k+1}$. Taking t^2 as a trial polynomial in Equation (2) yields $\theta_k \le \|A\|^2$; hence $\theta_k \to \tau > 0$.

Both vectors in Equation (4) have unit norm, and Equation (3) gives the formula. Finally, $1 - a/b \le \log(b/a)$ for $0 < a \le b$, so

$$\sum_{k=0}^{N} \|y_{k+2} - y_k\|^2 \le 2 \sum_{k=0}^{N} \log \frac{\theta_{k+1}}{\theta_k} = 2 \log \frac{\theta_{N+1}}{\theta_0}.$$

The right side is uniformly bounded. $\square$

The second monotone quantity uses the constant-normalized residual polynomials. Besides supplying another limiting scalar, it proves that the iteration never terminates under the grade hypothesis.

Lemma 4.4 (Energy-ratio monotonicity and grade persistence)**.** *Assume $d(A, r_0) \geq 3$. Every residual generated by* Restarted-CG(2) *is nonzero and has grade at least three. With*

$$E_k = \|r_k\|_{A^{-1}}^2, \qquad \rho_k = \frac{E_{k+1}}{E_k},$$

one has

$$0 < \rho_k \leq \rho_{k+1} \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \right)^2 < 1, \tag{5}$$

where $\lambda_{\min} = \min \Lambda(A, r_0)$ and $\lambda_{\max} = \max \Lambda(A, r_0)$. In particular, $\rho_k \to \rho_\infty$ for some $\rho_\infty \in (0, 1)$.

Proof. For a nonzero residual, the minimizing step has a residual polynomial of degree at most two with constant term one; if the residual has grade at most two, one may choose such a polynomial to annihilate it. Let $P_k(0) = 1$ denote the minimizing polynomial. Whenever $r_{k+1} \neq 0$, orthogonality gives

$$\langle r_k, r_{k+2} \rangle_{A^{-1}} = E_{k+1}. \tag{6}$$

Indeed,

$$\begin{aligned} \langle r_k, r_{k+2} \rangle_{A^{-1}} &= \langle P_{k+1}(A) r_k, r_{k+1} \rangle_{A^{-1}} \\ &= E_{k+1} + \langle (P_{k+1} - P_k)(A) r_k, r_{k+1} \rangle_{A^{-1}}. \end{aligned}$$

Since $P_{k+1}(0) = P_k(0) = 1$, their difference is $th(t)$ for a polynomial h of degree at most one. The final term is $\langle h(A) r_k, r_{k+1} \rangle = 0$, proving Equation (6).

If r_k has grade at least three, then $r_{k+1} \neq 0$, because a degree-two polynomial cannot annihilate it. If r_{k+1} had grade at most two, the following minimizing step would give $r_{k+2} = 0$, whereas Equation (6) would give $0 = E_{k+1} > 0$. Induction proves nonvanishing and grade at least three for every residual.

Cauchy–Schwarz applied to Equation (6) yields $E_{k+1}^2 \leq E_k E_{k+2}$, which is $\rho_k \leq \rho_{k+1}$. The energy decreases strictly at a nonzero step: the admissible direction r_k has directional derivative $\langle \nabla f(x_k), r_k \rangle = -\|r_k\|^2 < 0$, so x_k is not the affine-space minimizer. Thus $0 < \rho_k < 1$. Since the CG minimizer also minimizes the A^{-1}-norm of the residual over degree-two residual polynomials with constant term one, it does at least as well as

$$P(t) = 1 - \frac{2t}{\lambda_{\min} + \lambda_{\max}}.$$

On the initial cyclic subspace,

$$|P(\lambda)| \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

Evaluation componentwise in the A^{-1}-norm proves Equation (5), and monotone convergence gives the claimed limit. ☐

4.3 Omega-limit structure and limiting quartics

We next convert the two scalar limits into a rigid description of every subsequential limit.

Lemma 4.5 (Two-cycle structure of the omega-limit set)**.** *Let $\Omega = \omega(y_{2k})$. Then Ω is nonempty, compact, and connected. Every $y \in \Omega$ has grade three or four and satisfies*

$$T^2 y = y, \qquad \theta(y) = \theta(Ty) = \tau. \tag{7}$$

If

$$q_y(t) = \pi_{Ty}(t)\pi_y(t), \qquad \beta = \tau^2,$$

then

$$q_y(A)y = \beta y. \tag{8}$$

Proof. The sequence lies on the compact unit sphere, and Lemma 4.3 gives $\|y_{2k+2} - y_{2k}\| \to 0$. The limit-set facts recorded in Section 2 therefore show that Ω is nonempty, compact, and connected.

Take $y_{2k_j} \to y \in \Omega$. The limit $\theta(y_{2k_j}) \to \tau > 0$ rules out grade at most two. Indeed, a fixed monic quadratic g vanishing on such a support would satisfy

$$0 < \theta(y_{2k_j}) \le \|g(A)y_{2k_j}\| \longrightarrow \|g(A)y\| = 0,$$

contradicting Equation (2). Hence the moment system is nonsingular at y, and continuity gives

$$\theta(y) = \tau, \qquad y_{2k_j+1} \to Ty.$$

The same argument at Ty gives grade at least three, $\theta(Ty) = \tau$, and continuity there. Thus $y_{2k_j+2} \to T^2 y$; together with $y_{2k_j+2} - y_{2k_j} \to 0$, this proves Equation (7). Finally,

$$q_y(A)y = \pi_{Ty}(A)\big(\theta(y)Ty\big) = \theta(y)\theta(Ty)T^2y = \beta y.$$

Every eigenvalue in the support of y is therefore a root of the nonzero monic quartic $q_y - \beta$. The support has between three and four points. ☐

The next lemma makes the two-step polynomial independent of the chosen omega-limit point. The energy-ratio limit fixes the otherwise undetermined constant coefficient on a three-point support.

Lemma 4.6 (Convergence of the two-step quartic)**.** *There is a monic quartic*

$$q_\star(t) = t^4 + \alpha_3^\star t^3 + \alpha_2^\star t^2 + \alpha_1^\star t + \alpha_0^\star$$

such that the coefficients of

$$q_{2k}(t) = \pi_{y_{2k+1}}(t)\pi_{y_{2k}}(t)$$

converge to those of $q_\star$. There are also monic quadratics $\pi_\star^{\mathrm{e}}$ and $\pi_\star^{\mathrm{o}}$ such that

$$\pi_{y_{2k}} \longrightarrow \pi_\star^{\mathrm{e}}, \qquad \pi_{y_{2k+1}} \longrightarrow \pi_\star^{\mathrm{o}}, \qquad q_\star = \pi_\star^{\mathrm{o}}\pi_\star^{\mathrm{e}}$$

coefficientwise. Moreover,

$$\alpha_0^\star = \frac{\beta}{\rho_\infty}. \tag{9}$$

Proof. For $y \in \Omega$, write

$$q_y(t) = t^4 + \alpha_3(y)t^3 + \alpha_2(y)t^2 + \alpha_1(y)t + \alpha_0(y).$$

The coefficients are continuous in y, and Lemma 4.5 gives $q_y(\lambda_i) = \beta$ on $\operatorname{supp}(y)$.

Because $P_y = \pi_y/\pi_y(0)$, the actual two-step residual polynomial at the two-cycle y is $q_y/\alpha_0(y)$. By Equation (8), it multiplies y by the positive scalar $\beta/\alpha_0(y)$. Define the energy contraction functional

$$\mathcal{E}(z) = \frac{\|P_z(A)z\|_{A^{-1}}^2}{\|z\|_{A^{-1}}^2}.$$

Along a subsequence converging to y, continuity and Lemma 4.4 give $\mathcal{E}(y) = \mathcal{E}(Ty) = \rho_\infty$. Hence

$$\left(\frac{\beta}{\alpha_0(y)}\right)^2 = \rho_\infty^2.$$

All quantities are positive, so $\alpha_0(y) = \beta/\rho_\infty$ throughout Ω.

For a fixed four-element support, the four equations $q_y(\lambda_i) = \beta$ uniquely determine the four lower coefficients of the monic quartic. For a fixed three-element support $\{i_1, i_2, i_3\}$, the constant coefficient is already fixed, and the remaining system is nonsingular because

$$\det[\lambda_{i_a}^b]_{a,b=1}^3 = \lambda_{i_1}\lambda_{i_2}\lambda_{i_3} \prod_{1\le a<b\le 3} (\lambda_{i_b} - \lambda_{i_a}) \neq 0.$$

Only finitely many three- and four-element supports occur. Thus the continuous coefficient map has finite image on the connected set Ω, so it is constant there; call its value $q_\star$. Compactness then promotes subsequential convergence to convergence of the full sequence q_{2k}.

For every $y \in \Omega$, $q_\star = \pi_{Ty}\pi_y$, so π_y is a monic quadratic divisor of $q_\star$. A fixed quartic has finitely many monic divisors, whereas $y \mapsto \pi_y$ is continuous on connected Ω. Therefore π_y is one fixed polynomial $\pi_\star^{\mathrm{e}}$ on Ω, and compactness gives $\pi_{y_{2k}} \to \pi_\star^{\mathrm{e}}$. Since $q_{2k} = \pi_{y_{2k+1}}\pi_{y_{2k}}$ and all factors are monic, coefficient comparison yields $\pi_{y_{2k+1}} \to q_\star/\pi_\star^{\mathrm{e}} =: \pi_\star^{\mathrm{o}}$. □

The limiting quartic makes every coordinate outside its equality set summably small, even if its limiting two-step multiplier is -1.

Lemma 4.7 (Summability away from the quartic equality set)**.** *Let*

$$I_\star = \{i : q_\star(\lambda_i) = \beta\}.$$

For every $i \notin I_\star$,

$$\sum_{k=0}^{\infty} y_{2k,i}^2 < \infty. \tag{10}$$

Consequently, the total squared mass outside $I_\star$ is summable.

Proof. Put $z_k = y_{2k}$. Two applications of Equation (1) give

$$z_{k+1,i} = s_{k,i} z_{k,i}, \qquad s_{k,i} = \frac{q_{2k}(\lambda_i)}{\theta_{2k}\theta_{2k+1}}.$$

By Lemmas 4.3 and 4.6,

$$s_{k,i} \longrightarrow s_{\star,i} = \frac{q_\star(\lambda_i)}{\beta}.$$

If $i \notin I_\star$, then $s_{\star,i} \neq 1$. For some $\delta_i > 0$ and all sufficiently large k, $|s_{k,i} - 1| \geq \delta_i$, and therefore

$$|z_{k+1,i} - z_{k,i}|^2 \geq \delta_i^2 |z_{k,i}|^2.$$

The left side is summable by Lemma 4.3. This proves Equation (10); summing over the finite spectrum gives the final assertion. □

4.4 Collapse on three spectral coordinates

If all omega-limit points have grade three, the fixed quartic and the two orthogonality equations leave only finitely many possible directions.

Lemma 4.8 (Collapse of the three-coordinate regime)**.** *For*

$$I_\star = \{i : q_\star(\lambda_i) = \beta\}$$

one has $3 \le |I_\star| \le 4$*. If* $|I_\star| = 3$*, then* $\{y_{2k}\}$ *converges. More generally, the same conclusion holds if every vector in* Ω *has grade three. In either case, there is a vector* $y_\star$ *such that*

$$y_{2k} \to y_\star, \qquad y_{2k+1} \to T(y_\star).$$

Proof. Every $y \in \Omega$ has at least three support points, all in $I_\star$, by Lemmas 4.5 and 4.6. Since $q_\star - \beta$ is a nonzero monic quartic, $|I_\star| \le 4$. Thus $|I_\star| = 3$ implies that every vector in Ω has grade three.

Assume more generally that all vectors in Ω have grade three. For each such vector,

$$q_\star = \pi_{Ty}\pi_y.$$

Thus π_y is one of finitely many monic quadratic divisors of $q_\star$. Fix a possible support $S = \{i_1, i_2, i_3\} \subseteq I_\star$ and a possible divisor π. For a vector $y \in \Omega$ with this support and $\pi_y = \pi$, put

$$w_a = y_{i_a}^2 > 0, \qquad v_a = w_a \pi(\lambda_{i_a}).$$

Orthogonality and normalization become

$$\sum_{a=1}^{3} v_a = 0, \qquad \sum_{a=1}^{3} \lambda_{i_a} v_a = 0, \qquad \sum_{a=1}^{3} w_a = 1.$$

None of $\pi(\lambda_{i_a})$ is zero, because its product with the other factor equals $\beta > 0$. The nullspace of the first two equations is one-dimensional, spanned by

$$\big(\lambda_{i_3} - \lambda_{i_2}, -(\lambda_{i_3} - \lambda_{i_1}), \lambda_{i_2} - \lambda_{i_1}\big).$$

Hence the positive weights $w_a = v_a/\pi(\lambda_{i_a})$ are unique after normalization. There are finitely many supports, divisors, and coordinate sign choices, so Ω is finite. It is nonempty and connected by Lemma 4.5, hence is a singleton $\{y_\star\}$. Precompactness then gives $y_{2k} \to y_\star$, and continuity of T gives the odd convergence. $\square$

4.5 Four-node barycentric algebra

The remaining case has four equality nodes. We first isolate the exact one-step algebra on a four-node face; its barycentric parameter is the coordinate later used to control drift along a continuum of two-cycles. For a polynomial h, the notation $[t^j]h(t)$ denotes the coefficient of t^j in h.

Lemma 4.9 (Exact four-node barycentric transport)**.** *Fix* $0 < \xi_1 < \xi_2 < \xi_3 < \xi_4$ *and positive weights* $w_1, \dots, w_4$ *summing to one. Let* π *be the monic quadratic satisfying*

$$\sum_i w_i \pi(\xi_i) = 0, \qquad \sum_i \xi_i w_i \pi(\xi_i) = 0,$$

put $H = \sum_i w_i \pi(\xi_i)^2$, *and define*

$$w_i' = \frac{w_i \pi(\xi_i)^2}{H}.$$

Let π' *and* H' *be the analogous monic quadratic and squared projection norm for* w'. *Set*

$$L(t) = \prod_{i=1}^{4}(t - \xi_i), \qquad D_i = L'(\xi_i), \qquad q(t) = \pi(t)\pi'(t).$$

There are unique real numbers ζ, ζ' *such that*

$$w_i \pi(\xi_i) = H\frac{\xi_i - \zeta}{D_i}, \qquad w_i'\pi'(\xi_i) = H'\frac{\xi_i - \zeta'}{D_i}. \tag{11}$$

If

$$\delta = [t^3]q(t) + \sum_{i=1}^{4} \xi_i,$$

then

$$q(t) = L(t) + H' + \delta\,\frac{L(t) - L(\zeta)}{t - \zeta}, \qquad \zeta' - \zeta = \delta\,\frac{L(\zeta)}{H'}. \tag{12}$$

In particular, $\delta = 0$ *implies that two normalized polynomial steps return every signed vector with squared coordinates* w_i *to itself.*

Proof. The polynomial π cannot vanish at two of the four nodes. Otherwise, the two orthogonality equations on the remaining two distinct nodes force both remaining quantities $w_i\pi(\xi_i)$ to vanish, so π would vanish at all four nodes. Hence w' has at least three positive coordinates, and π' and $H' > 0$ are well defined.

Lagrange interpolation gives, for every polynomial g of degree at most three,

$$[t^3]g(t) = \sum_{i=1}^{4} \frac{g(\xi_i)}{D_i}. \tag{13}$$

The vector $v_i = w_i\pi(\xi_i)$ annihilates the rows $(1, \ldots, 1)$ and $(\xi_1, \ldots, \xi_4)$. By Equation (13), every vector in this two-dimensional nullspace has a unique representation

$$v_i = \frac{a\xi_i + b}{D_i}.$$

Multiplying by $\pi(\xi_i)$ and summing, then applying Equation (13) to $(at+b)\pi(t)$, gives $a = \sum_i w_i\pi(\xi_i)^2 = H$. Writing $b = -H\zeta$ proves the first identity in Equation (11); the second is identical.

The weight update and Equation (11) imply

$$(\xi_i - \zeta)q(\xi_i) = H'(\xi_i - \zeta')$$

at every node. Hence the monic degree-five polynomial

$$(t - \zeta)q(t) - H'(t - \zeta')$$

is divisible by $L(t)$. Comparison of its degree-four coefficient shows that the quotient is $t - \zeta + \delta$, so

$$(t - \zeta)q(t) - H'(t - \zeta') = (t - \zeta + \delta)L(t). \tag{14}$$

Putting $t = \zeta$ gives the second identity in Equation (12); substituting it into Equation (14) and dividing by $t - \zeta$ gives the first.

If $\delta = 0$, then $q = L + H'$, and $q(\xi_i) = H' > 0$. After two weight updates,

$$w_i'' = \frac{w_i q(\xi_i)^2}{HH'} = w_i \frac{H'}{H}.$$

Both w and w'' sum to one, so $H' = H$ and $w'' = w$. The signed two-step multiplier is $q(\xi_i)/\sqrt{HH'} = 1$, proving the last assertion. $\square$

Assume from now on that $|I_\star| = 4$. Lemmas 4.3 and 4.6 give $\theta_n \to \tau > 0$, convergence of the even and odd monic factors, and the limiting quartic $q_\star$, while Lemma 4.7 gives

$$\sum_k \sum_{j \notin I_\star} y_{2k,j}^2 < \infty.$$

We retain the contribution of every coordinate outside the four-node equality set.

Definition 4.10 (Four-point orbit data)**.** *A* four-point orbit *is the reduced diagonal iteration $y_{n+1} = T(y_n)$ in the case $|I_\star| = 4$. For such an orbit, write the four indices as $\iota_1, \dots, \iota_4$, ordered so that*

$$\xi_1 := \lambda_{\iota_1} < \xi_2 := \lambda_{\iota_2} < \xi_3 := \lambda_{\iota_3} < \xi_4 := \lambda_{\iota_4}.$$

For $n \geq 0$, set

$$p_n = \pi_{y_n}, \qquad \theta_n = \theta(y_n), \qquad H_n = \theta_n^2, \qquad q_n = p_n p_{n+1},$$
$$v_{n,i} = y_{n,\iota_i}^2 \quad (1 \leq i \leq 4), \qquad w_{n,j} = y_{n,j}^2 \quad (j \notin I_\star).$$

Let

$$p_{\mathrm{e}} = \pi_\star^{\mathrm{e}}, \qquad p_{\mathrm{o}} = \pi_\star^{\mathrm{o}}, \qquad L(t) = \prod_{i=1}^{4}(t - \xi_i), \qquad D_i = L'(\xi_i),$$

and put

$$\beta = \tau^2, \qquad \delta_n = [t^3]q_n + \sum_{i=1}^{4} \xi_i, \qquad \epsilon_n = \sum_{j \notin I_\star} w_{n,j}.$$

The quadratic factors converge separately by Lemma 4.6; on odd indices they are merely interchanged. Thus $q_n \to q_\star$ along the full sequence. Since $q_\star - \beta$ and L are monic quartics with the same four roots,

$$q_\star(t) = L(t) + \beta = p_{\mathrm{o}}(t) p_{\mathrm{e}}(t). \tag{15}$$

In particular, $H_n \to \beta$ and $\delta_n \to 0$.

Definition 4.11 (Tail and resonance data)**.** *Define*

$$K_n = \sqrt{H_n H_{n+1}} = \theta_n \theta_{n+1}, \qquad \Delta_n = \|y_{n+2} - y_n\|^2,$$
$$\sigma_n = \frac{H_{n+1}}{K_n} - 1, \qquad e_n = \frac{\beta}{K_n} - 1, \qquad b_{n,i} = \frac{q_n(\xi_i)}{K_n} - 1.$$

Choose N_0 beyond the first zero time of every coordinate that is ever annihilated. Among indices $j \notin I_\star$ for which $w_{n,j} > 0$ for all $n \geq N_0$, define

$$\mathcal{R} = \{j : q_\star(\lambda_j) = -\beta\}, \qquad \mathcal{S} = \{j : |q_\star(\lambda_j)| < \beta\},$$

and put

$$\epsilon_n^{\mathrm{R}} = \sum_{j \in \mathcal{R}} w_{n,j}, \qquad \epsilon_n^{\mathrm{S}} = \sum_{j \in \mathcal{S}} w_{n,j}.$$

The following division functional extracts the scalar coefficient transported from one quadratic factor to the next.

Definition 4.12 (Division functional)**.** *For $z \in \mathbb{R}$, set*

$$R_z(t) = \frac{L(t) - L(z)}{t - z}.$$

For a monic quadratic p, let $X(p, z)$ be the coefficient of t in the remainder of $R_z(t)$ upon division by $p(t)$.

If

$$L(t) = t^4 - St^3 + e_2 t^2 - e_3 t + e_4, \qquad p(t) = t^2 + at + b,$$

then direct division gives

$$X(p, z) = z^2 - (a + S)z + a^2 + aS - b + e_2. \tag{16}$$

Before deriving the forced transport law, we transfer the even-index summability from Lemma 4.7 to both parities.

Lemma 4.13 (Full-parity outside-mass summability)**.** *Every four-point orbit satisfies*

$$\sum_{n=0}^{\infty} \epsilon_n < \infty. \tag{17}$$

Proof. By Lemma 4.7, $\sum_k \epsilon_{2k} < \infty$. The coefficients of p_{2k} are bounded, the spectrum is finite, and H_{2k} is bounded away from zero. The exact one-step squared-weight update therefore gives, for all sufficiently large k,

$$\epsilon_{2k+1} = \sum_{j \notin I_\star} \frac{w_{2k,j} p_{2k}(\lambda_j)^2}{H_{2k}} \leq C\epsilon_{2k}.$$

Adding the finitely many initial terms proves the claim. $\square$

The outside coordinates cannot simply be discarded: a multiplier tending to -1 need not decay geometrically. Lagrange correction incorporates all of them into an exact barycentric identity.

Definition 4.14 (Corrected barycentric coordinate)**.** *For $x \notin \{\xi_1, \ldots, \xi_4\}$, define*

$$\ell_i(x) = \frac{L(x)}{D_i(x - \xi_i)} \qquad (1 \leq i \leq 4).$$

Put

$$a_{n,i} = v_{n,i} p_n(\xi_i) + \sum_{j \notin I_\star} w_{n,j} p_n(\lambda_j) \ell_i(\lambda_j).$$

The corrected barycentric coordinate ζ_n is the real number satisfying

$$a_{n,i} = H_n \frac{\xi_i - \zeta_n}{D_i} \qquad (1 \leq i \leq 4). \tag{18}$$

Lemma 4.15 (Forced barycentric transport)**.** *The coordinate ζ_n in Definition 4.14 exists and is unique. Define*

$$F_n(t) = \sum_{j \notin I_\star} w_{n,j} p_n(\lambda_j) L(\lambda_j) \frac{q_n(\lambda_j) - q_n(t)}{\lambda_j - t},$$

where the divided difference is interpreted by polynomial continuation at $t = \lambda_j$. Then

$$(t - \zeta_n) q_n(t) + \frac{F_n(t)}{H_n} - H_{n+1}(t - \zeta_{n+1}) = (t - \zeta_n + \delta_n) L(t), \tag{19}$$

$$\zeta_{n+1} - \zeta_n = \frac{\delta_n L(\zeta_n) - F_n(\zeta_n)/H_n}{H_{n+1}}, \tag{20}$$

$$q_n(t) = L(t) + H_{n+1} + \delta_n \frac{L(t) - L(\zeta_n)}{t - \zeta_n} - \frac{F_n(t) - F_n(\zeta_n)}{H_n(t - \zeta_n)}. \tag{21}$$

Let

$$C_n(t) = \frac{F_n(t) - F_n(\zeta_n)}{H_n(t - \zeta_n)}, \qquad Y_n(p) = [t] \operatorname{rem}_p C_n(t),$$

again using polynomial continuation. The exact defect transport law is

$$\delta_{n+1} X(p_{n+1}, \zeta_{n+1}) - Y_{n+1}(p_{n+1}) = \delta_n X(p_{n+1}, \zeta_n) - Y_n(p_{n+1}). \tag{22}$$

Uniformly for all sufficiently large n,

$$\|F_n\|_{\text{coeff}} + \|C_n\|_{\text{coeff}} + |Y_n(p_{n+1})| \le C\epsilon_n, \tag{23}$$

where $\|P\|_{\text{coeff}}$ is the largest absolute coefficient of P. Finally, if $p_n(t) = t^2 + c_{1,n}t + c_{0,n}$, then

$$\zeta_n = \sum_{i=1}^{4} \xi_i + c_{1,n} - \sum_{j=1}^{m} \lambda_j y_{n+1,j}^2. \tag{24}$$

Proof. Lagrange interpolation gives

$$\sum_{i=1}^{4} \xi_i^r \ell_i(x) = x^r \qquad (0 \le r \le 3). \tag{25}$$

The two orthogonality equations for p_n imply

$$\sum_i a_{n,i} = 0, \qquad \sum_i \xi_i a_{n,i} = 0.$$

Every four-vector satisfying these equations has the form $(\kappa\xi_i + \nu)/D_i$. Since $t^2 - p_n(t)$ is linear and p_n is orthogonal to every linear polynomial,

$$\sum_i \xi_i^2 a_{n,i} = \sum_{j=1}^{m} \lambda_j^2 y_{n,j}^2 p_n(\lambda_j) = H_n.$$

The identities $\sum_i \xi_i^2/D_i = 0$ and $\sum_i \xi_i^3/D_i = 1$ show that $\kappa = H_n$, proving existence and uniqueness of ζ_n.

The exact squared-weight update is

$$y_{n+1,j}^2 = \frac{y_{n,j}^2 p_n(\lambda_j)^2}{H_n}. \tag{26}$$

Use it in the corrected formula at time $n+1$ and substitute the formula at time n. At each active node ξ_i, this gives

$$H_{n+1}\frac{\xi_i - \zeta_{n+1}}{D_i} = q_n(\xi_i)\frac{\xi_i - \zeta_n}{D_i} + \sum_{j \notin I_\star} \frac{w_{n,j} p_n(\lambda_j)}{H_n}\big(q_n(\lambda_j) - q_n(\xi_i)\big)\ell_i(\lambda_j).$$

Consequently, the left side of Equation (19) vanishes at all four roots of L. It is a monic polynomial of degree five, and comparison of the t^4 coefficient shows that its quotient by L is $t - \zeta_n + \delta_n$. This proves Equation (19). Evaluation at $t = \zeta_n$ gives Equation (20); substitution back gives Equation (21).

Reduce Equation (21) modulo p_{n+1} and take the coefficient of t in the remainder. Both $q_n = p_n p_{n+1}$ and $q_{n+1} = p_{n+1}p_{n+2}$ are divisible by p_{n+1}. Comparing the identities at n and $n+1$ proves Equation (22).

The spectrum is fixed, the coefficients of p_n, q_n are bounded, and H_n is bounded away from zero. Formula Equation (24), proved below, shows that ζ_n is bounded. Thus every coefficient of F_n, C_n, and the indicated remainder is at most $C\epsilon_n$, proving Equation (23).

To prove Equation (24), take the third moment of Equation (18):

$$H_n\left(\sum_i \xi_i - \zeta_n\right) = \sum_{j=1}^m \lambda_j^3 y_{n,j}^2 p_n(\lambda_j).$$

Orthogonality and the displayed form of p_n give

$$\sum_{j=1}^m \lambda_j y_{n,j}^2 p_n(\lambda_j)^2 = \sum_{j=1}^m \lambda_j^3 y_{n,j}^2 p_n(\lambda_j) + c_{1,n}H_n.$$

After division by H_n, the left side is the spectral mean of y_{n+1} by Equation (26), which proves Equation (24). □

4.6 Lyapunov tails in the four-coordinate regime

It remains to collapse Ω when its positive equality set has four nodes. The four-point notation and the exact forced transport identities of Definition 4.10 and lemma 4.15 now apply, because Lemmas 4.3, 4.6 and 4.7 supply their conditional hypotheses.

The scalar defect transport is parabolic. The following elementary dichotomy will later turn geometric forcing into convergence of the barycentric coordinate.

Lemma 4.16 (Parabolic sign-or-summability dichotomy). *Let $u_n \to 0$, and suppose that for constants $C > 0$ and $0 < \gamma < 1$,*

$$|u_{n+1} - u_n| \le C(u_n^2 + \gamma^n)$$

for every sufficiently large n. Then either u_n is eventually one-signed, or $\sum_n |u_n| < \infty$.

Proof. Choose α with $\sqrt{\gamma} < \alpha < 1$. For all sufficiently large n, once $|u_n| \geq \alpha^n$, its sign never changes. Indeed, as long as $|u_k| \geq \alpha^k$, one has $\gamma^k = o(u_k^2)$, and smallness gives

$$|u_{k+1} - u_k| \leq C_1 u_k^2 < \frac{1}{2}|u_k|, \qquad \frac{1}{|u_{k+1}|} \leq \frac{1}{|u_k|} + 2C_1.$$

If $m > n$ were the first later index with $|u_m| < \alpha^m$, iteration would yield

$$|u_m| \geq \frac{1}{\alpha^{-n} + 2C_1(m-n)} > \alpha^m$$

for all sufficiently large n. Writing $\ell = m - n$, the final inequality is equivalent to

$$\alpha^\ell + 2C_1 \ell \alpha^{n+\ell} < 1,$$

uniformly for $\ell \geq 1$, a contradiction. Therefore infinitely many sign changes force $|u_n| < \alpha^n$ eventually, and the latter sequence is summable. □

Definition 4.17 (Four-positive points and surviving coordinates). *An omega-limit point is* four-positive *if its four squared coordinates on $I_\star$ are positive. A coordinate is* surviving *if it is nonzero at arbitrarily large times. Since a zero coordinate remains zero, every surviving coordinate is nonzero at all times after the index N_0 in Definition 4.11.*

A resonant outside coordinate has limiting two-step multiplier -1. Its sign change makes its mass visible in the exact Lyapunov increase.

Lemma 4.18 (Active sign geometry and resonant-tail coercivity). *Assume that Ω contains a four-positive vector. Then*

$$\operatorname{sgn} p_{\mathrm{e}}(\xi_i) = \operatorname{sgn} p_{\mathrm{o}}(\xi_i) = (+,-,-,+)_i. \tag{27}$$

Every cluster value of ζ_n belongs to $[\xi_2, \xi_3]$, and every resonant node λ_j, $j \in \mathcal{R}$, belongs to $(\xi_1, \xi_2) \cup (\xi_3, \xi_4)$. Moreover, for all sufficiently large n,

$$\tau - \theta_n = \frac{1}{2}\sum_{\ell=n}^{\infty} \theta_{\ell+1}\Delta_\ell \geq c \sum_{\ell=n}^{\infty} \epsilon_\ell^{\mathrm{R}} \tag{28}$$

for a constant $c > 0$.

Proof. The four-positive point and its one-step image make p_{e} and p_{o} orthogonal quadratics for positive weights on all four active nodes. Each has two simple real roots in (ξ_1, ξ_4). To see this, if such a quadratic had fewer than two sign-changing roots in that interval, let g be the product of the factors $t - r$ over its sign-changing roots. Then $\deg g \leq 1$, while the product of the quadratic with g has one nonzero sign on all four nodes, contradicting orthogonality to every linear polynomial.

By Equation (15), none of these roots lies outside $[\xi_1, \xi_4]$ or in the middle gap, where $L + \beta > 0$. If one factor placed both roots in the same outer gap, it would be positive at all four active nodes, contradicting its orthogonality to 1. Each factor therefore has one root in each outer gap and has the sign pattern Equation (27).

Formula Equation (24) makes ζ_n bounded. Along a convergent subsequence, $\epsilon_n \to 0$ by Lemma 4.13, so the correction in Equation (18) vanishes. Passing to the limit and using Equation (27) and nonnegativity of the active weights forces every cluster value of ζ_n into $[\xi_2, \xi_3]$. For $j \in \mathcal{R}$,

$$L(\lambda_j) = q_\star(\lambda_j) - \beta = -2\beta < 0,$$

and the sign of the four-root polynomial L places λ_j in an outer gap.

Orthogonality of p_n to every linear polynomial gives

$$\sum_{j=1}^{m} y_{n,j}^2 p_n(\lambda_j) p_{n+1}(\lambda_j) = \sum_{j=1}^{m} y_{n,j}^2 p_n(\lambda_j)^2 = H_n.$$

Consequently,

$$\langle y_n, y_{n+2} \rangle = \frac{\theta_n}{\theta_{n+1}}, \qquad \theta_{n+1} - \theta_n = \frac{\theta_{n+1}}{2}\Delta_n. \tag{29}$$

Summation from n to infinity proves the equality in Equation (28). For $j \in \mathcal{R}$,

$$\frac{y_{\ell+2,j}}{y_{\ell,j}} = \frac{q_\ell(\lambda_j)}{K_\ell} \longrightarrow -1.$$

The finite resonant set thus satisfies

$$|y_{\ell+2,j} - y_{\ell,j}|^2 \geq w_{\ell,j}$$

for all sufficiently large ℓ. Since $\theta_{\ell+1} \to \tau > 0$, substitution into the exact tail identity proves the lower bound. $\square$

The resonant mass is now charged to the Lyapunov tail, while strict-gap coordinates decay geometrically. Together these estimates control every outside coordinate.

Lemma 4.19 (Lyapunov control of displacement and outside-mass tails)**.** *If Ω contains a four-positive vector, then there are constants $C > 0$ and $0 < \gamma < 1$ such that, for all sufficiently large n,*

$$\sum_{\ell=n}^{\infty} \Delta_\ell + \sum_{\ell=n}^{\infty} \epsilon_\ell \leq C\big(\tau - \theta_n + \gamma^n\big). \tag{30}$$

Moreover,

$$\sigma_n = \frac{\Delta_n}{2 - \Delta_n} \leq C\Delta_n, \qquad e_n \geq \frac{\tau - \theta_n}{\tau}. \tag{31}$$

Proof. Equation Equation (29) yields

$$\tau - \theta_n = \frac{1}{2}\sum_{\ell=n}^{\infty} \theta_{\ell+1}\Delta_\ell. \tag{32}$$

Because $\theta_\ell \to \tau > 0$,

$$\sum_{\ell=n}^{\infty} \Delta_\ell \leq C(\tau - \theta_n). \tag{33}$$

Lemma 4.18 likewise gives

$$\sum_{\ell=n}^{\infty} \epsilon_\ell^{\mathrm{R}} \leq C(\tau - \theta_n). \tag{34}$$

For every surviving outside coordinate, the exact recurrence is

$$w_{n+2,j} = w_{n,j}\frac{q_n(\lambda_j)^2}{H_n H_{n+1}}. \tag{35}$$

By Lemma 4.7, $w_{2k,j} \to 0$. If $|q_\star(\lambda_j)| > \beta$, the multiplier in Equation (35) is eventually greater than one on each parity, contradicting this convergence on the even parity and its one-step consequence on the odd parity. Equality $q_\star(\lambda_j) = \beta$ would put j in $I_\star$. Hence every surviving outside coordinate lies in $\mathcal{R} \cup \mathcal{S}$.

For $j \in \mathcal{S}$, the multiplier in Equation (35) is eventually bounded by a number strictly smaller than one. Applying the recurrence separately to the even and odd parities and using finiteness of the spectrum gives one $0 < \gamma < 1$ such that

$$\epsilon_n^{\mathrm{S}} \le C\gamma^n, \qquad \sum_{\ell=n}^{\infty} \epsilon_\ell^{\mathrm{S}} \le C\gamma^n. \tag{36}$$

Nonsurviving coordinates are zero after N_0. Combining Equation (33)–Equation (36) proves Equation (30).

Finally, Equation (29) gives

$$\sigma_n = \frac{\theta_{n+1}}{\theta_n} - 1 = \frac{\Delta_n}{2 - \Delta_n}.$$

Also $\theta_{n+1} \le \tau$, so

$$e_n = \frac{\tau^2 - \theta_n\theta_{n+1}}{\theta_n\theta_{n+1}} \ge \frac{\tau - \theta_n}{\tau}.$$

This proves Equation (31). □

The next estimate converts displacement and outside mass into summable variation of the cubic quartic defect. The factor $|L(\zeta_n)|$ is essential near a grade-three endpoint, where one active weight may vanish.

Definition 4.20 (Endpoint defect objects)**.** *For $z \in \mathbb{R}$, put*

$$B_i(z) = R_z(\xi_i) = \frac{L(z)}{z - \xi_i},$$

with $B_i(\xi_i) = D_i$ by continuity. Retain F_n, C_n, and Y_n from Lemma 4.15.

Lemma 4.21 (Summable variation of the forced quartic defect)**.** *Assume that Ω contains a four-positive vector. For all sufficiently large n,*

$$q_n(t) = L(t) + H_{n+1} + \delta_n R_{\zeta_n}(t) - C_n(t) \tag{37}$$

and

$$\delta_{n+1}X(p_{n+1}, \zeta_{n+1}) - Y_{n+1}(p_{n+1}) = \delta_n X(p_{n+1}, \zeta_n) - Y_n(p_{n+1}). \tag{38}$$

There is a constant $C > 0$ such that

$$\delta_n^2|L(\zeta_n)| \le C(\Delta_n + \epsilon_n), \tag{39}$$

$$|\delta_{n+1} - \delta_n| \le C(\Delta_n + \epsilon_n + \epsilon_{n+1}), \tag{40}$$

$$|\delta_n| \le C\big(\tau - \theta_n + \gamma^n\big), \tag{41}$$

where γ is as in Lemma 4.19.

Proof. Equations Equation (37) and Equation (38) are Equation (21) and Equation (22).

We first record uniform nonvanishing of X. Let p be either limiting quadratic factor and let $\widehat{p}$ be the other, so that $p\widehat{p} = L + \beta$. If u_1, u_2 are the roots of p, then the coefficient of t in the linear interpolation of R_z at these roots is

$$\frac{R_z(u_2) - R_z(u_1)}{u_2 - u_1} = \frac{L(z) + \beta}{p(z)} = \widehat{p}(z). \tag{42}$$

Thus $X(p, z) = \widehat{p}(z)$. Both roots of $\widehat{p}$ lie in the outer gaps by Lemma 4.18, so $\widehat{p}$ has no zero on a fixed neighborhood of $[\xi_2, \xi_3]$. Every cluster value of ζ_n lies in that interval, and the parity factors converge. Consequently, for all sufficiently large n, each of

$$|X(p_{n+1}, \zeta_n)|, \qquad |X(p_{n+1}, \zeta_{n+1})|$$

is bounded below by a fixed positive constant. The same functions are uniformly Lipschitz in their second argument there.

Rearrange Equation (38), use Equation (23) at n and $n+1$, and use Equation (20). This gives

$$|\zeta_{n+1} - \zeta_n| \le C\big(|\delta_n||L(\zeta_n)| + \epsilon_n\big) \tag{43}$$

and then

$$|\delta_{n+1} - \delta_n| \le C\delta_n^2|L(\zeta_n)| + C(\epsilon_n + \epsilon_{n+1}). \tag{44}$$

It remains to prove Equation (39). Evaluation of Equation (37) at the active nodes gives

$$b_{n,i} = \sigma_n + \frac{\delta_n}{K_n} B_i(\zeta_n) - \frac{C_n(\xi_i)}{K_n}. \tag{45}$$

Cover a fixed neighborhood of $[\xi_2, \xi_3]$ by small endpoint neighborhoods and a compact interior region. On the compact interior, $B_1(z) = L(z)/(z - \xi_1)$ is bounded away from zero directly, since L is positive in the open middle gap. Also $v_{n,1}$ is uniformly positive there. Otherwise, a subsequence with $v_{n,1} \to 0$ and $\zeta_n \to z$ in that compact interior would make the vanishing outside correction in Equation (18) give

$$0 = \beta\frac{\xi_1 - z}{D_1},$$

which is impossible. Since

$$\Delta_n \ge v_{n,1} b_{n,1}^2, \tag{46}$$

solving Equation (45) at $i = 1$, using $\sigma_n = O(\Delta_n)$ and $C_n = O(\epsilon_n)$, yields $\delta_n^2 \le C(\Delta_n + \epsilon_n)$ on the interior. Boundedness of $L(\zeta_n)$ proves Equation (39) there.

Now suppose that ζ_n is near an endpoint ξ_a, where $a \in \{2, 3\}$, and put

$$\omega_n = v_{n,a}, \qquad d_n = b_{n,a}.$$

The corrected barycentric equation at $i = a$ is

$$H_n \frac{\xi_a - \zeta_n}{D_a} = \omega_n p_n(\xi_a) + \sum_{j \notin I_\star} w_{n,j} p_n(\lambda_j) \ell_a(\lambda_j). \tag{47}$$

All polynomial values are bounded and H_n is bounded away from zero, so

$$|\zeta_n - \xi_a| \le C(\omega_n + \epsilon_n), \qquad |L(\zeta_n)| \le C(\omega_n + \epsilon_n). \tag{48}$$

Moreover, $B_a(\zeta_n)$ is bounded away from zero because $B_a(\xi_a) = D_a \neq 0$. Solving Equation (45) at $i = a$ gives

$$|\delta_n| \le C(|d_n| + \sigma_n + \epsilon_n). \tag{49}$$

Here $d_n \to 0$, and

$$\omega_n d_n^2 = |y_{n+2,\iota_a} - y_{n,\iota_a}|^2 \le \Delta_n. \tag{50}$$

Using Equation (48)–Equation (50), $\sigma_n = O(\Delta_n)$, and smallness of these quantities,

$$\begin{aligned}
\delta_n^2 |L(\zeta_n)| &\le C(d_n^2 + \sigma_n^2 + \epsilon_n^2)(\omega_n + \epsilon_n) \\
&\le C\omega_n d_n^2 + C\sigma_n^2 + C\epsilon_n \\
&\le C(\Delta_n + \epsilon_n).
\end{aligned}$$

This proves Equation (39) near both endpoints and hence everywhere for large n.

Substitution into Equation (44) proves Equation (40). We know $\delta_n \to 0$ from Equation (15), while $\sum_n (\Delta_n + \epsilon_n) < \infty$ by Lemma 4.19. Therefore the variation estimate may be summed backward from infinity:

$$|\delta_n| \le C \sum_{\ell=n}^{\infty} (\Delta_\ell + \epsilon_\ell).$$

Lemma 4.19 now proves Equation (41). □

The defect-tail bound controls the exceptional active multiplier during visits near either endpoint of the middle gap. All other active multipliers are square-summable on those visits.

Lemma 4.22 (Endpoint multiplier absorption)**.** *Assume that Ω contains a four-positive vector, and fix γ from Lemma 4.19. There are $0 < \varepsilon_0 < (\xi_3 - \xi_2)/3$, N, and $C > 0$ such that, for $a \in \{2, 3\}$, the endpoint-visit set*

$$\mathcal{N}_a = \{n \ge N : |\zeta_n - \xi_a| < \varepsilon_0\}$$

has the following properties. Put $\omega_n = v_{n,a}$ and $d_n = b_{n,a}$ for $n \in \mathcal{N}_a$. Then

$$|d_n| \le C\big(\tau - \theta_n + \gamma^n\big). \tag{51}$$

For every $\eta > 0$, there are $C_\eta > 0$ and $N_\eta \ge N$ such that

$$d_n^2 \le \eta e_n + C_\eta \gamma^{2n} \qquad (n \in \mathcal{N}_a,\ n \ge N_\eta). \tag{52}$$

For every $i \neq a$,

$$\sum_{n \in \mathcal{N}_a} b_{n,i}^2 < \infty. \tag{53}$$

Proof. Equation Equation (45) at $i = a$, together with uniform boundedness away from zero of B_a, gives

$$|d_n| \le C(\sigma_n + |\delta_n| + \epsilon_n). \tag{54}$$

By Lemma 4.19, $\Delta_n + \epsilon_n \le C(\tau - \theta_n + \gamma^n)$, and $\sigma_n = O(\Delta_n)$. Equation Equation (41) proves Equation (51).

Put $g_n = \tau - \theta_n$. By Equation (31), $e_n \ge g_n/\tau$, while Equation (51) gives

$$d_n^2 \le C(g_n^2 + \gamma^{2n}).$$

Since $g_n \to 0$, for every $\eta > 0$ one has $Cg_n^2 \leq \eta e_n$ for all sufficiently large n. This proves Equation (52).

For Equation (53), let

$$\widehat{C}_{n,i} = \frac{C_n(\xi_i)}{K_n}.$$

Solving Equation (45) at $i = a$ and substituting into the same formula at $i \neq a$ gives

$$b_{n,i} = \sigma_n + \frac{B_i(\zeta_n)}{B_a(\zeta_n)}(d_n - \sigma_n + \widehat{C}_{n,a}) - \widehat{C}_{n,i}. \tag{55}$$

By Equation (48),

$$\left|\frac{B_i(\zeta_n)}{B_a(\zeta_n)}\right| \leq C|\zeta_n - \xi_a| \leq C(\omega_n + \epsilon_n). \tag{56}$$

Since $\widehat{C}_{n,i} = O(\epsilon_n)$ and $d_n \to 0$,

$$|b_{n,i}| \leq C\big(\sigma_n + \epsilon_n + \omega_n|d_n|\big). \tag{57}$$

Now $\sigma_n \to 0$, and eventually

$$\sigma_n \leq 2\log(1+\sigma_n) = 2\log\frac{\theta_{n+1}}{\theta_n}.$$

Thus $\sum_n \sigma_n < \infty$ by telescoping. Also $\sum_n \epsilon_n < \infty$, and

$$\sum_{n\in\mathcal{N}_a} \omega_n d_n^2 \leq \sum_{n\in\mathcal{N}_a} \Delta_n < \infty.$$

Squaring Equation (57) and using $\omega_n^2 d_n^2 \leq \omega_n d_n^2$ proves Equation (53). $\square$

4.7 Exclusion of resonance and four-point collapse

We now use an exact logarithmic cocycle to exclude every surviving multiplier--1 coordinate. Cubic interpolation cancels all terms that are linear in the four active multiplier errors.

Lemma 4.23 (Exclusion of multiplier-minus-one resonances)**.** *If Ω contains a four-positive vector, then $\mathcal{R} = \varnothing$.*

Proof. Suppose to the contrary that $\lambda = \lambda_j$ for some $j \in \mathcal{R}$, and put

$$\ell_i = \ell_i(\lambda) = \frac{L(\lambda)}{D_i(\lambda - \xi_i)}.$$

Lagrange interpolation gives $\sum_i \ell_i = 1$. Every active coordinate is nonzero at every finite time: if one vanished, it would remain zero, contradicting the existence of a four-positive omega-limit point. For all sufficiently large n, $q_n(\xi_i) > 0$ and $q_n(\lambda) < 0$, so

$$\Psi_{n,\lambda} = \log w_{n,j} + \sum_{i=1}^{4} \ell_i \log v_{n,i} \tag{58}$$

is well defined. The exact two-step squared-weight recurrence gives

$$\frac{\Psi_{n+2,\lambda} - \Psi_{n,\lambda}}{2} = \log\frac{|q_n(\lambda)|}{K_n} + \sum_{i=1}^{4} \ell_i \log\frac{q_n(\xi_i)}{K_n}. \tag{59}$$

The polynomial $q_n - q_\star$ has degree at most three. Interpolating it at the four active nodes, and using $q_\star(\lambda) = -\beta$ and $q_\star(\xi_i) = \beta$, yields

$$\frac{|q_n(\lambda)|}{K_n} = 1 + 2e_n - \sum_{i=1}^{4} \ell_i b_{n,i}. \tag{60}$$

Thus, with $\Phi_{n,\lambda} = (\Psi_{n+2,\lambda} - \Psi_{n,\lambda})/2$,

$$\Phi_{n,\lambda} = \log\left(1 + 2e_n - \sum_i \ell_i b_{n,i}\right) + \sum_i \ell_i \log(1 + b_{n,i}). \tag{61}$$

Taylor expansion at the origin cancels every term linear in $b_{n,i}$:

$$\Phi_{n,\lambda} = 2e_n + O\left(e_n^2 + \sum_{i=1}^{4} b_{n,i}^2\right). \tag{62}$$

All logarithm arguments tend to one. Because $e_n \geq 0$ and $e_n^2 = o(e_n)$, there is a constant C_λ such that

$$\Phi_{n,\lambda} \geq e_n - C_\lambda \sum_{i=1}^{4} b_{n,i}^2 \tag{63}$$

for all sufficiently large n.

Choose the endpoint neighborhoods from Lemma 4.22. Their complement in a fixed neighborhood of $[\xi_2, \xi_3]$ is compactly contained in the middle gap. During visits to this compact interior, all four active weights are uniformly positive; otherwise the limiting corrected barycentric equations would place the cluster value at an endpoint. Hence

$$\Delta_n \geq \sum_{i=1}^{4} v_{n,i} b_{n,i}^2 \tag{64}$$

and $\sum_n \Delta_n < \infty$ make $\sum_i b_{n,i}^2$ summable over the interior visits.

At a visit in $\mathcal{N}_a$, the errors $b_{n,i}^2$ for $i \neq a$ are summable by Equation (53). Apply Equation (52) with $\eta < (2C_\lambda)^{-1}$. Then Equation (63) gives

$$\Phi_{n,\lambda} \geq \frac{1}{2} e_n - C_\lambda \sum_{i \neq a} b_{n,i}^2 - C\gamma^{2n} \qquad (n \in \mathcal{N}_a) \tag{65}$$

for all sufficiently large n. Every large index is either an interior visit or lies in one of the two endpoint neighborhoods, because every cluster value of ζ_n lies in $[\xi_2, \xi_3]$. It follows that the negative parts of $\Phi_{n,\lambda}$ are summable on each parity. Since $\Psi_{n+2,\lambda} - \Psi_{n,\lambda} = 2\Phi_{n,\lambda}$, this summability bounds $\Psi_{n,\lambda}$ below on each parity.

Choose an even subsequence converging to the four-positive omega-limit point. Along it, all four active logarithms in Equation (58) remain bounded, whereas $w_{n,j} \to 0$ by Lemma 4.7. Hence $\Psi_{n,\lambda} \to -\infty$ along that subsequence, contradicting the parity-wise lower bound. Therefore $\mathcal{R} = \varnothing$. □

With resonances absent, every surviving transverse coordinate has a strict gap. The resulting geometric forcing and the parabolic dichotomy make the corrected barycentric coordinate converge.

Lemma 4.24 (Strict-gap four-point collapse)**.** *Suppose*

$$|q_\star(\lambda_j)| < \beta$$

for every $j \notin I_\star$ whose coordinate survives. Then Ω is a singleton and y_{2k} converges.

Proof. If every vector in Ω has grade three, the claim follows from Lemma 4.8. Otherwise, Ω contains a four-positive vector. No active coordinate can then have been annihilated at a finite time.

The strict inequalities, full-sequence convergence $q_n \to q_\star$, and $H_n \to \beta$ give constants $C_0 > 0$ and $0 < \gamma < 1$ such that

$$\epsilon_n \le C_0 \gamma^n \tag{66}$$

for all sufficiently large n. This follows by applying Equation (35) separately on the two parities; coordinates annihilated at a finite time remain zero.

By Lemma 4.18, the limiting active sign pattern is $(+,-,-,+)$, and every cluster value of ζ_n belongs to $[\xi_2, \xi_3]$. The corrected barycentric equations also give

$$\operatorname{dist}(\zeta_n, [\xi_2, \xi_3]) \le C\epsilon_n. \tag{67}$$

Indeed, let s_i be the eventual sign of $p_n(\xi_i)$. From Equation (18) and boundedness of the outside correction,

$$s_i H_n \frac{\xi_i - \zeta_n}{D_i} \ge -C\epsilon_n \qquad (1 \le i \le 4).$$

The four exact half-line inequalities intersect in $[\xi_2, \xi_3]$, and their $O(\epsilon_n)$ relaxations imply Equation (67). By Equation (42), $X(p_{n+1}, z)$ is bounded away from zero for large n and z near this interval.

Equation Equation (20) and Equation (23) give

$$|\zeta_{n+1} - \zeta_n| \le C(|\delta_n| + \epsilon_n).$$

Use this in Equation (22), subtract the two X terms, and use their uniform Lipschitz dependence on the second argument. Since $\delta_n \to 0$,

$$|\delta_{n+1} - \delta_n| \le C\delta_n^2 + C\gamma^n. \tag{68}$$

Lemma 4.16 applies. If $\sum_n |\delta_n| < \infty$, then Equation (20) and Equation (66) give finite total variation of ζ_n.

Otherwise δ_n is eventually one-signed. The polynomial L is nonnegative on $[\xi_2, \xi_3]$, while Equation (67) shows that the negative part of $L(\zeta_n)$ is $O(\epsilon_n)$. After orienting the real line according to the fixed sign of δ_n, Equation (20) shows that all increments in the adverse direction have a summable total, because $F_n = O(\epsilon_n)$ and $\sum_n \epsilon_n < \infty$. A bounded real sequence with summable increments in one direction converges: adding the tail of those adverse increments makes it monotone up to a vanishing correction. Thus ζ_n converges in both alternatives.

For even $n = 2k$, Equation (18) gives

$$v_{2k,i} p_{2k}(\xi_i) = H_{2k} \frac{\xi_i - \zeta_{2k}}{D_i} + O(\epsilon_{2k}).$$

Because $p_{2k}(\xi_i) \to p_{\mathrm{e}}(\xi_i) \neq 0$, all active squared coordinates converge, while the outside squared coordinates tend to zero. Finally,

$$\frac{y_{2k+2,\iota_i}}{y_{2k,\iota_i}} = \frac{q_{2k}(\xi_i)}{K_{2k}} > 0$$

for every sufficiently large k. Hence the sign of each nonvanishing active coordinate eventually stabilizes, and a coordinate whose squared weight tends to zero itself tends to zero. Therefore y_{2k} converges, and Ω is a singleton. □

The preceding two lemmas classify every possible outside coordinate and complete the four-node case without a genericity or strict-gap assumption.

Lemma 4.25 (Collapse of the four-point omega-limit set)**.** *Assume $|I_\star| = 4$. Then Ω is a singleton. Consequently, y_{2k} converges and, by continuity of T, so does y_{2k+1}.*

Proof. If every vector in Ω has grade three, Lemma 4.8 gives the conclusion. Otherwise, Ω contains a grade-four vector, necessarily four-positive on $I_\star$.

Let $j \notin I_\star$ be a surviving coordinate. By Equation (35) and Lemma 4.7, the inequality $|q_\star(\lambda_j)| > \beta$ is impossible: its squared two-step multiplier would eventually exceed one, whereas the coordinate mass tends to zero. The equality $q_\star(\lambda_j) = \beta$ is impossible by the definition of $I_\star$, and Lemma 4.23 excludes $q_\star(\lambda_j) = -\beta$. Thus every surviving outside coordinate satisfies $|q_\star(\lambda_j)| < \beta$; every nonsurviving coordinate is eventually zero. Lemma 4.24 applies, so Ω is a singleton. Its point has grade three or four by Lemma 4.5, and continuity of T there gives convergence of the odd subsequence. □

4.8 Conclusion

Proof of Theorem 1.1. Lemma 4.4 proves that every residual is nonzero and retains grade at least three. By Lemmas 4.1 and 4.2, the normalized residuals are isometric to the diagonal monic-quadratic orbit $y_{k+1} = T(y_k)$.

Lemmas 4.3, 4.5 and 4.6 produce a nonempty connected even omega-limit set Ω and an equality set $I_\star$ with $3 \leq |I_\star| \leq 4$. If every omega-limit vector has grade three, Lemma 4.8 makes Ω a singleton. Otherwise, some omega-limit vector has grade four, so $|I_\star| = 4$, and Lemma 4.25 again makes Ω a singleton. Therefore $y_{2k} \to y_{\mathrm{e}}$. By Lemma 4.5, y_{e} has grade three or four, T is continuous there, and

$$y_{2k+1} = T(y_{2k}) \longrightarrow y_{\mathrm{o}} := T(y_{\mathrm{e}}), \qquad T(y_{\mathrm{o}}) = y_{\mathrm{e}}.$$

The proof of Lemma 4.5 gives grade at least three for y_{o}. Moreover, with $q(t) = \pi_{y_{\mathrm{o}}}(t)\pi_{y_{\mathrm{e}}}(t)$, polynomial commutativity and $q(A)y_{\mathrm{e}} = \beta y_{\mathrm{e}}$ imply $q(A)y_{\mathrm{o}} = \beta y_{\mathrm{o}}$; hence its grade is at most four. The fixed spectral isometry in Lemma 4.2 transfers both limits and their grades to the original residual directions. The energy-ratio estimate is Equation (5), and the per-restart resource bound follows from Lemma 4.1. This proves the theorem. □

References

[Aka59] Hirotugu Akaike. On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Annals of the Institute of Statistical Mathematics*, 11:1–16, 1959. 1, 2

[CG89] A. T. Chronopoulos and C. W. Gear. *s*-step iterative methods for symmetric linear systems. *Journal of Computational and Applied Mathematics*, 25(2):153–168, 1989. 2

[FLT23] Vance Faber, Jörg Liesen, and Petr Tichý. On the Forsythe conjecture. *BIT Numerical Mathematics*, 63, 2023. Article 49. 1, 2

[FM51] George E. Forsythe and Theodore S. Motzkin. Asymptotic properties of the optimum gradient method. *Bulletin of the American Mathematical Society*, 57(3):183, 1951. Abstract 57-3-231. 1

[For68] George E. Forsythe. On the asymptotic directions of the s-dimensional optimum gradient method. *Numerische Mathematik*, 11:57–76, 1968. 1, 2

[HS52] Magnus R. Hestenes and Eduard Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952. 1

[PWZ09] Luc Pronzato, Henry P. Wynn, and Anatoly A. Zhigljavsky. A dynamical-system analysis of the optimum s-gradient algorithm. In Luc Pronzato and Anatoly Zhigljavsky, editors, *Optimal Design and Related Areas in Optimization and Statistics*, volume 28 of *Springer Optimization and Its Applications*, pages 39–80. Springer, New York, NY, 2009. 2

[SS86] Youcef Saad and Martin H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986. 2

[VL10] Eugene Vecharynski and Julien Langou. The cycle-convergence of restarted GMRES for normal matrices is sublinear. *SIAM Journal on Scientific Computing*, 32(1):186–196, 2010. 2

[Zab79] A. F. Zabolotskaya. Asymptotic behaviour of the s-dimensional optimal gradient method in hilbert space. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 19(1):233–238, 1979. 2

[ZB84] P. Ph. Zhuk and L. N. Bondarenko. On a conjecture of G. Forsythe. *Mathematics of the USSR-Sbornik*, 49(2):427–445, 1984. English translation; Russian original published in 1983. 2