Barzilai–Borwein Fails Superlinear Convergence on an Open Set of Quadratics for Every Dimension $n \geq 4$

Dawei Li Xiaotian Jiang Mingyi Hong
University of Minnesota
{li004678,jian0851,mhong}@umn.edu

July 14, 2026

Abstract

Barzilai–Borwein (BB) method has shown strong practical performance in continuous optimization, yet its convergence dynamics remains poorly understood. In particular, a central unresolved question is whether BB converges superlinearly for almost every strictly convex quadratic problem and initialization. We provide a negative answer to this question. Specifically, for every finite dimension $n \geq 4$, we construct a nonempty open, hence positive-Lebesgue-measure, family of strictly convex quadratic problems and initial points for which the long Barzilai–Borwein method (BB1) converges but cannot converge root-superlinearly. More precisely, with the explicit constants

$$\underline{\rho} = 10^{-6}, \qquad \overline{\rho} = 0.61,$$

every spectral component of the gradient is bounded above and below by the corresponding geometric sequence. Consequently, the gradient norm and the energy norm of the error satisfy two-sided geometric estimates with the same rates, while the objective gap satisfies the corresponding estimates with squared rates. In particular, all three quantities are bounded below by geometric sequences, ruling out superlinear convergence. The construction is highly nontrivial, based on a computer-assisted proof of a nonresonant, attracting seven-cycle of the projectivized BB dynamics in dimension four.

1 Introduction

Gradient descent is one of the oldest and most widely used methods in continuous optimization. Each iteration requires only a gradient and a few vector operations, making it attractive in large-scale problems. Its performance, however, can deteriorate on ill-conditioned settings: exact line search may produce zigzag iterates, whereas a good constant step requires spectral information that is rarely available in advance.

The Barzilai–Borwein (BB) method (Barzilai and Borwein, 1988) addresses this difficulty without changing the negative-gradient direction. Instead, it extracts a scalar curvature estimate from the two most recent iterates. This change of step-size rule, rather than search direction, can substantially improve the practical behavior of a basic gradient iteration (Fletcher, 2005; Birgin et al., 2014).

For a differentiable objective $f : \mathbb{R}^n \to \mathbb{R}$, write $g_k = \nabla f(x_k)$, $s_{k-1} = x_k - x_{k-1}$, and $y_{k-1} = g_k - g_{k-1}$. The two classical BB step sizes are (Barzilai and Borwein, 1988)

$$\alpha_k^{\text{BB1}} = \frac{s_{k-1}^\top s_{k-1}}{s_{k-1}^\top y_{k-1}}, \qquad \alpha_k^{\text{BB2}} = \frac{s_{k-1}^\top y_{k-1}}{y_{k-1}^\top y_{k-1}}, \qquad x_{k+1} = x_k - \alpha_k g_k. \tag{1}$$

The first choice makes the scalar matrix $(\alpha_k^{\mathrm{BB1}})^{-1}I$ the least-squares solution of the secant equation; the second makes $\alpha_k^{\mathrm{BB2}}I$ a least-squares approximation of the inverse Hessian. They are therefore often called the long and short BB steps, respectively. The formulation of the two BB steps are different, but they possess similar properties in training. BB runs at essentially the cost of gradient descent while carrying curvature information without forming a matrix and, in its basic form, without performing a line search. On a quadratic, the reciprocal BB step is a Rayleigh quotient of the Hessian; this is also the origin of the name *spectral gradient method*.

Despite this simple construction, BB generates nonlinear, history-dependent dynamics even on quadratic objectives. A longstanding open question is whether BB converges superlinearly for almost every strictly convex quadratic problem and initialization, or whether failure of root-superlinear convergence can persist on a set of positive measure rather than only on exceptional instances. More precisely, the question is whether

$$\lim_{k\to\infty} \|g_k\|^{1/k} = 0$$

for almost every strictly convex quadratic problem and initialization. We aim to resolve this question in this paper.

Significance and applications. The BB step has become an algorithmic primitive rather than a single method. Raydan combined it with a non-monotone globalization strategy for large-scale unconstrained optimization (Raydan, 1997); related non-monotone frameworks were developed in Grippo and Sciandrone (2002) and Zhang and Hager (2004), and the spectral projected gradient method brought the same step selection to optimization over convex sets (Birgin et al., 2000, 2001; Dai and Fletcher, 2005b; Andreani et al., 2005). BB-type spectral steps now appear in support-vector-machine training (Serafini et al., 2005), sparse reconstruction and compressed sensing (Figueiredo et al., 2007; Wright et al., 2009; Van Den Berg and Friedlander, 2009; Wen et al., 2010), image restoration (Wang and Ma, 2007; Bonettini et al., 2009), nonnegative matrix factorization (Huang et al., 2015), optimization with orthogonality constraints (Wen and Yin, 2013), Riemannian optimization (Iannazzo and Porcelli, 2018), stochastic and variance-reduced methods (Tan et al., 2016), distributed optimization with locally computed step sizes (Gao et al., 2022), and PDE-constrained optimization in Hilbert space (Azmi and Kunisch, 2020, 2022); more recent safeguarded and adaptive constructions seek global guarantees on broader function classes (Burdakov et al., 2019; Zhou et al., 2026). These descendants alter the pure iteration in essential ways, such as projection, regularization, line searches, stochastic averaging, safeguarding, but all of them inherit its step-size mechanism. Understanding the unmodified deterministic iteration is therefore valuable both in its own right and as a foundation for interpreting its many descendants.

Earlier convergence and rate theory. The known convergence theory of BB mostly focuses on low-dimensional problems. Barzilai and Borwein (1988) proved R-superlinear convergence for two-dimensional strictly convex quadratics. Raydan (1993) established global convergence for arbitrary finite-dimensional strictly convex quadratics, and Dai and Liao (2002) proved an R-linear rate in arbitrary dimension, together with a local consequence for sufficiently smooth nonquadratic objectives near a nondegenerate minimizer. In dimension two, Dai (2013) related the rate to the initial spectral balance and showed that superlinear behavior holds for almost every initialization, while linear behavior occurs on an exceptional set. Dai and Fletcher (2005a) developed an asymptotic framework for gradient methods with retarded spectral information and, by computing the asymptotics of BB and its relatives, observed a transition from superlinear to linear behavior at some dimension $n \geq 4$ depending on the method. Their statements are asymptotic computations and numerical observations, and no rigorous lower-rate mechanism in dimension $n \geq 4$ was identified.

For general dimension, rate bounds have been substantially less informative about the observed fast behavior. A recent refinement by Li and Sun (2021) proves, for BB1 on an arbitrary finite-dimensional strongly convex quadratic with condition number $\kappa = \lambda_n/\lambda_1$, an R-linear upper factor $1 - 1/\kappa$. This is comparable in condition-number scaling to the exact-line search steepest-descent factor. The same work gives a lower-rate example supported only on the two extreme eigenspaces, with exact factor $(\kappa - 1)/(\kappa + 1)$, ruling out superlinear convergence. However, for $n \geq 2$ its initial point lies in a Lebesgue-null set, so an arbitrarily small perturbation could possibly accelerate the convergence; the same paper explicitly raises the convergence rate under generic initialization as an open question.

Parallel lines of work have sought either to globalize BB or to design spectral rules with more transparent asymptotics. These include gradient methods with retards (Friedlander et al., 1998), alternate and cyclic BB schemes (Dai, 2003; Dai et al., 2006), convex combinations of the long and short BB steps (Dai et al., 2018), and steps engineered for two-dimensional termination or favorable spectral limits (Yuan, 2006; Huang et al., 2021, 2022). This literature gives considerable evidence that low-dimensional spectral mechanisms govern the performance of many gradient methods. What has been missing for the original BB1 method is an answer to the following question:

> In dimension $n \geq 4$, is every linear-rate BB1 example an unstable exception, or does genuinely full-dimensional linear-rate behavior persist on a robust family?

Contribution of the present lower bound. This paper answers this question for BB1. For every finite dimension $n \geq 4$ and every linear term $b \in \mathbb{R}^n$, we construct a nonempty open set $\mathcal{A}_n \subset \mathbb{S}^n_{++}$ of simple-spectrum matrices such that, for every $A \in \mathcal{A}_n$, there is a nonempty open set $\mathcal{X}_n(A)$ of initial points for which the admissibly initialized BB1 iteration is well defined and converges to $x_* = A^{-1}b$. More precisely, if $\Pi_i(A)$ denotes the orthogonal spectral projector associated with the ith eigenvalue, then

$$10^{-6k}\,\|\Pi_i(A)g_0\| \;\leq\; \|\Pi_i(A)g_k\| \;\leq\; 0.61^k\,\|\Pi_i(A)g_0\|, \qquad i = 1, \ldots, n, \quad k \geq 0. \tag{2}$$

Here "admissibly initialized" means that the first step is the inverse gradient Rayleigh quotient,

$$\alpha_0 = \frac{g_0^\top g_0}{g_0^\top A g_0}, \qquad x_1 = x_0 - \alpha_0 g_0,$$

so the orbit is an actual BB1 trajectory rather than an arbitrarily prescribed pair of projective states. The joint set $\Omega_n = \{(A, x_0) : A \in \mathcal{A}_n,\ x_0 \in \mathcal{X}_n(A)\}$ is open and nonempty in $\mathbb{S}^n_{++} \times \mathbb{R}^n$. It therefore has positive Lebesgue measure, as does every initial-point fiber $\mathcal{X}_n(A)$. Summing (2) over the orthogonal spectral components gives the same two-sided geometric estimate for the gradient norm and the energy norm of the error; the objective gap satisfies the corresponding squared estimate. In particular,

$$\liminf_{k\to\infty} \|g_k\|^{1/k} \geq 10^{-6} > 0,$$

so none of these trajectories is root-superlinearly or Q-superlinearly convergent. An almost-everywhere superlinearity statement on the joint problem-data space is therefore false in every dimension $n \geq 4$; in particular, the question of Li and Sun (2021) about generic initialization is answered in the negative on an open set of matrices.

The result strengthens the earlier lower-rate picture in three ways.

(i) *Robustness.* The slow behavior is not confined to a specially balanced endpoint subspace or to a zero-measure initialization. It persists under simultaneous perturbations of the spectrum and the initial point.

(ii) *A dynamical mechanism.* The proof identifies a nonresonant, attracting period-seven orbit of the projectivized four-dimensional BB1 dynamics. A certified fifteen-step connection reaches its basin from the admissible-initialization manifold, and a spectral translation makes every radial multiplier contractive. Attraction of the projective cycle, rather than invariance of a coordinate subspace, creates the open basin and the lower geometric rate.

(iii) *Full-dimensional persistence.* Four spectral modes remain active along the asymptotic cycle. A uniform transverse Floquet estimate then embeds the construction into each finite dimension $n > 4$ while keeping every additional component nonzero and geometrically controlled. The normalized spectral measure consequently approaches a seven-periodic measure rather than collapsing to the two endpoint modes familiar from exact steepest descent (Akaike, 1959; Forsythe, 1968; Pronzato et al., 2006).

The periodic orbit is certified by interval arithmetic and a contraction argument; it is not inferred from a long floating-point simulation. This computer-assisted component establishes exact existence, nonresonance, and attraction inside explicit rational boxes; openness, positive measure, and persistence in higher dimension then follow from analytic stability arguments.

The theorem is an existence result, not a universal rate formula. It concerns BB1, not every BB variant. The set $\mathcal{A}_n$ is a small open neighborhood of one explicitly constructed spectrum—three separated eigenvalues together with, for $n > 4$, the remaining eigenvalues clustered in a short interval near the top—so the phrase "every finite dimension" asserts that such a neighborhood exists for each n, not that its size is uniform in n. The constants 10^{-6} and 0.61 are conservative separation margins rather than claimed optimal asymptotic factors. The contribution is structural: it proves that the dimension-four transition from low-dimensional superlinear behavior to linear-rate behavior can be generated by a stable recurrent attractor and can occupy positive measure. In this sense, the lower bound complements the classical and recent R-linear upper bounds. The upper theory controls how slow BB1 can be; the present result shows that a positive linear root factor is actually realized robustly, and explains the orbit geometry that sustains it.

AI-Assisted Development. The sketch of the proof of the main result of this paper was obtained by GPT 5.6 Sol. GPT was given an equivalent problem formulation (component-wise quotient formulation) and asked to construct the lower bound example. The prompt is provided in Appendix B. The construction by GPT is highly non-trivial. Human verification and polishing were done afterwards.

2 The BB1 method and the main theorem

Let $\mathbb{S}^n_{++}$ denote the open cone of real symmetric positive-definite $n \times n$ matrices. It is an open subset of the Euclidean space $\mathbb{S}^n$, which has dimension $n(n+1)/2$. Fix $A \in \mathbb{S}^n_{++}$ and $b \in \mathbb{R}^n$, and consider

$$f_{A,b}(x) = \frac{1}{2}x^T Ax - b^T x, \qquad x_* := A^{-1}b, \qquad g(x) := \nabla f_{A,b}(x) = Ax - b. \tag{3}$$

The unique minimizer is x_* because $A \succ 0$.

We study the long Barzilai–Borwein method, usually called BB1 Barzilai and Borwein (1988); classical convergence analyses on strictly convex quadratics include Raydan (1993); Dai and Liao (2002). The initialization is part of the theorem and is therefore stated explicitly. Given $x_0 \neq x_*$, put $g_0 = Ax_0 - b$ and

$$\alpha_0 := \frac{g_0^T g_0}{g_0^T A g_0}, \qquad x_1 := x_0 - \alpha_0 g_0. \tag{4}$$

For $k \geq 1$, provided the quantities are defined, set

$$\begin{aligned} s_{k-1} &:= x_k - x_{k-1}, \quad y_{k-1} := g_k - g_{k-1}, \\ \alpha_k &:= \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}, \quad x_{k+1} := x_k - \alpha_k g_k, \end{aligned} \qquad g_k := Ax_k - b. \tag{5}$$

The choice (4) is the inverse Rayleigh quotient of the initial gradient. It makes the first two reciprocal step sizes equal.

Main Result. For a simple-spectrum matrix A, let

$$0 < \lambda_1(A) < \cdots < \lambda_n(A)$$

be its ordered eigenvalues and let $\Pi_i(A)$ be the orthogonal projector onto the one-dimensional eigenspace corresponding to $\lambda_i(A)$. The orientation of an eigenvector is irrelevant because all conclusions below use $\|\Pi_i(A)g_k\|$.

Theorem 2.1 (Positive-measure lower rate for BB1)**.** *Fix an integer $n \geq 4$ and a vector $b \in \mathbb{R}^n$. There exist a nonempty open set $\mathcal{A}_n \subset \mathbb{S}^n_{++}$ consisting of simple-spectrum matrices and, for every $A \in \mathcal{A}_n$, a nonempty open set $\mathcal{X}_n(A) \subset \mathbb{R}^n \setminus \{x_*\}$ with the following properties.*

For every $x_0 \in \mathcal{X}_n(A)$, the BB1 iteration (4)–(5) is well-defined for every $k \geq 0$, converges to x_*, and satisfies, for every $i \in \{1, \ldots, n\}$ and every $k \geq 0$,

$$\underline{\rho}^k \|\Pi_i(A)g_0\| \leq \|\Pi_i(A)g_k\| \leq \overline{\rho}^k \|\Pi_i(A)g_0\|, \qquad \underline{\rho} = 10^{-6}, \quad \overline{\rho} = 0.61. \tag{6}$$

In particular, none of the initial spectral components is zero.

The union

$$\Omega_n := \{(A, x) : A \in \mathcal{A}_n,\ x \in \mathcal{X}_n(A)\} \subset \mathbb{S}^n_{++} \times \mathbb{R}^n \tag{7}$$

is nonempty and open. Consequently Ω_n has positive Lebesgue measure in $\mathbb{S}^n \times \mathbb{R}^n$, and each fiber $\mathcal{X}_n(A)$ has positive Lebesgue measure in $\mathbb{R}^n$.

Moreover, writing $e_k = x_k - x_*$ and $\|e\|_A = (e^T A e)^{1/2}$, one has

$$\underline{\rho}^k \|g_0\| \leq \|g_k\| \leq \overline{\rho}^k \|g_0\|, \tag{8}$$

$$\underline{\rho}^k \|e_0\|_A \leq \|e_k\|_A \leq \overline{\rho}^k \|e_0\|_A, \tag{9}$$

$$\underline{\rho}^{2k}(f(x_0) - f(x_*)) \leq f(x_k) - f(x_*) \leq \overline{\rho}^{2k}(f(x_0) - f(x_*)). \tag{10}$$

Thus

$$\liminf_{k\to\infty} \|g_k\|^{1/k} \geq 10^{-6}, \qquad \liminf_{k\to\infty} \|e_k\|_A^{1/k} \geq 10^{-6}, \tag{11}$$

and these BB1 orbits are not root-superlinearly or Q-superlinearly convergent.

The proof occupies Sections 4–9. The constants are conservative and are not claimed to be optimal. The result is an existence theorem for an open family; it does not assert the same lower bound for every positive spectrum.

Why the BB dynamics are difficult. Even for the quadratic objective (3), BB1 generates a nonlinear delayed dynamical system. Let $0 < \lambda_1 \le \cdots \le \lambda_n$ be the eigenvalues of A, let $v_1, \ldots, v_n$ be an orthonormal eigenbasis, and write $g_k = \sum_i d_k^i v_i$. Then for BB1 on this quadratic,

$$\alpha_k = \frac{\sum_{j=1}^n (d_{k-1}^j)^2}{\sum_{j=1}^n \lambda_j (d_{k-1}^j)^2}, \qquad d_{k+1}^i = (1 - \lambda_i \alpha_k) d_k^i, \tag{12}$$

equivalently,

$$d_{k+1}^i = d_k^i \, \frac{\sum_{j=1}^n (\lambda_j - \lambda_i)(d_{k-1}^j)^2}{\sum_{j=1}^n \lambda_j (d_{k-1}^j)^2}. \tag{13}$$

Equation (13) displays the first source of difficulty. Diagonalization completely decouples fixed-step gradient descent, but it does not decouple BB1. The update of the ith component d^i from time k to time $k+1$ depends on every component at time $k-1$. Thus, although the objective is diagonal in the eigenbasis, its spectral modes remain coupled through a delayed Rayleigh quotient. This coupled and delayed feedback produces nonlinear behavior. Depending on the Rayleigh quotient selected by the preceding gradient, the multiplier of a spectral component can be negative, close to zero, or greater than one in absolute value. Consequently, components can change sign, high-eigenvalue components can grow, and different eigendirections may take turns dominating the gradient, so neither the objective value nor the gradient norm is monotone. (On nonquadratic objectives, the situation is more delicate – $s_{k-1}^\top y_{k-1}$ may fail to remain positive, and the unsafeguarded method can take excessively long steps or fail to converge even under strong convexity (Fletcher, 2005; Burdakov et al., 2019) – but they are not the subject of this paper.)

The second difficulty is determining the actual asymptotic rate. As a component multiplier can be close to zero, a spectral component may become arbitrarily small in a single iteration. An upper estimate $\|g_k\| \le Cq^k$ therefore cannot distinguish genuinely linear convergence from faster convergence, such as superlinear. Such an upper bound also does not determine which spectral modes survive, whether the normalized state approaches a fixed point or a cycle, or what lower asymptotic rate is actually attained. Ruling out superlinear convergence needs a lower bound, or an equally precise description of the asymptotic orbit.

There is a third difficulty in answering an almost-everywhere question. A single specially constructed lower bound example could be an isolated exception and hence belong to a set of Lebesgue measure zero (Dai, 2013; Li and Sun, 2021). In practice, such an example may depart from the theoretical trajectory and lead to faster convergence rate due to numerical instability He et al. (2025). A robust lower bound example should additionally prove that the convergence rate persists under perturbations of both the spectrum and the initialization. This is the aspect of BB dynamics for which the classical theory has been least complete.

3 Outline of the proof

The lower bound is not obtained by estimating a generic BB1 trajectory. The proof constructs one rigorously certified orbit of a low-dimensional dynamical system, verifies that this orbit stays uniformly away from every resonance, and then lets stability and continuity enlarge the single orbit into an open family. Figure 2 displays the chain of implications; this section walks through it once.

Reduction to a factor bound (§4). On a quadratic, BB1 acts on each spectral component of the gradient through the exact multiplier identity $d_{k+1}^i = \big((a_k - \lambda_i)/a_k\big)\, d_k^i$, where $a_k = \alpha_k^{-1}$ is

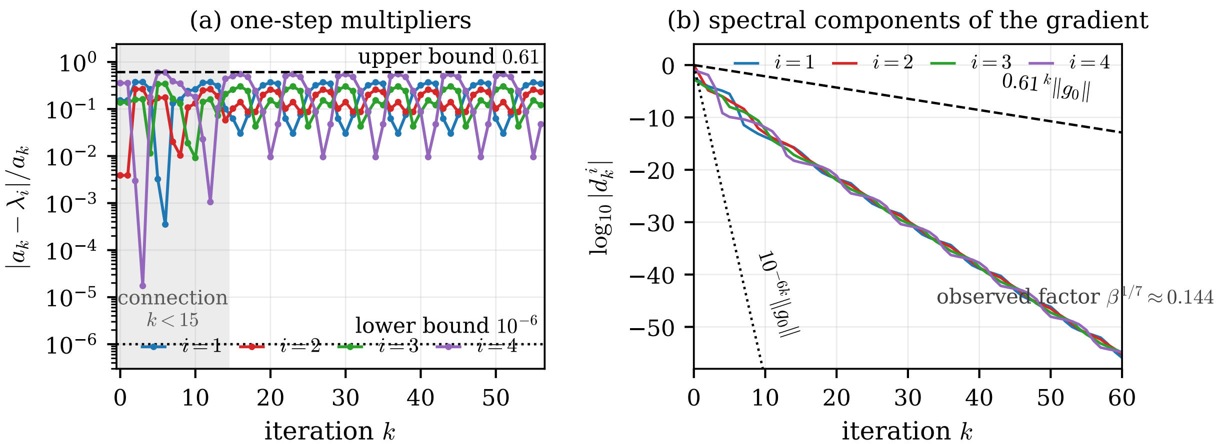

Figure 1: Numerical realization of the factor bounds along the certified orbit at the central shifted spectrum λ^*, computed in 200-digit arithmetic. *(a)* The one-step multipliers $|a_k - \lambda_i|/a_k$: after the fifteen-step connection (shaded) they are seven-periodic and remain inside the certified band $[10^{-6}, 0.61]$; the transient minimum, approximately 1.7×10^{-5}, occurs at $k = 3$. *(b)* The four spectral components of the gradient decay in parallel between the two geometric envelopes; the observed asymptotic per-step factor is $\beta^{1/7} \approx 0.144$, where $\beta \approx 1.29 \times 10^{-6}$ is the contraction of the raw coordinates over one seven-cycle. The figure is an illustration; the rigorous statements are those of Certificate 5.1.

the delayed Rayleigh quotient, so that $|d_k^i| = |d_0^i| \prod_{t<k} |a_t - \lambda_i|/a_t$. All conclusions of Theorem 2.1 therefore follow from a single uniform one-step inequality,

$$10^{-6} \le \frac{|a_k - \lambda_i|}{a_k} \le 0.61 \qquad \text{for every } i \text{ and every } k \ge 0: \tag{14}$$

the lower bound is a nonresonance condition (a_k never comes too close to an eigenvalue), the upper bound forces convergence. Since the gradient itself tends to zero, its scale is removed: the pair (a_k, p_k) of the reciprocal step size and the normalized squared spectral weights evolves under a rational map F_λ on $\mathbb{R} \times \Delta_{n-1}$, with the admissible first step encoded by an initialization map D_λ. If this projective state approaches a periodic orbit that is separated from the spectrum, all late multipliers in (14) lie in a fixed compact subinterval of $(0, 1)$, and the finite initial segment is handled by continuity.

The certified periodic object (§5). The construction is carried out in dimension four. The unknowns are a free eigenvalue $\bar\lambda_3$, a projective state z, and an admissible initial weight r; they satisfy eight rational equations, namely $F_{\bar\lambda}^7(z) = z$ (a seven-cycle) together with $F_{\bar\lambda}^{15}(D_{\bar\lambda}(r)) = z$ (an exact fifteen-step landing). The landing block is forced by a dimension count: the projective state space is four-dimensional while the admissibly initialized states form a three-dimensional hypersurface, so an attracting cycle need not meet the forward orbit of any genuine BB1 initialization. A Newton-like map $H(u) = u - RG(u)$ with an exact rational preconditioner R is then evaluated in directed interval arithmetic on a box X of radius 10^{-55}; the certified inclusions $H(X) \subset \operatorname{int} X$ and $\sup_X \|DH\|_\infty < 1$ make Banach's theorem produce an exact root u_* (Lemma 5.2). The same run certifies that every normalization denominator (namely Z_t) is positive, that the cycle and the connection stay separated from all eigenvalues, and that an explicit Lyapunov pair (P, Q_7) is positive definite. Two logically distinct contractions appear here: H only certifies the root; attraction is proved separately.

Attraction, rates, and openness in dimension four (§6–§7). The certified inequality $P - A_7^T P A_7 \succ 0$ gives $\|A_7\|_P < 1$, so the seven-cycle is Schur stable and possesses an open basin (Lemma 6.3). The rate constants come from two independent mechanisms. A translation $\lambda \mapsto \lambda + 4\mathbf{1}$ commutes with the projective dynamics (Lemma 7.1) and moves the spectrum into (4.99, 8.01) with width below 3.02; since every a_k is a convex combination of eigenvalues, this alone yields the upper factor $3.02/4.99 < 0.61$. The certified separation yields the lower factor $10^{-5}/8.01 > 10^{-6}$. Because 1 is not an eigenvalue of A_7, the implicit-function theorem continues the cycle to all nearby spectra, and finite-time continuity of the fifteen-step map carries all nearby admissible initializations into the trapping tube: one trajectory becomes an open product $L_4 \times P_4$ (Proposition 7.2).

Every dimension, and the lift to BB1 problems (§8–§9). For $n > 4$, the extra eigenvalues are placed in $(7.49, 7.51)$ with zero initial weight, so the four-dimensional cycle survives on an invariant simplex face. The seven-step derivative is block upper triangular there, and the certificate bounds every transverse multiplier by $\tau(\nu) < 0.049$, so the boundary cycle attracts in the new directions as well; tilting the extra weights to small positive values and continuing in all n eigenvalues gives open sets $L_n \times P_n$ for each fixed finite n (Proposition 8.1), with the rate constants—though not the neighborhood sizes—uniform in n. Finally, three changes of variables lift the projective family to problem data: the orthant polar map $d_0^i = \sigma_i r \sqrt{p_{0,i}}$, the continuity of the spectral-projector weights $p(A, g)$ on simple-spectrum matrices, and the diffeomorphism $(A, x) \mapsto (A, Ax - b)$. The resulting set Ω_n of problems and initial points is open and nonempty, hence of positive Lebesgue measure, and Theorem 2.1 follows.

4 Exact reduction of BB1 to a projective rational map

We first derive the scalar recurrence without suppressing the exceptional index $k = 0$.

Lemma 4.1 (The BB1 step on a quadratic)**.** *Suppose $A \succ 0$ and $g_{k-1} \neq 0$. If $x_k = x_{k-1} - \alpha_{k-1} g_{k-1}$ with $\alpha_{k-1} \neq 0$, then the BB1 quotient in (5) is positive and equals*

$$\alpha_k = \frac{g_{k-1}^T g_{k-1}}{g_{k-1}^T A g_{k-1}}. \tag{15}$$

In particular, the initialization (4) gives $\alpha_1 = \alpha_0$.

Proof. Because $g_k = Ax_k - b$ and $g_{k-1} = Ax_{k-1} - b$, subtraction gives

$$y_{k-1} = g_k - g_{k-1} = A(x_k - x_{k-1}) = As_{k-1}.$$

The preceding gradient step gives $s_{k-1} = -\alpha_{k-1} g_{k-1}$. Therefore

$$\begin{aligned} s_{k-1}^T s_{k-1} &= (-\alpha_{k-1} g_{k-1})^T (-\alpha_{k-1} g_{k-1}) = \alpha_{k-1}^2 g_{k-1}^T g_{k-1}, \\ s_{k-1}^T y_{k-1} &= s_{k-1}^T A s_{k-1} = \alpha_{k-1}^2 g_{k-1}^T A g_{k-1}. \end{aligned}$$

Since $A \succ 0$ and $g_{k-1} \neq 0$, the denominator $g_{k-1}^T A g_{k-1}$ is strictly positive. Since $\alpha_{k-1}^2 > 0$, cancellation yields (15); both its numerator and denominator are positive. Taking $k = 1$ and using (4) proves $\alpha_1 = \alpha_0$. □

Exact reduction (§4)

$$d^i_{k+1} = \frac{a_k - \lambda_i}{a_k} d^i_k, \qquad (a_k, p_k) = F^k_\lambda\big(D_\lambda(p_0)\big)$$

goal: $10^{-6} \leq |a_k - \lambda_i| / a_k \leq 0.61$ for all i, k

↓ *interval Newton + Banach (Lem. 5.2)*

Certified periodic object (§5, dimension 4)

eight equations $G(u_*) = 0$:

seven-cycle $F^7(z_*) = z_* \oplus$ admissible landing $F^{15}(D(r_*)) = z_*$;

certified $Z_t > 0$, separations $\left|a_t - \bar{\lambda}_i\right|$, Lyapunov pair (P, Q_7), transverse $\tau(\bar{\nu})$

↓ *discrete Lyapunov inequality (Lem. 6.2, 6.3)*

Attraction (§6)

$P \succ 0,\ \ P - A_7^T P A_7 \succ 0 \implies \|A_7\|_P < 1 \implies$ the cycle has an open basin

↓ *translation conjugacy + IFT (Lem. 7.1, Prop. 7.2)*

Rates and openness in dimension 4 (§7)

translation $\lambda \mapsto \lambda + 4\mathbf{1} \Rightarrow$ upper factor 0.61;

certified separation $\Rightarrow$ lower factor 10^{-6};

hyperbolicity $\Rightarrow$ open family $L_4 \times P_4$

↓ *transverse Floquet bound (Prop. 8.1)*

Every finite dimension (§8)

boundary cycle on an invariant simplex face;

transverse multipliers $\tau(\nu) < 0.049$

$\Rightarrow$ attraction persists $\Rightarrow$ open $L_n \times P_n$ for each $n \geq 4$

↓ *three changes of variables (§9)*

Lift and measure (§9)

Θ_σ, projector weights $p(A, g)$, $(A, x) \mapsto (A, Ax - b)$ diffeomorphisms

$\Rightarrow \Omega_n$ open and nonempty $\Rightarrow$ positive measure $\Rightarrow$ Theorem 2.1

Figure 2: Map of the proof of Theorem 2.1. The shaded box is the computer-assisted step (Certificate 5.1); every other stage is a pen-and-paper argument that consumes its certified strict inequalities. Arrow labels name the tool carrying each implication.

Fix an orthogonal diagonalization

$$A = Q\Lambda Q^T, \qquad \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \qquad 0 < \lambda_1 < \cdots < \lambda_n, \tag{16}$$

and define the spectral gradient coordinates

$$d_k := Q^T g_k, \qquad d^i_k = (Q^T g_k)_i. \tag{17}$$

Because Q is orthogonal,

$$g_k^T g_k = \sum_{j=1}^n (d^j_k)^2, \qquad g_k^T A g_k = d_k^T \Lambda d_k = \sum_{j=1}^n \lambda_j (d^j_k)^2. \tag{18}$$

The gradient update is

$$g_{k+1} = A(x_k - \alpha_k g_k) - b = (Ax_k - b) - \alpha_k A g_k = (I - \alpha_k A) g_k.$$

Multiplication by Q^T and use of $Q^T A Q = \Lambda$ give

$$d_{k+1}^i = (1 - \alpha_k \lambda_i) d_k^i. \tag{19}$$

For $k = 0$, substitution of (4) and (18) into (19) yields

$$d_1^i = d_0^i \frac{\sum_{j=1}^n (\lambda_j - \lambda_i)(d_0^j)^2}{\sum_{j=1}^n \lambda_j (d_0^j)^2}. \tag{20}$$

For $k \geq 1$, Lemma 4.1 instead uses g_{k-1}, and hence

$$d_{k+1}^i = d_k^i \frac{\sum_{j=1}^n (\lambda_j - \lambda_i)(d_{k-1}^j)^2}{\sum_{j=1}^n \lambda_j (d_{k-1}^j)^2}. \tag{21}$$

Equations (20)–(21) are exactly the delayed recurrence certified below.

We now remove the common magnitude of d_k. Whenever $d_k \neq 0$, define

$$x_{k,i} := (d_k^i)^2, \qquad S_k := \sum_{j=1}^n x_{k,j}, \qquad p_{k,i} := \frac{x_{k,i}}{S_k}. \tag{22}$$

Then p_k belongs to the simplex

$$\Delta_{n-1} := \left\{ p \in \mathbb{R}^n : p_i \geq 0, \ \sum_{i=1}^n p_i = 1 \right\}.$$

For $p \in \Delta_{n-1}$ and $a \in \mathbb{R}$, define

$$\mu_\lambda(p) := \sum_{i=1}^n \lambda_i p_i, \tag{23}$$

$$Z_\lambda(a, p) := \sum_{i=1}^n p_i (a - \lambda_i)^2, \tag{24}$$

$$(T_{\lambda,a}(p))_i := \frac{p_i (a - \lambda_i)^2}{Z_\lambda(a, p)}, \tag{25}$$

$$F_\lambda(a, p) := (\mu_\lambda(p), T_{\lambda,a}(p)), \qquad D_\lambda(p) := (\mu_\lambda(p), p). \tag{26}$$

The map $T_{\lambda,a}$ is used only if $Z_\lambda(a, p) > 0$.

Lemma 4.2 (Exact projective state and multiplier identities)**.** *Suppose $d_0 \neq 0$, and let p_0 be defined by (22). Set*

$$a_0 := \mu_\lambda(p_0), \qquad a_k := \mu_\lambda(p_{k-1}) \quad (k \geq 1). \tag{27}$$

As long as the recurrence is defined,

$$(a_k, p_k) = F_\lambda^k(D_\lambda(p_0)) \qquad (k \geq 0), \tag{28}$$

and, for every i and $k \geq 0$,

$$d_{k+1}^i = d_k^i \frac{a_k - \lambda_i}{a_k}, \tag{29}$$

$$\left| d_k^i \right| = \left| d_0^i \right| \prod_{t=0}^{k-1} \frac{|a_t - \lambda_i|}{a_t}. \tag{30}$$

The empty product for $k = 0$ is one. In particular, $a_1 = a_0 = \mu_\lambda(p_0)$.

Proof. For $k \geq 1$, the scalar quotient in (21) is

$$\frac{\sum_j (\lambda_j - \lambda_i)(d_{k-1}^j)^2}{\sum_j \lambda_j (d_{k-1}^j)^2} = \frac{\sum_j \lambda_j x_{k-1,j} - \lambda_i \sum_j x_{k-1,j}}{\sum_j \lambda_j x_{k-1,j}}$$
$$= \frac{S_{k-1} \sum_j \lambda_j p_{k-1,j} - \lambda_i S_{k-1}}{S_{k-1} \sum_j \lambda_j p_{k-1,j}}$$
$$= \frac{\mu_\lambda(p_{k-1}) - \lambda_i}{\mu_\lambda(p_{k-1})} = \frac{a_k - \lambda_i}{a_k}.$$

The same computation applied to (20), with p_0 in place of p_{k-1}, gives the formula for $k = 0$. This proves (29).

Squaring (29) gives

$$x_{k+1,i} = x_{k,i} \frac{(a_k - \lambda_i)^2}{a_k^2}. \tag{31}$$

Summing (31) over i and using $x_{k,i} = S_k p_{k,i}$ yields

$$S_{k+1} = \sum_i x_{k,i} \frac{(a_k - \lambda_i)^2}{a_k^2} = \frac{S_k}{a_k^2} \sum_i p_{k,i}(a_k - \lambda_i)^2 = \frac{S_k}{a_k^2} Z_\lambda(a_k, p_k).$$

If $Z_\lambda(a_k, p_k) > 0$, division of (31) by this identity gives

$$p_{k+1,i} = \frac{p_{k,i}(a_k - \lambda_i)^2}{Z_\lambda(a_k, p_k)} = (T_{\lambda,a_k}(p_k))_i.$$

Definition (27) also gives $a_{k+1} = \mu_\lambda(p_k)$. Hence

$$F_\lambda(a_k, p_k) = (a_{k+1}, p_{k+1}).$$

Since $(a_0, p_0) = D_\lambda(p_0)$, induction on k proves (28). Finally, taking absolute values in (29) and multiplying the identities for $t = 0, \ldots, k-1$ proves (30).

Because all $\lambda_i > 0$ and p is a probability vector, $\mu_\lambda(p) \geq \lambda_1 > 0$. Thus a_k never causes division by zero. The only remaining possible singularity is $Z_\lambda(a_k, p_k) = 0$; the construction below keeps every orbit in a region where Z is strictly positive. $\square$

5 The certified four-dimensional periodic object

The only computer-assisted part of the proof is the validation of one exact zero of eight rational equations and of several strict inequalities at that zero. The validation is performed by the supplied script `interval_cert.py`, provided in Appendix A. The script uses only Python's standard `decimal` module and contains all entries of the rational preconditioner and Lyapunov matrices.

5.1 The eight equations

Begin with the unshifted four-point spectrum

$$\bar{\lambda} = (1,\ 1.8786699041860466,\ \bar{\lambda}_3,\ 4). \tag{32}$$

The displayed second eigenvalue is a terminating decimal and is interpreted as the exact corresponding rational number. Use the direct state chart

$$z = (a, p_1, p_2, p_3), \qquad p_4 = 1 - p_1 - p_2 - p_3,$$

and the initialization chart

$$r = (r_1, r_2, r_3), \qquad r_4 = 1 - r_1 - r_2 - r_3.$$

The unknown is

$$u = (\bar{\lambda}_3, a, p_1, p_2, p_3, r_1, r_2, r_3) \in \mathbb{R}^8.$$

Define, wherever the iterates are well-defined,

$$G(u) := \begin{pmatrix} F_{\bar{\lambda}}^{7}(z) - z \\ F_{\bar{\lambda}}^{15}(D_{\bar{\lambda}}(r)) - z \end{pmatrix} \in \mathbb{R}^8. \tag{33}$$

The first four equations require a point fixed by the seven-step map. The last four require an exact fifteen-step connection from an admissible BB initial state $D_{\bar{\lambda}}(r)$. The landing condition is essential: a periodic state in the four-dimensional (a, p) space need not lie in the forward orbit of the three-dimensional admissible-initialization hypersurface $a = \mu(p)$.

Let $c = (c_1, \ldots, c_8)$ be the exact terminating-decimal vector

$$\begin{aligned}
c_1 &= 2.7127814910653372282743021917166259612213120097331194358155052455762957581\,1,\\
c_2 &= 1.3334496725238812991898003496838493380028119662647089762866386716511125365\,6,\\
c_3 &= 0.8281085983625578302847062799336426754477549077553810635365940227268902274\,54,\\
c_4 &= 0.1670430513702266313898094627595023197810382863138141770388781875255927909\,29,\\
c_5 &= 0.0047179573503674129528125339395019941688067981035232679520880506589411846\,44,\\
c_6 &= 0.0000134113750442893933116720965327494222816004773939806475307653666721805974\,245,\\
c_7 &= 0.9891717057690781490472697115151355358471282234279160684741637435269262337\,31,\\
c_8 &= 0.0000034926002843914558913346790346936487552351777485127115413021962990464337\,9924.
\end{aligned}$$

Set

$$X := c + [-10^{-55}, 10^{-55}]^8. \tag{34}$$

At the center, and uniformly throughout X, the eliminated probabilities satisfy

$$\begin{aligned}
p_4 &\in [0.000130392916848125372671723367353010602400007827281491\,1,\\
&\qquad 0.000130392916848125372671723367353010602400007827281491\,8], && (35)\\
r_4 &\in [0.010811390255593170103527281709297021081834940916941437\,8,\\
&\qquad 0.010811390255593170103527281709297021081834940916941438\,5]. && (36)
\end{aligned}$$

Thus p and r are in the relative interior of Δ_3, and $1 < 1.8786699041860466 < \bar{\lambda}_3 < 4$ throughout X.

5.2 What the interval calculation proves

Computer-assisted certificate 5.1 (Validated root, stability, and separation)**.** With G, c, and X defined above, the supplied directed-rounding calculation proves all of the following statements.

(C1) There is an explicitly recorded exact rational matrix $R \in \mathbb{R}^{8\times 8}$ such that, with $H(u) = u - RG(u)$,

$$H(X) \subset c - RG(c) + \big(I - R[DG(X)]\big)(X - c) \subset \operatorname{int} X. \tag{37}$$

The largest coordinate ratio of the final box radius to 10^{-55} is less than

$$3.876379 \times 10^{-20}.$$

(C2) The same derivative enclosure satisfies

$$\sup_{u\in X}\|DH(u)\|_\infty \le \|I - R[DG(X)]\|_\infty < 7.200691\times 10^{-27} < 1. \tag{38}$$

(C3) Every normalization denominator occurring in the seven cycle steps and the fifteen connection steps is positive. More explicitly, the same outward evaluation gives

$$\min_{0\le t<7} Z_t > 0.1516357118221540, \qquad \min_{0\le t<15} Z_t^{\mathrm{conn}} > 0.0009510722601452239. \tag{39}$$

(C4) For every exact $u \in X$, the finite connection and cycle obey the uniform separation bounds

$$\min_{\substack{0\le t\le 15\\ 1\le i\le 4}} \left|a_t - \bar\lambda_i\right| > 1.3886903675\times 10^{-4}, \tag{40}$$

$$\min_{\substack{0\le t<7\\ 1\le i\le 4}} \left|a_t - \bar\lambda_i\right| > 7.5691389311\times 10^{-2}. \tag{41}$$

Both statements remain true, with the same displayed lower bounds, when an additional test eigenvalue $\bar\nu \in [3.49, 3.51]$ is included in the minimum.

(C5) Let $A_7 = D_z F_{\bar\lambda}^7(z)$, with z ranging over the z-part of X. The script records an exact symmetric rational matrix P and proves, for the exact root obtained below,

$$P \succ 0, \qquad Q_7 := P - A_7^T P A_7 \succ 0. \tag{42}$$

The outward-rounded LDL^T pivots for P have the positive lower bounds

$$977656.6634223271, \quad 65683881.36341604, \quad 9.217413243679623, \quad 323.9590422006289, \tag{43}$$

and the row diagonal-dominance margins for Q_7 have the lower bounds

$$\begin{aligned} &0.9999999999999999999976374, \quad 0.9999999999999999994270558,\\ &0.9999999999999999994222700, \quad 0.9999999999999999993948258. \end{aligned} \tag{44}$$

(C6) For every $\bar\nu \in [3.49, 3.51]$, the seven-step transverse multiplier defined in (67) below satisfies

$$0.0230006289035345 < \tau(\bar\nu) < 0.0482036188628888 < 1. \tag{45}$$

(C7) For the first two phases of the cycle,

$$a_0 - a_1 > 0.1782019618055700634. \tag{46}$$

These machine statements are rigorous real inequalities by the standard directed-rounding validation principle (Krawczyk, 1969; Rump, 2010). A decimal literal is first converted to a `Decimal` exactly. For each addition, subtraction, multiplication, and reciprocal, the script evaluates once with rounding toward $-\infty$ and once with rounding toward $+\infty$; if the input intervals contain the exact real inputs, the output interval therefore contains the exact result of that elementary operation. Structural induction over an arithmetic expression then shows that its interval evaluation encloses its entire real range over the input box. The automatic-differentiation class stores a pair (v, d) of an interval value and a vector of interval partial derivatives and propagates them by the exact

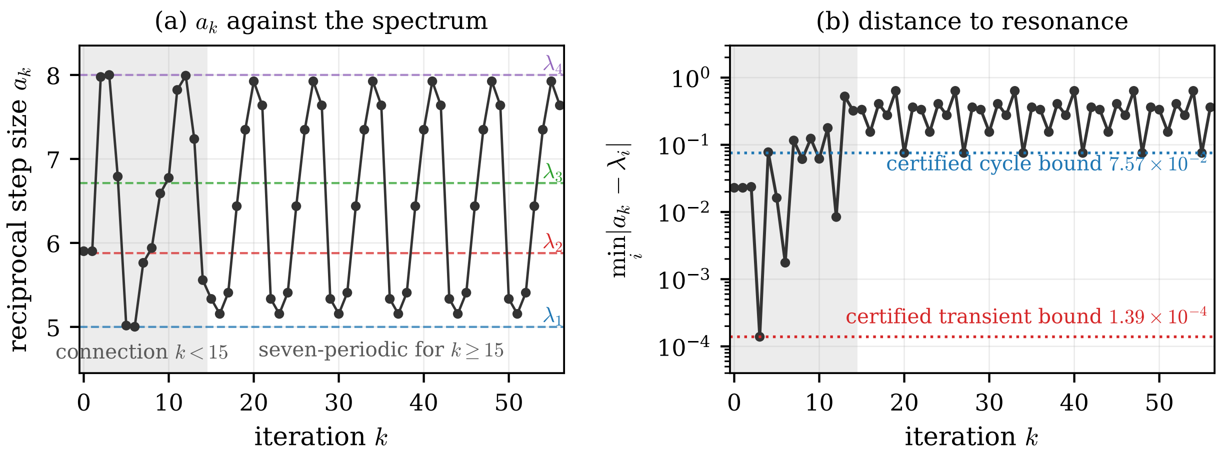

Figure 3: The certified projective orbit at the central parameter. *(a)* The reciprocal step size a_k against the eigenvalues: the admissible initialization lands on the seven-cycle after exactly fifteen steps and then repeats. *(b)* The distance from a_k to the nearest eigenvalue. The certified separation bounds of Certificate 5.1(C4) are sharp for this orbit: the transient minimum at $k = 3$ and the cycle minimum match the certified lower bounds to the displayed digits.

identities $D(u+v) = Du + Dv$, $D(uv) = vDu + uDv$, $D(u^{-1}) = -u^{-2}Du$; the same induction, applied simultaneously to values and derivatives, shows that the computed matrix $[DG(X)]$ contains $DG(u)$ entrywise for every $u \in X$. Each reciprocal operation asserts that its denominator interval excludes zero, and the explicit positive bounds (39) strengthen this check. All magnitudes used are far from the decimal exponent limits, so no overflow or underflow is involved.

The centered enclosure in (37) follows from the identity

$$\begin{aligned} H(u) - c &= H(c) - c + \int_0^1 DH(c + t(u-c))(u-c)\,dt \\ &= -RG(c) + \int_0^1 (I - RDG(c + t(u-c)))(u-c)\,dt. \end{aligned}$$

valid because X is convex and H is C^1 on a neighborhood of X. Every derivative matrix in the integral belongs entrywise to $I - R[DG(X)]$, and interval matrix–box multiplication encloses every product with $u - c \in X - c$ as well as its average over $t \in [0,1]$. The second inclusion in (37) is the directed-rounding output.

We next turn the interval statements into exact analytic conclusions.

Lemma 5.2 (Existence and uniqueness of an exact solution)**.** *There is a unique $u_* \in X$ satisfying $G(u_*) = 0$.*

Proof. The box X is a nonempty closed subset of $(\mathbb{R}^8, \|\cdot\|_\infty)$ and hence a complete metric space. Certificate 5.1(C1) gives $H(X) \subset \operatorname{int} X \subset X$, so H maps X into itself, and by Certificate 5.1(C3) every denominator is nonzero on an open neighborhood of X, so H is C^1 there. For $u, v \in X$ the segment $[v, u]$ lies in the convex set X, so the mean-value inequality together with Certificate 5.1(C2) gives

$$\|H(u) - H(v)\|_\infty \leq (7.200691 \times 10^{-27}) \|u - v\|_\infty .$$

Banach's fixed-point theorem yields a unique $u_* \in X$ with $H(u_*) = u_*$.

The bound (38) also forces R to be invertible: if $y^T R = 0$ for some $y \neq 0$, then $y^T(I - RDG(u)) = y^T$, so 1 would be an eigenvalue of $(I - RDG(u))^T$, contradicting $\rho(I - RDG(u)) \leq \|I - RDG(u)\|_\infty < 1$. Hence $H(u_*) = u_*$ reads $RG(u_*) = 0$ and gives $G(u_*) = 0$. Conversely, every zero of G in X is a fixed point of H, so u_* is the only such zero. $\square$

Write the exact components as

$$u_* = (\bar{\lambda}_3^*, z_*, r_*), \qquad z_* = (a_*, p_*). \tag{47}$$

Unpacking $G(u_*) = 0$ gives the exact, rather than numerical, identities

$$F_{\bar{\lambda}}^7(z_*) = z_*, \qquad F_{\bar{\lambda}}^{15}(D_{\bar{\lambda}}(r_*)) = z_*. \tag{48}$$

The third eigenvalue lies in the radius-10^{-55} interval centered at c_1, so in particular

$$\bar{\lambda}_3^* = 2.71278149106533722827430219171662596122\ldots.$$

Equation (46) shows $F_{\bar{\lambda}}(z_*) \neq z_*$. The least period of z_* divides 7; since 7 is prime and the period is not 1, it equals 7.

6 Floquet stability and nonlinear attraction

Let

$$A_7 := D_z F_{\bar{\lambda}}^7(z_*). \tag{49}$$

The derivative exists because all seven denominators are positive. We now derive every stability implication of Certificate 5.1(C5).

Lemma 6.1 (The certified matrices are positive definite)**.** *The exact matrices P and $Q_7 = P - A_7^T P A_7$ satisfy $P \succ 0$ and $Q_7 \succ 0$.*

Proof. The interval LDL^T recursion is performed without pivoting; every division is valid because each preceding pivot interval is positive. It encloses an exact factorization $P = LDL^T$ with L unit lower triangular and the four diagonal entries of D bounded below by the positive numbers in (43); hence $P \succ 0$. The matrix Q_7 is symmetric, and the interval calculation proves strict row diagonal dominance,

$$(Q_7)_{ii} - \sum_{j \neq i} |(Q_7)_{ij}| \geq m_i > 0, \tag{50}$$

with the margins m_i bounded below by (44); in particular every diagonal entry is positive. By Gershgorin's circle theorem, every eigenvalue of Q_7 is positive, so $Q_7 \succ 0$. $\square$

Lemma 6.2 (Schur stability of the seven-step derivative)**.** *Every complex eigenvalue of A_7 has modulus strictly smaller than one.*

Proof. Let $v \in \mathbb{C}^4 \setminus \{0\}$ and $\zeta \in \mathbb{C}$ satisfy $A_7 v = \zeta v$. A real symmetric positive-definite matrix defines a positive Hermitian form on $\mathbb{C}^4$, so $v^* P v > 0$ and $v^* Q_7 v > 0$. Using that A_7 and P are real,

$$v^* Q_7 v = v^* P v - (A_7 v)^* P (A_7 v) = (1 - |\zeta|^2) v^* P v.$$

Both sides are strictly positive, so $|\zeta| < 1$. $\square$

Lemma 6.3 (Nonlinear attraction in an adapted norm)**.** *The point z_* is a locally attracting fixed point of $\Phi := F_\lambda^7$.*

Proof. Define $\|x\|_P = (x^T P x)^{1/2}$. Since $P \succ 0$, this is a norm. On the compact P-unit sphere $S_P = \{x : x^T P x = 1\}$ the continuous function $x \mapsto x^T Q_7 x$ is strictly positive by Lemma 6.1 and therefore has a positive minimum $\delta > 0$; by homogeneity,

$$x^T Q_7 x \geq \delta x^T P x \qquad \text{for every } x \in \mathbb{R}^4. \tag{51}$$

Since $Q_7 = P - A_7^T P A_7$, (51) gives

$$\|A_7 x\|_P^2 = x^T (P - Q_7) x \leq (1 - \delta) \|x\|_P^2 ,$$

and the left side is nonnegative, so $0 < \delta \leq 1$ and

$$\|A_7\|_P \leq q_0 := \sqrt{1 - \delta} < 1. \tag{52}$$

All normalization denominators at the cycle phases are positive by (39) and remain positive on a small Euclidean neighborhood of each phase, so Φ is C^1 near z_* with continuous derivative and $D\Phi(z_*) = A_7$. Choose any q with $q_0 < q < 1$. Continuity of $D\Phi$ in the induced P-norm gives $\varepsilon > 0$ such that

$$\|D\Phi(z)\|_P \leq q \quad \text{whenever} \quad \|z - z_*\|_P \leq \varepsilon. \tag{53}$$

The closed P-ball in (53) is convex, so for any z in this ball the mean-value inequality gives

$$\|\Phi(z) - z_*\|_P = \|\Phi(z) - \Phi(z_*)\|_P \leq q \|z - z_*\|_P .$$

Thus the ball is forward invariant under Φ, and induction gives $\|\Phi^m(z) - z_*\|_P \leq q^m \|z - z_*\|_P \to 0$. $\square$

7 Translation and an open four-dimensional family

The projective dynamics depends only on differences $a - \lambda_i$, whereas the original BB multiplier also contains the denominator a. A spectral translation exploits this distinction.

Lemma 7.1 (Translation conjugacy)**.** *For $s \in \mathbb{R}$, let $\lambda' = \lambda + s\mathbf{1}$ and $C_s(a, p) = (a + s, p)$. Wherever the maps are defined,*

$$F_{\lambda'} \circ C_s = C_s \circ F_\lambda, \qquad D_{\lambda'} = C_s \circ D_\lambda. \tag{54}$$

Proof. Because $\sum_i p_i = 1$,

$$\mu_{\lambda'}(p) = \sum_i (\lambda_i + s) p_i = \mu_\lambda(p) + s.$$

Also $(a + s) - (\lambda_i + s) = a - \lambda_i$, so

$$Z_{\lambda'}(a + s, p) = \sum_i p_i ((a + s) - (\lambda_i + s))^2 = Z_\lambda(a, p),$$

and every normalized weight in (25) is unchanged. Substitution into (26) proves both identities. $\square$

Take $s = 4$. The shifted central spectrum is

$$\lambda^* = (5,\ 5.8786699041860466,\ 6.7127814910653372228\ldots,\ 8). \tag{55}$$

Translation leaves the projective weights, all differences, all Z values, and all state derivatives unchanged; it adds four only to the scalar a. Let $z_*^+ = C_4(z_*)$. Then

$$F_{\lambda^*}^7(z_*^+) = z_*^+, \qquad F_{\lambda^*}^{15}(D_{\lambda^*}(r_*)) = z_*^+. \tag{56}$$

Proposition 7.2 (Uniform factor bounds in dimension four)**.** *There are nonempty open neighborhoods*

$$L_4 \subset \{\lambda \in \mathbb{R}^4 : 0 < \lambda_1 < \cdots < \lambda_4\} \quad \textit{and} \quad P_4 \subset \operatorname{int}\Delta_3$$

of λ^* *and* r_**, respectively, such that for every* $(\lambda, p_0) \in L_4 \times P_4$*, the orbit* $(a_k, p_k) = F_\lambda^k(D_\lambda(p_0))$ *is defined for all* k *and*

$$10^{-6} \le \frac{|a_k - \lambda_i|}{a_k} \le 0.61 \qquad (1 \le i \le 4,\ k \ge 0). \tag{57}$$

Proof. We divide the argument into continuation, uniform attraction, finite-time entry, and the numerical factor bounds.

Step 1: continuation of the seven-cycle. Define $\Psi(\lambda, z) := F_\lambda^7(z) - z$. All seven Z denominators are positive at the cycle by (39) and, by continuity, positive nearby, so Ψ is C^1 near (λ^*, z_*^+). Identity (56) gives $\Psi(\lambda^*, z_*^+) = 0$, and

$$D_z\Psi(\lambda^*, z_*^+) = A_7 - I$$

is invertible because 1 is not an eigenvalue of A_7 (Lemma 6.2). The implicit-function theorem gives an open neighborhood U_λ of λ^* and a C^1 periodic-point function $z(\lambda)$ satisfying

$$F_\lambda^7(z(\lambda)) = z(\lambda), \qquad z(\lambda^*) = z_*^+. \tag{58}$$

Step 2: uniform local attraction. At the central point, $P - A_7^T P A_7 \succ 0$. Positive definiteness is open in the entries: if $Q_7 \succ 0$, its minimum on the Euclidean unit sphere is some $m > 0$, and every symmetric perturbation E with $\|E\|_2 < m/2$ satisfies $x^T(Q_7 + E)x \ge m/2$ for every Euclidean unit x. The matrix $D_z F_\lambda^7(z(\lambda))$ depends continuously on λ, so after shrinking U_λ, the strict inequality

$$P - D_z F_\lambda^7(z(\lambda))^T P D_z F_\lambda^7(z(\lambda)) \succ 0$$

holds for every $\lambda \in U_\lambda$. Repeating the compact-unit-sphere argument from Lemma 6.3, now over the closure of a sufficiently small parameter ball, gives a common contraction constant $q < 1$ and a common radius $\varepsilon > 0$ such that the P-ball

$$B_\lambda := \{z : \|z - z(\lambda)\|_P < \varepsilon\} \tag{59}$$

is mapped into itself by F_λ^7 for every parameter in that smaller ball. Shrinking ε if necessary, all seven intermediate images $F_\lambda^j(B_\lambda)$, $0 \le j < 7$, stay in neighborhoods where $Z > 0$ and where every cycle-phase separation remains strict.

Step 3: admissible initial conditions enter the trapping tube. At the central parameter, (56) gives exact entry after fifteen steps. The map

$$(\lambda, r) \longmapsto F_\lambda^{15}(D_\lambda(r)) \tag{60}$$

is continuous near (λ^*, r_*): it is a finite composition of rational maps, and every denominator on the central connection is positive by (39). Since B_{λ^*} is an open neighborhood of z_*^+, continuity gives a joint open neighborhood W of (λ^*, r_*) whose image under (60) lies in the corresponding parameter-dependent trapping ball; W contains a product $L_4 \times P_4$ of open neighborhoods, and (35) and (36) allow P_4 to be chosen inside $\operatorname{int}\Delta_3$.

For any (λ, p_0) in this product, the first fifteen iterates are defined by finite-time continuity, the fifteenth state lies in B_λ, and thereafter the seven-step iterates remain in B_λ with the intermediate states in its seven-phase trapping tube. Hence the orbit is defined for all time.

Step 4: lower factor bound. At the central parameter and initial weight, the finite connection has absolute separation greater than $1.3886903675 \times 10^{-4}$ by (40); the cycle has the much larger separation (41). All relevant functions are continuous, so we may shrink the trapping tube and $L_4 \times P_4$ so that

$$|a_k - \lambda_i| > 10^{-5} \quad \text{for all } i \text{ and all } k \geq 0. \tag{61}$$

This is a finite continuity assertion for $0 \leq k \leq 15$ and a uniform seven-phase assertion in the forward-invariant trapping tube for $k \geq 15$.

Shrink L_4 further so that $\lambda_4 < 8.01$. Every scalar a_k along an admissible orbit is a convex combination of the eigenvalues, by $a_0 = \mu_\lambda(p_0)$ and $a_k = \mu_\lambda(p_{k-1})$ for $k \geq 1$; thus $a_k \leq \lambda_4 < 8.01$, and with (61),

$$\frac{|a_k - \lambda_i|}{a_k} > \frac{10^{-5}}{8.01} > 10^{-6}.$$

Step 5: upper factor bound. Shrink L_4 once more so that

$$\lambda_1 > 4.99, \qquad \lambda_4 < 8.01, \qquad \lambda_4 - \lambda_1 < 3.02. \tag{62}$$

For every convex combination $a_k \in [\lambda_1, \lambda_4]$ and every λ_i in the same interval,

$$|a_k - \lambda_i| \leq \lambda_4 - \lambda_1 < 3.02, \qquad a_k \geq \lambda_1 > 4.99,$$

so

$$\frac{|a_k - \lambda_i|}{a_k} < \frac{3.02}{4.99} < 0.61.$$

This completes all parts of (57). □

8 Extension to every finite dimension

Fix $n > 4$. We embed the four active eigenvalues at indices $1, 2, 3, n$ and choose any distinct central extra eigenvalues

$$7.49 < \nu_4 < \nu_5 < \cdots < \nu_{n-1} < 7.51. \tag{63}$$

Such a choice exists for every finite n: for example, equally spaced points in the interval $(7.49, 7.51)$ suffice. The resulting central ordered spectrum is

$$\lambda^{*,n} = (5,\ 5.8786699041860466,\ 6.7127814910653372 28\ldots, \nu_4, \ldots, \nu_{n-1},\ 8). \tag{64}$$

Give every extra coordinate zero projective weight. The corresponding face of Δ_{n-1} is invariant because each update has the form

$$p_i' = \frac{p_i(a - \lambda_i)^2}{Z(a, p)};$$

if $p_i = 0$, then $p_i' = 0$. Thus the four-dimensional seven-cycle persists as a boundary cycle of the n-dimensional projective map.

Use independent local coordinates

$$(a, p_1, p_2, p_3, q_4, \ldots, q_{n-1}),$$

where q_ℓ is the weight at ν_ℓ, and eliminate the active weight at the eigenvalue 8:

$$p_n = 1 - p_1 - p_2 - p_3 - \sum_{\ell=4}^{n-1} q_\ell.$$

At a boundary-cycle phase, $q = 0$ and

$$Z_t = \sum_{j \in \{1,2,3,n\}} p_{t,j}(a_t - \lambda_j)^2 > 0.$$

For an extra weight,

$$q'_\ell = \frac{q_\ell (a - \nu_\ell)^2}{Z(a,p,q)}. \tag{65}$$

Differentiate (65) at $q = 0$ by the quotient rule gives

$$\begin{aligned}
\left.\frac{\partial q'_\ell}{\partial q_\ell}\right|_{q=0} &= \left.\frac{(a-\nu_\ell)^2 Z - q_\ell (a-\nu_\ell)^2 (\partial Z/\partial q_\ell)}{Z^2}\right|_{q=0} = \frac{(a-\nu_\ell)^2}{Z}, \\
\left.\frac{\partial q'_\ell}{\partial q_m}\right|_{q=0} &= -\left.\frac{q_\ell (a-\nu_\ell)^2 (\partial Z/\partial q_m)}{Z^2}\right|_{q=0} = 0 \qquad (m \neq \ell).
\end{aligned}$$

Every derivative of q'_ℓ with respect to a base variable a, p_1, p_2, p_3 also contains the prefactor q_ℓ and is zero at $q = 0$. Thus the one-step derivative at phase t is block upper triangular:

$$\begin{pmatrix} B_t & C_t \\ 0 & \operatorname{diag}(\gamma_t(\nu_4), \dots, \gamma_t(\nu_{n-1})) \end{pmatrix}, \qquad \gamma_t(\nu) = \frac{(a_t - \nu)^2}{Z_t}. \tag{66}$$

Products of block upper-triangular matrices are block upper triangular, and their diagonal blocks are the products of the diagonal blocks. Therefore the full seven-step derivative M_n has the form

$$M_n = \begin{pmatrix} A_7 & * \\ 0 & \operatorname{diag}(\tau(\nu_4), \dots, \tau(\nu_{n-1})) \end{pmatrix}, \qquad \tau(\nu) := \prod_{t=0}^{6} \frac{(a_t - \nu)^2}{Z_t}. \tag{67}$$

The certificate evaluates the unshifted interval $\bar{\nu} \in [3.49, 3.51]$. Translation by four leaves every difference and Z_t unchanged, so (45) holds for every $\nu \in [7.49, 7.51]$, including every extra eigenvalue in (63).

The characteristic polynomial of a block upper-triangular matrix is the product of the characteristic polynomials of its diagonal blocks. Thus, the eigenvalues of M_n are those of A_7 together with the scalars $\tau(\nu_\ell)$. Lemma 6.2 and (45) imply

$$\rho(M_n) < 1. \tag{68}$$

Since $\rho(M_n) < 1$, we have $\left\| M_n^k \right\|_2 \le C\gamma^k$ for any $\gamma \in (\rho(M_n), 1)$ and some finite C, so the series

$$\widetilde{P}_n := \sum_{k=0}^{\infty} (M_n^T)^k M_n^k \tag{69}$$

converges, satisfies $\widetilde{P}_n \succeq I \succ 0$, and obeys the exact identity $\widetilde{P}_n - M_n^T \widetilde{P}_n M_n = I$. It therefore supplies an adapted Lyapunov norm in the full state space, even though the four-dimensional matrix P does not include the new coordinates, and thus the nonlinear attraction proof of Lemma 6.3 applies verbatim in the full state space.

The boundary cycle lies on a simplex face, but the rational map is smooth on an ordinary ambient Euclidean neighborhood because every cycle denominator is positive, so its attracting neighborhood intersects the simplex in a relative neighborhood containing interior points. To see this explicitly, start from the boundary initialization r_*, assign sufficiently small positive values to all extra weights, and subtract their sum from the active weight $r_{*,4}$ at eigenvalue 8, which is possible because $r_{*,4} > 0.0108$ by (36). The resulting weight lies in $\operatorname{int}\Delta_{n-1}$ and can be arbitrarily close to the boundary weight. The boundary admissible state lands on the cycle after fifteen steps; by continuity of the finite iterate, every sufficiently close interior weight lands in the attracting neighborhood. Choose one such interior weight and call it $r_*^{(n)}$.

We must also continue the cycle under perturbations of all n eigenvalues. The map $\Psi_n(\lambda, z) = F_\lambda^7(z) - z$ is C^1 near the central boundary cycle because the denominators are positive, and its state derivative $M_n - I$ is invertible by (68). The implicit-function theorem therefore continues the periodic point for every sufficiently small perturbation of all n eigenvalues, and the strict Lyapunov inequality, the finite-time entry property, and all separation inequalities persist after the parameter and initial neighborhoods are shrunk.

The certificate's separation test includes every $\bar{\nu} \in [3.49, 3.51]$, hence after translation every $\nu \in [7.49, 7.51]$. The factor-bound calculation in Steps 4–5 of Proposition 7.2 therefore applies to active and extra coordinates alike. We have proved the following.

Proposition 8.1 (Uniform factor bounds for every finite n)**.** *For every finite $n \geq 4$, there are nonempty open sets*

$$L_n \subset \{\lambda \in \mathbb{R}^n : 0 < \lambda_1 < \cdots < \lambda_n\}, \qquad P_n \subset \operatorname{int}\Delta_{n-1}, \tag{70}$$

such that every admissible projective orbit initialized by $(\lambda, p_0) \in L_n \times P_n$ is defined for all time and satisfies

$$10^{-6} \leq \frac{|a_k - \lambda_i|}{a_k} \leq 0.61 \qquad (1 \leq i \leq n,\ k \geq 0). \tag{71}$$

There is no hidden uniformity claim as $n \to \infty$: for each fixed finite n, finitely many distinct extra eigenvalues are selected and then a possibly n-dependent open neighborhood is taken. The rate constants 10^{-6} and 0.61, however, are the same for every finite n.

9 Lifting the projective construction to BB problems and measure

We first obtain componentwise gradient bounds for a diagonal quadratic. Fix $\lambda \in L_n$ and choose any $p_0 \in P_n$, any radius $r > 0$, and any sign vector $\sigma \in \{\pm 1\}^n$. Set

$$d_0^i := \sigma_i r \sqrt{p_{0,i}}. \tag{72}$$

Then $\sum_i (d_0^i)^2 = r^2$ and $(d_0^i)^2 / \sum_j (d_0^j)^2 = p_{0,i}$. Conversely, on any fixed open orthant, the map

$$\Theta_\sigma : (0, \infty) \times \operatorname{int}\Delta_{n-1} \to \{d \in \mathbb{R}^n : \operatorname{sign}(d_i) = \sigma_i\}, \qquad \Theta_\sigma(r, p)_i = \sigma_i r \sqrt{p_i}, \tag{73}$$

has the smooth inverse $r = \|d\|_2$, $p_i = d_i^2 / \|d\|_2^2$, and is therefore a diffeomorphism. In particular, the image of $(1, 2) \times P_n$ is a nonempty open set of initial spectral gradients.

Combining Proposition 8.1 with Lemma 4.2 gives, coordinate by coordinate,

$$\left|d_0^i\right| 10^{-6k} \leq \left|d_k^i\right| = \left|d_0^i\right| \prod_{t=0}^{k-1} \frac{|a_t - \lambda_i|}{a_t} \leq \left|d_0^i\right| 0.61^k.$$

This is the spectral version of (6). Because $P_n \subset \operatorname{int}\Delta_{n-1}$, every d_0^i is nonzero. The lower bound then makes every d_k^i, and hence every g_k, nonzero. Lemma 4.1 shows inductively that every BB denominator $s_{k-1}^T y_{k-1} = s_{k-1}^T A s_{k-1}$ is strictly positive. Thus the BB1 iteration is well-defined for all k, rather than merely being a formal solution of the scalar recurrence. The upper bound tends to zero, so $g_k \to 0$ and, because A is invertible, $x_k - x_* = A^{-1} g_k \to 0$.

We next allow arbitrary eigenvectors and prove openness in the full matrix space. Define

$$\mathcal{A}_n := \{A \in \mathbb{S}^n_{++} : \lambda_1(A) < \cdots < \lambda_n(A),\ (\lambda_1(A), \ldots, \lambda_n(A)) \in L_n\}. \tag{74}$$

This set is nonempty: it contains the diagonal matrix with any spectrum in L_n. It is open because the ordered eigenvalues of a real symmetric matrix are continuous functions of the matrix entries (Weyl's inequality gives $|\lambda_i(A) - \lambda_i(B)| \leq \|A - B\|_2$) and L_n is open.

For $A \in \mathcal{A}_n$ and $g \neq 0$, define the orientation-free spectral weights

$$p_i(A, g) := \frac{g^T \Pi_i(A) g}{g^T g} = \frac{\|\Pi_i(A) g\|^2}{\|g\|^2}. \tag{75}$$

They are nonnegative and sum to one because the spectral projectors are orthogonal and sum to I.

The map $(A, g) \mapsto p(A, g)$ is continuous on simple-spectrum matrices with $g \neq 0$: on the simple-spectrum set,

$$\Pi_i(A) = \prod_{j \neq i} \frac{A - \lambda_j(A) I}{\lambda_i(A) - \lambda_j(A)}. \tag{76}$$

To verify (76), apply the polynomial on its right to an eigenvector of A: it equals one on the ith eigenspace and zero on every other eigenspace. Every denominator is nonzero because the spectrum is simple. Therefore, continuity of the eigenvalues makes the right side locally continuous in A, and (75) is a quotient of continuous functions with positive denominator $g^T g$, proving the claim.

Define the open set of matrix–gradient pairs

$$\Gamma_n := \{(A, g) : A \in \mathcal{A}_n,\ 1 < \|g\| < 2,\ p(A, g) \in P_n\}. \tag{77}$$

It is open by the continuity just established. It is nonempty: take a diagonal A with spectrum in L_n, any $p \in P_n$, a sign vector σ, and $g_i = (3/2)\sigma_i \sqrt{p_i}$; then $\|g\| = 3/2$ and $p(A, g) = p$.

For the fixed vector b, the map

$$\mathcal{J}_b : \mathbb{S}^n_{++} \times \mathbb{R}^n \to \mathbb{S}^n_{++} \times \mathbb{R}^n, \qquad \mathcal{J}_b(A, x) = (A, Ax - b), \tag{78}$$

is a smooth bijection with smooth inverse $\mathcal{J}_b^{-1}(A, g) = (A, A^{-1}(g + b))$, matrix inversion being smooth on $\mathbb{S}^n_{++}$; it is therefore a diffeomorphism. Set

$$\Omega_n := \mathcal{J}_b^{-1}(\Gamma_n), \qquad \mathcal{X}_n(A) := \{x : (A, x) \in \Omega_n\}. \tag{79}$$

The set Ω_n is nonempty and open, and every fiber $\mathcal{X}_n(A)$ is nonempty and open: for a fixed $A \in \mathcal{A}_n$, one can choose a spectral-coordinate gradient from (72) and then put $x = A^{-1}(g + b)$. Every nonempty open subset of a finite-dimensional Euclidean space has positive Lebesgue measure; applied in dimensions $n(n+1)/2 + n$ and n, this proves the positive-measure assertions of Theorem 2.1.

It remains to prove the norm, objective, and non-superlinear conclusions. Because Q is orthogonal, (6) gives

$$10^{-12k} \|g_0\|^2 \leq \|g_k\|^2 = \sum_i \left|d_k^i\right|^2 \leq 0.61^{2k} \|g_0\|^2,$$

and taking nonnegative square roots proves (8). Since $g_k = A e_k$, one has, in the eigenbasis,

$$\|e_k\|_A^2 = g_k^T A^{-1} g_k = \sum_{i=1}^n \frac{|d_k^i|^2}{\lambda_i}, \tag{80}$$

and applying the component bounds term by term proves (9). Completing the square in (3) gives $f(x)-f(x_*) = \frac{1}{2}\|x - x_*\|_A^2$, so squaring (9) and multiplying by 1/2 proves (10). The lower bounds in (11) follow by taking kth roots, for example $\|g_k\|^{1/k} \geq 10^{-6}\|g_0\|^{1/k} \to 10^{-6}$. Finally, Q-superlinear convergence of a positive null sequence ($u_{k+1}/u_k \to 0$) implies $u_k^{1/k} \to 0$, so (11) rules out both root-superlinear and Q-superlinear convergence. All assertions of Theorem 2.1 are proved.

10 Scope and limitations

The theorem establishes a robust lower-geometric rate on a constructed open family. Four qualifications are important.

1. It is a result for BB1, not BB2. The initial step is the explicit inverse gradient Rayleigh quotient (4). An arbitrary first step has a related projective description but is not the admissible-initialization map certified here.

2. The conclusion holds on a nonempty open, positive-measure family of spectra and initial points. It is not asserted for every positive spectrum or for almost every spectrum.

3. The computer-assisted step proves existence of one exact orbit inside a tiny rational box. The displayed decimal for λ_3^* is a locator for that exact root, not a claim that the eigenvalue is itself a terminating decimal or has a closed form.

4. The constants 10^{-6} and 0.61 are safety margins. The proof uses a minimum transient separation of approximately 1.39×10^{-4} and an upper multiplier bound close to 0.6 at the central translated spectrum. No optimality is claimed.

References

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):1–16, 1959.

Roberto Andreani, Ernesto G Birgin, José Mario Martínez, and Jinyun Yuan. Spectral projected gradient and variable metric methods for optimization with linear inequalities. *IMA Journal of Numerical Analysis*, 25(2):221–252, 2005.

Behzad Azmi and Karl Kunisch. Analysis of the barzilai-borwein step-sizes for problems in hilbert spaces. *Journal of Optimization Theory and Applications*, 185(3):819–844, 2020.

Behzad Azmi and Karl Kunisch. On the convergence and mesh-independent property of the barzilai–borwein method for pde-constrained optimization. *IMA Journal of Numerical Analysis*, 42(4): 2984–3021, 2022.

Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

Ernesto G Birgin, José Mario Martínez, and Marcos Raydan. Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211, 2000.

Ernesto G Birgin, José Mario Martínez, and Marcos Raydan. Algorithm 813: Spg—software for convex-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 27(3): 340–349, 2001.

Ernesto G Birgin, Jose Mario Martínez, and Marcos Raydan. Spectral projected gradient methods: review and perspectives. *Journal of Statistical Software*, 60:1–21, 2014.

Silvia Bonettini, Riccardo Zanella, and Luca Zanni. A scaled gradient projection method for constrained image deblurring. *Inverse problems*, 25(1):015002, 2009.

Oleg Burdakov, Yuhong Dai, and Na Huang. Stabilized barzilai-borwein method. *Journal of Computational Mathematics*, pages 916–936, 2019.

Yu-Hong Dai. Alternate step gradient method. *Optimization*, 52(4-5):395–415, 2003.

Yu-Hong Dai. A new analysis on the barzilai-borwein gradient method. *Journal of the operations Research Society of China*, 1(2):187–198, 2013.

Yu-Hong Dai and Roger Fletcher. On the asymptotic behaviour of some new gradient methods. *Mathematical Programming*, 103(3):541–559, 2005a.

Yu-Hong Dai and Roger Fletcher. Projected barzilai-borwein methods for large-scale box-constrained quadratic programming. *Numerische Mathematik*, 100(1):21–47, 2005b.

Yu-Hong Dai and Li-Zhi Liao. R-linear convergence of the barzilai and borwein gradient method. *IMA Journal of numerical analysis*, 22(1):1–10, 2002.

Yu-Hong Dai, William W Hager, Klaus Schittkowski, and Hongchao Zhang. The cyclic barzilai—borwein method for unconstrained optimization. *IMA Journal of Numerical Analysis*, 26(3): 604–627, 2006.

Yu-Hong Dai, Yakui Huang, and Xin-Wei Liu. A family of spectral gradient methods for optimization. *arXiv preprint arXiv:1812.02974*, 2018.

Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.

Roger Fletcher. On the barzilai-borwein method. In *Optimization and control with applications*, pages 235–256. Springer, 2005.

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

Ana Friedlander, José Mario Martínez, Brigida Molina, and Marcus Raydan. Gradient method with retards and generalizations. *SIAM Journal on Numerical Analysis*, 36(1):275–289, 1998.

Juan Gao, Xin-Wei Liu, Yu-Hong Dai, Yakui Huang, and Peng Yang. Achieving geometric convergence for distributed optimization with barzilai-borwein step sizes. *Science China. Information Sciences*, 65(4):149204, 2022.

Luigi Grippo and Marco Sciandrone. Nonmonotone globalization techniques for the barzilai-borwein gradient method. *Computational Optimization and Applications*, 23(2):143–169, 2002.

Chang He, Wenzhi Gao, Bo Jiang, Madeleine Udell, and Shuzhong Zhang. New results on the polyak stepsize: Tight convergence analysis and universal function classes. *arXiv preprint arXiv:2512.06231*, 2025.

Ya-Kui Huang, Yu-Hong Dai, and Xin-Wei Liu. Equipping the barzilai–borwein method with the two dimensional quadratic termination property. *SIAM Journal on Optimization*, 31(4):3068–3096, 2021.

Yakui Huang, Hongwei Liu, and Shuisheng Zhou. Quadratic regularization projected barzilai–borwein method for nonnegative matrix factorization. *Data mining and knowledge discovery*, 29 (6):1665–1684, 2015.

Yakui Huang, Yu-Hong Dai, Xin-Wei Liu, and Hongchao Zhang. On the asymptotic convergence and acceleration of gradient methods. *Journal of Scientific Computing*, 90(1):7, 2022.

Bruno Iannazzo and Margherita Porcelli. The riemannian barzilai–borwein method with nonmonotone line search and the matrix geometric mean computation. *IMA Journal of Numerical Analysis*, 38(1):495–517, 2018.

Rudolf Krawczyk. Newton-algorithmen zur bestimmung von nullstellen mit fehlerschranken. *Computing*, 4(3):187–201, 1969.

Dawei Li and Ruoyu Sun. On a faster R-linear convergence rate of the barzilai-borwein method, 2021.

Luc Pronzato, Henry P Wynn, and Anatoly A Zhigljavsky. Asymptotic behaviour of a family of gradient algorithms in $\mathbb{R}^d$ and Hilbert spaces. *Mathematical programming*, 107(3):409–438, 2006.

Marcos Raydan. On the barzilai and borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13(3):321–326, 1993.

Marcos Raydan. The barzilai and borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, 7(1):26–33, 1997.

Siegfried M Rump. Verification methods: Rigorous results using floating-point arithmetic. In *Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, pages 3–4, 2010.

Thomas Serafini, Gaetano Zanghirati, and Luca Zanni. Gradient projection methods for quadratic programs and applications in training support vector machines. *Optimization Methods and Software*, 20(2-3):353–378, 2005.

Conghui Tan, Shiqian Ma, Yu-Hong Dai, and Yuqiu Qian. Barzilai-borwein step size for stochastic gradient descent. volume 29, 2016.

Ewout Van Den Berg and Michael P Friedlander. Probing the pareto frontier for basis pursuit solutions. *Siam journal on scientific computing*, 31(2):890–912, 2009.

Yanfei Wang and Shiqian Ma. Projected barzilai–borwein method for large-scale nonnegative image restoration. *Inverse Problems in Science and Engineering*, 15(6):559–583, 2007.

Zaiwen Wen and Wotao Yin. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1):397–434, 2013.

Zaiwen Wen, Wotao Yin, Donald Goldfarb, and Yin Zhang. A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation. *SIAM Journal on Scientific Computing*, 32(4):1832–1857, 2010.

Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on signal processing*, 57(7):2479–2493, 2009.

Ya-xiang Yuan. A new stepsize for the steepest descent method. *Journal of Computational Mathematics*, pages 149–156, 2006.

Hongchao Zhang and William W Hager. A nonmonotone line search technique and its application to unconstrained optimization. *SIAM journal on Optimization*, 14(4):1043–1056, 2004.

Danqing Zhou, Shiqian Ma, and Junfeng Yang. Adabb: Adaptive barzilai-borwein method for convex optimization. *Mathematics of Operations Research*, 51(1):715–745, 2026.

A Computer-Assisted Certificate

The computer-assisted component of the proof is implemented by the following Python script. It uses only Python's standard `decimal` module.

```
from decimal import Decimal, getcontext, localcontext, ROUND_FLOOR, ROUND_CEILING

getcontext().prec = 120
D = Decimal

def opdown(fn):
    with localcontext() as c:
        c.prec=120; c.rounding=ROUND_FLOOR
        return +fn()
def opup(fn):
    with localcontext() as c:
        c.prec=120; c.rounding=ROUND_CEILING
        return +fn()

class IV:
    __slots__=('lo','hi')
    def __init__(self,lo,hi=None):
        self.lo=D(str(lo)) if not isinstance(lo,D) else lo
        self.hi=self.lo if hi is None else (D(str(hi)) if not isinstance(hi,D) else hi)
        assert self.lo <= self.hi
    def __add__(a,b):
        b=asiv(b); return IV(opdown(lambda:a.lo+b.lo),opup(lambda:a.hi+b.hi))
    __radd__=__add__
    def __neg__(a): return IV(-a.hi,-a.lo)
    def __sub__(a,b): return a+(-asiv(b))
    def __rsub__(a,b): return asiv(b)+(-a)
    def __mul__(a,b):
        b=asiv(b)
        los=[opdown(lambda x=x,y=y:x*y) for x in (a.lo,a.hi) for y in (b.lo,b.hi)]
        his=[opup(lambda x=x,y=y:x*y) for x in (a.lo,a.hi) for y in (b.lo,b.hi)]
        return IV(min(los),max(his))
    __rmul__=__mul__
    def recip(a):
        assert not (a.lo <= 0 <= a.hi)
        valslo=[opdown(lambda x=x:D(1)/x) for x in (a.lo,a.hi)]
        valshi=[opup(lambda x=x:D(1)/x) for x in (a.lo,a.hi)]
        return IV(min(valslo),max(valshi))
    def __truediv__(a,b): return a*asiv(b).recip()
    def __rtruediv__(a,b): return asiv(b)*a.recip()
    def __repr__(a): return f'[{a.lo},{a.hi}]'

def asiv(x): return x if isinstance(x,IV) else IV(x)
def sumup(vals):
    s=D(0)
    for v in vals: s=opup(lambda s=s,v=v:s+v)
    return s

class AD:
    __slots__=('v','d')
    def __init__(self,v,d=None):
        self.v=asiv(v); self.d=[IV(0) for _ in range(8)] if d is None else d
    def __add__(a,b):
        b=asad(b); return AD(a.v+b.v,[x+y for x,y in zip(a.d,b.d)])
    __radd__=__add__
    def __neg__(a): return AD(-a.v,[-x for x in a.d])
    def __sub__(a,b): return a+(-asad(b))
    def __rsub__(a,b): return asad(b)+(-a)
    def __mul__(a,b):
        b=asad(b); return AD(a.v*b.v,[x*b.v+a.v*y for x,y in zip(a.d,b.d)])
    __rmul__=__mul__
    def recip(a): return AD(a.v.recip(),[-x/(a.v*a.v) for x in a.d])
    def __truediv__(a,b): return a*asad(b).recip()
    def __rtruediv__(a,b): return asad(b)*a.recip()

```

```
def asad(x): return x if isinstance(x,AD) else AD(x)
def var(box,j):
    der=[IV(0) for _ in range(8)];der[j]=IV(1);return AD(box,der)

L2=AD('1.8786699041860466')
ONE=AD(1); FOUR=AD(4)

def fmap_with_den(z,la):
    m=z[0]; p=[z[1],z[2],z[3],ONE-z[1]-z[2]-z[3]]
    terms=[p[i]*(m-la[i])*(m-la[i]) for i in range(4)]
    den=sum(terms,AD(0))
    mean=sum((la[i]*p[i] for i in range(4)),AD(0))
    return [mean]+[terms[i]/den for i in range(3)],den

def fmap(z,la):
    out,_=fmap_with_den(z,la)
    return out

def fpow(z,la,n):
    for _ in range(n): z=fmap(z,la)
    return z

def f_and_j(boxes):
    x=[var(boxes[i],i) for i in range(8)]
    la=[ONE,L2,x[0],FOUR]
    z=x[1:5]
    a=fpow(z,la,7)
    p=x[5:8]; p4=ONE-sum(p,AD(0))
    m0=sum((la[i]*([p[0],p[1],p[2],p4][i]) for i in range(4)),AD(0))
    b=fpow([m0]+p,la,15)
    out=[a[i]-z[i] for i in range(4)]+[b[i]-z[i] for i in range(4)]
    return [o.v for o in out],[o.d for o in out]

C=[D(s) for s in [
'2.7127814910653372282743021917166259612213120097331194358155052455762957581 1',
'1.3334496725238812991898003496838493380028119662647089762866386716511125365 6',
'0.8281085983625578302847062799336426754477549077553810635365940227268902274 54',
'0.1670430513702266313898094627595023197810382863138141770388781875255927909 29',
'0.0047179573503674129528125339395019941688067981035232679520880506589411846 44',
'0.0000134113750442893933116720965327494222816004773939806475307653666721805974245',
'0.9891717057690781490472697115151355358471282234279160684741637435269262337 31',
'0.0000034926002843914558913346790346936487552351777485127115413021962990464337 9924']]

RSTR='''
-0.810526348848592888901084652460863262098889862944445915997922,307.242673381334385271539228546921025279578294613703728839309,307.741025629901050999029082768910651032865241495000968766894,312.782742571536891626096323271366901386778871589158283516632,-0.0137368376192333684238175487126343869576306541180311710900804,-421.401532972233614339176178157325006069741606160158648367929,-420.720628332827512110952534958393390951741568565011272813875,-417.510650350147764867647073770427680507322497423363849322798
-2.533812407435946415505136654312482193250919599564062402797 72,219.146844709590120107670663557533796892315501194180693740921,221.314903815662600782780829104204280970210726156970789478092,238.4330611388820966065109075544046309702814650794113064747 34,-0.00661561375369757644998764130282880318501507149754201226640 02,-202.945528995481851107027954492147980696679854674457223124 06,-202.617607662910774712346795218836979798249918528977551841755,-201.071693306206914696638563015692889394595107198327734687951
1.21923182241183846078543106866408237572050290347089038684799,-283.490269055458601705948958623004565413293682800367378712469,-284.215637538925680074135617356259178167657374408793138161663,-294.814697516715266368336657507713365927445251871977984552338,0.00510430037289669853951692869951609336796693289217752688876014,156.583346292001534198852820229253143092303058113580534397193,156.330337418986871290621900890772993639761529519065013423077,155.137581686689234202670907269428105368197596348629742204233
-1.26108044272173568710731141145266481650790341533600008992776,313.942924362985217723054543998906451668776939851163166266309,314.691544869619096433878614662418685401023724831284472124436,326.12764070745284415061672026001723134524840696137617430052,-0.00557213932631546351238737770698205496104361101774514176877326,-170.935124890502247958801827793221206576836705555576496966117,-170.658926275954081329267222464116649321236505862472492490618,-169.356847511562687722423100405437534964897323010662551564416
0.0430781956042790019140979633277378904016462354528942551105 03,-31.2979228333921281928882113925871982705271157312435670381131,-31.3219714710519066420310359824923010174212504037635542365452,-32.1662172009335501562092089499137302982952674704895954239567,0.000478099922934842210827835744935335196661930745438524440498953,14.6665517947567543301073202404399084503649597758063756119065,14.6428534396731771022785373940457640590873359509893549270 27,14.531132658756835739832753728374049724120361151869572985717
```

```
0.000226520043874099667313273699369756542509219592106852411297113,-0.0225135215590217304102381601094316680754744631503520593497794,-0.0226850497610954124630365608985981365370252310505830687908151,-0.0252253072713914488396223981850409872844821931535349461685 19,-0.000064641629038638976572344598739863152114839699637262146 1687481,-0.0407002504446717796356496112619123261021708978107613923529244,-0.040803751508573006261031724880533814595619281910822129727265 1,-0.0395655532419233021327540805034448469654922061236616605213023
-0.0591009778675161000065292747250136710540662107925102493119043,4.525428637470154004681198684575136861825405109680853337672 12,4.5708376271796282978337702766393546845959286011088156970192,5.252779805830252696321420814517451148234860848503853243936 74,0.0182267787018476751330387775829333443438138046050174382552172,13.7294182492651964487088572696615033597888320575657007499638,13.754769271794975636070159437239004779426662361833920781069 3,13.3953732522325690444026485860547372641458783131975069959 7
0.0000756712444366593089630182277257919839520913409911366332434105,-0.0081989760042423075326402649186583787661810381054243883229077 5,-0.0082547513572781010519963902420167902104881633655288816704791 3,-0.0091573580862574489622738016909791090939734170343188184899298 6,-0.0000197290639090762745552792690106575683366414360898273244501419,-0.0159791300873645197383620839265853550901431071050090962014869,-0.0160185329798007737064568268873709203041685752673075101630795,-0.0155338627526389510516621091918715132796619616194482426763043'''
R=[[IV(s) for s in line.split(',')] for line in RSTR.strip().splitlines()]

PSTR='''
977656.6634223271235507758960608141,227877753.5872023336353569232362594,226076347.8588968549246065233368533 3,213421596.3597830638102310653398526 2
227877753.5872023336353569232362594,53180721626.9996721233616937965385 79,52760940596.8204241335778931413892 99,49812202173.8342030467813803524301 7
226076347.8588968549246065233368533 3,52760940596.8204241335778931413892 99,52344478955.7791923121537767899753 03,49419059318.1086758667066629570640 5
213421596.3597830638102310653398526 2,49812202173.8342030467813803524301 7,49419059318.1086758667066629570640 5,46657455589.3577296571128749691689 3'''
P=[[IV(s) for s in line.split(',')] for line in PSTR.strip().splitlines()]

def mm(A,B):
    return [[sum((A[i][k]*B[k][j] for k in range(len(B))),IV(0)) for j in range(len(B[0]))] for i in range(len(A))]
def mv(A,v): return [sum((A[i][j]*v[j] for j in range(len(v))),IV(0)) for i in range(len(A))]
def tr(A): return [list(x) for x in zip(*A)]

def ldl_pivots(A):
    n=len(A); L=[[IV(0) for _ in range(n)] for _ in range(n)]; piv=[]
    for i in range(n):
        L[i][i]=IV(1)
        di=A[i][i]-sum((L[i][k]*L[i][k]*piv[k] for k in range(i)),IV(0))
        piv.append(di)
        for j in range(i+1,n):
            num=A[j][i]-sum((L[j][k]*L[i][k]*piv[k] for k in range(i)),IV(0))
            L[j][i]=num/di
    return piv

if __name__=='__main__':
    rad=D('1e-55')
    boxes=[IV(c-rad,c+rad) for c in C]
    _,J=f_and_j(boxes)
    fx,_=f_and_j([IV(c) for c in C])
    rf=mv(R,fx)
    RJ=mm(R,J)
    B=[[IV(1 if i==j else 0)-RJ[i][j] for j in range(8)] for i in range(8)]
    dx=[IV(-rad,rad) for _ in range(8)]
    bd=mv(B,dx)
    kc=[IV(C[i])-rf[i]+bd[i] for i in range(8)]
    print('Krawczyk ratios |K-c|/radius:')
    for i in range(8):
        assert C[i]-rad < kc[i].lo and kc[i].hi < C[i]+rad
        ratio=max(abs(kc[i].lo-C[i]),abs(kc[i].hi-C[i]))/rad
        print(i,ratio)
    bnorm=max(sumup(max(abs(v.lo),abs(v.hi)) for v in row) for row in B)
    assert bnorm < 1
    print('contraction infinity row-sum bound:',bnorm)
    # Period derivative A = derivative of first four outputs wrt z plus I.
    A=[[J[i][1+j]+(1 if i==j else 0) for j in range(4)] for i in range(4)]
    print('A intervals widths / centers:')
    for row in A: print(' '.join(str((v.lo+v.hi)/2) for v in row))
```

```
print('P LDL lower pivots:')
ppiv=ldl_pivots(P); assert all(x.lo>0 for x in ppiv)
print(*(x.lo for x in ppiv))
Q=[[P[i][j]-mm(mm(tr(A),P),A)[i][j] for j in range(4)] for i in range(4)]
print('Q diagonal-dominance lower margins:')
for i in range(4):
    off=sumup(max(abs(Q[i][j].lo),abs(Q[i][j].hi)) for j in range(4) if j!=i)
    margin=opdown(lambda:Q[i][i].lo-off); assert margin>0
    print(margin)
# Uniform transverse Floquet multiplier for any inserted eigenvalue nu in [3.49,3.51].
xv=[var(boxes[i],i) for i in range(8)]
la=[ONE,L2,xv[0],FOUR]; zz=xv[1:5]; nu=AD(IV('3.49','3.51')); tau=AD(1)

cycle_state=xv[1:5]; cycle_z_min=D('Infinity')
for _ in range(7):
    cycle_state,den=fmap_with_den(cycle_state,la)
    cycle_z_min=min(cycle_z_min,den.v.lo)
p=xv[5:8]; p4=ONE-sum(p,AD(0))
conn_state=[sum((la[i]*[p[0],p[1],p[2],p4][i] for i in range(4)),AD(0))]+p
conn_z_min=D('Infinity')
for _ in range(15):
    conn_state,den=fmap_with_den(conn_state,la)
    conn_z_min=min(conn_z_min,den.v.lo)
print('minimum cycle normalization denominator:',cycle_z_min)
print('minimum connection normalization denominator:',conn_z_min)
assert cycle_z_min>D('0.1516357118221540')
assert conn_z_min>D('0.0009510722601452239')

cycle_next,_=fmap_with_den(xv[1:5],la)
phase_gap=(xv[1]-cycle_next[0]).v
print('first cycle phase difference a0-a1:',phase_gap)
assert phase_gap.lo>D('0.1782019618055700634')

for _ in range(7):
    m=zz[0]; pp=[zz[1],zz[2],zz[3],ONE-zz[1]-zz[2]-zz[3]]
    den=sum((pp[i]*(m-la[i])*(m-la[i]) for i in range(4)),AD(0))
    tau=tau*(m-nu)*(m-nu)/den
    zz=fmap(zz,la)
print('transverse tau [3.49,3.51]:',tau.v)
assert D(0)<tau.v.lo and tau.v.hi<1
# No multiplier vanishes during the exceptional initial segment or on the cycle.
def sep(a,b):
    a=a.v if isinstance(a,AD) else asiv(a); b=b.v if isinstance(b,AD) else asiv(b)
    if a.hi < b.lo: return opdown(lambda:b.lo-a.hi)
    if b.hi < a.lo: return opdown(lambda:a.lo-b.hi)
    return D(0)
xx=[sum((la[i]*[p[0],p[1],p[2],p4][i] for i in range(4)),AD(0))]+p
transient_sep=D('Infinity')
for _ in range(16):
    transient_sep=min(transient_sep,*(sep(xx[0],ell) for ell in la),sep(xx[0],nu))
    xx=fmap(xx,la)
cycle_sep=D('Infinity'); xx=xv[1:5]
for _ in range(7):
    cycle_sep=min(cycle_sep,*(sep(xx[0],ell) for ell in la),sep(xx[0],nu))
    xx=fmap(xx,la)
print('minimum initial-segment m-lambda separation:',transient_sep)
print('minimum cycle separation (including nu box):',cycle_sep)
assert transient_sep>0 and cycle_sep>0
```

B Prompt Used for Proof Verification

The following prompt was submitted to GPT 5.6 Sol.

```
Current task statement
```

```
For $n\geq4$, given $n$ positive real numbers $\lambda_1\leq ... \leq \lambda_n$. Consider the sequence $\{d^1_k, ..., d^n_k\} (k\geq 0)$ where $d^i_k$ are real numbers for $1\leq i\leq n$, satisfying

$$\begin{equation}
d^i_{k+1}=d^i_k\cdot\left(
\frac{
\sum_{j=1}^n(\lambda_j-\lambda_i)(d^j_{k-1})^2
} {
\sum_{j=1}^n\lambda_j(d^j_{k-1})^2
}
\right), ~~i=1, \cdots, n.
\end{equation}$$
Specifically, in the first iteration, the relation becomes
$$\begin{equation}
d^i_1=d^i_0\cdot\left(\frac{\sum_{j=1}^n(\lambda_j-\lambda_i)(d^j_0)^2} {\sum_{j=1}^n\lambda_j(d^j_0)^2}\right), ~~i=1, \cdots, n.
\end{equation}$$

Now, given $2n$ real numbers $\lambda_1, ... \lambda_n, d_0^1, ..., d_0^n$, characterize the decreasing rate of the sequence $|d_k^i|$. In particular, prove or disprove the following:

For a positive measure of $(\lambda_1, ..., \lambda_n)\in\mathbb{R}^n$ and a positive measure of $(d_0^1, ..., d_0^n)\in\mathbb{R}^n$, $|d_k^i|$ converges at a linear rate, i.e., there exists positive constants $C_i$ and $0<\rho_i<1$ only related to the $2n$ initial real numbers, such that $|d_k^i|\geq C_i\rho_i^k$.

Assume for purposes of this task that a complete affirmative proof or counterexample exists. Partial progress does not count unless it implies exactly the resolution above. In particular, proofs for zero measure of $(\lambda_1, ..., \lambda_n)\in\mathbb{R}^n$ or zero measure of $(d_0^1, ..., d_0^n)\in\mathbb{R}^n$, $|d_k^i|$, constructions of counter-examples with zero measure of $(\lambda_1, ..., \lambda_n)\in\mathbb{R}^n$ or zero measure of $(d_0^1, ..., d_0^n)\in\mathbb{R}^n$, $|d_k^i|$, reductions to another
unproved conjecture, computational verification through any fixed $n$, and candidate
counterexamples without a complete certificate are insufficient.

Use multiagent v2 aggressively and dynamically. You have up to 64 concurrent agents available. Do not use a fixed assignment such as "N agents for strategy X." Instead, manage the search using the following heuristics:

- Begin with a genuinely diverse portfolio of approaches. Agents should explore substantially different formulations, invariants, reductions, algebraic viewpoints, structural inductions, decompositions, flow formulations, transition systems, embeddings, extremal arguments, and computational sanity checks.

- Do not tell most agents the currently favored approach. Preserve independence during early rounds so that agents do not all converge to the same attractive but incomplete reduction.

- Maintain an explicit registry of approach families. Group agents by the mathematical idea they are using, not by superficial wording. If many agents converge to one family, redirect some of them toward underexplored formulations.

- Do not allow one approach to dominate merely because it gives elegant reductions. A route that ends at a lemma equivalent in strength to the original conjecture is not close to completion unless it supplies a genuinely new proof of that lemma.

- When an approach stalls at a theorem-strength missing lemma, mark that route as blocked. Only continue assigning agents to it if someone proposes a materially new mechanism, invariant, or construction.
```

- Keep several incompatible proof routes alive through multiple rounds. Cross-pollinate ideas only after independent agents have developed them far enough to expose their real strengths and gaps.

- Require agents to return concrete lemmas, constructions, equations, or counterexamples to proposed sublemmas. Reject status reports, vague optimism, and claims that an unproved global compatibility statement is "routine."

- The root agent should repeatedly synthesize, challenge, redirect, and launch new rounds. Do not stop after the first wave fails. Produce a complete proof if one survives audit; otherwise report only the strongest rigorously proved derivation and its exact remaining gap.

Do not return merely because current approaches fail or agents report theorem-strength gaps.

Continue launching new rounds, reopening blocked approaches only when there is a genuinely new mechanism, and searching for fresh formulations.

Return only when a complete affirmative proof has been found and survives adversarial audit. Do not return a reduction, partial result, isolated missing lemma, "best effort" summary, or explanation of why the problem is difficult.

Spend at least 8 hours on this before even thinking of returning or giving up.

Public search may be used only for ordinary mathematical background or standard named theorems, not to search for a solution to this exact conjecture or benchmark. Do not search the public web merely to determine whether this problem is open, and do not answer that it is open.