GEOMETRY DETERMINES LOCAL ACCELERATION IN GRIFFIN–LIM PHASE RETRIEVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Acceleration in Griffin–Lim phase retrieval is commonly tuned heuristically, despite substantial instance-to-instance variation in local geometry. We show that this geometry determines the local acceleration problem. At a regular feasible solution, the transverse Griffin–Lim Jacobian is self-adjoint positive semidefinite, with eigenvalues equal to the squared cosines of the principal angles between the signal-consistency subspace and the tangent space of the magnitude manifold. This geometry–spectrum correspondence turns local acceleration into an explicit spectral design problem. For the three-parameter AGLA family, we derive the exact modal dynamics of the linearization and solve the resulting minimax spectral-radius problem in closed form. The geometry-optimal design achieves a strictly smaller optimal local factor than FGLA, and hence GLA, on every nondegenerate transverse spectral interval.

When the local spectrum is unknown, however, acceleration becomes an information problem. We show that passive trajectory observations cannot identify spectral directions they do not excite. Under explicit local active-subspace and consistency conditions, trajectory-based Ritz estimates nevertheless recover the excited dominant spectral edge, yielding an online method that conditionally attains the corresponding conservative geometry-designed asymptotic rate. Real-audio STFT experiments support the predicted geometry–spectrum relationship and reveal a complementary robustness–rate separation: online adaptation does not improve paired finite-window local contraction rates on average, but substantially improves empirical convergence success under the frozen initialization protocol. Across three real-audio evaluations, Online attains 73.30–74.94% success, versus 54.87–55.04% for oracle-tuned FGLA. On a disjoint 4,752-run MUSAN evaluation, the same frozen method attains 73.30%, compared with 52.06% for a literature-configured NAGLA and 47.62% for a published AGLA configuration; a larger 11,856-run MUSAN evaluation reproduces the robustness–rate separation. Together, these results identify local geometry as the organizing principle for Griffin–Lim acceleration, characterize the spectral information available to passive adaptation, and separate geometry-optimal local acceleration from what can be realized online.

1 INTRODUCTION

Phase retrieval from magnitude measurements is a canonical nonconvex inverse problem. For STFT reconstruction, Griffin–Lim alternates between signal consistency and prescribed magnitude (Griffin and Lim, 1984). Its accelerated variants modify this basic iteration through inertia: FGLA uses fixed momentum (Perraudin et al., 2013), AGLA introduces a three-parameter family with global convergence guarantees in a sufficient parameter region (Nenov et al., 2024), and recent analyses establish local guarantees for FGLA and NAGLA (Mezza et al., 2026; Cicognani et al., 2026). These results clarify when accelerated Griffin–Lim iterations converge, but leave open a different question: how much acceleration is permitted by the geometry of a specific phase-retrieval instance? We address this question within the AGLA family: *given the local geometry of a regular solution, what is the best achievable local contraction, and which inertial parameters attain it?*

Our starting point is that this is fundamentally a spectral question. Near a regular feasible solution, the Griffin–Lim map linearizes to a product of tangent-space projectors. After quotienting out the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

unavoidable global-phase direction, the transverse Jacobian is self-adjoint positive semidefinite, and its eigenvalues are precisely the squared cosines of the principal angles between the signal-consistency subspace and the tangent space of the magnitude manifold. The same spectrum simultaneously determines local curvature and the modal dynamics of inertial Griffin–Lim iterations. It therefore provides an instance-dependent description of both the difficulty of the local problem and the acceleration that its geometry permits. This leads to a different design principle from stability-oriented parameter selection: we optimize the inertial iteration directly against the geometry-induced spectral interval, derive its exact minimax contraction factor, and then determine how much of the required spectral information can be recovered from the trajectory when the geometry is unavailable.

Contributions. Our results form a geometry–information–acceleration chain. We first identify the complete transverse Griffin–Lim Jacobian spectrum with squared principal-angle cosines, linking local geometry directly to curvature and conditioning. Using this spectral representation, we derive the exact modal recurrence of the three-parameter AGLA linearization and solve its spectral-radius minimax problem over $[\lambda_-, \lambda_+]$ in closed form, obtaining a geometry-optimal local factor that is strictly smaller than that of optimally tuned FGLA, and hence GLA, on every nondegenerate transverse spectral interval. When the spectrum is unknown, we show that passive observations cannot identify unexcited spectral directions, while under explicit local excitation, residual-consistency, and conditioning assumptions, trajectory-active Ritz estimates consistently recover the excited spectrum and its dominant edge when represented. Finally, with consistent dominant-edge estimation and entry into the regular local regime, the resulting nonautonomous method converges orbitally and conditionally recovers the corresponding conservative geometry-designed asymptotic root rate.

Real-audio STFT experiments support this picture. Independently measured geometry predicts the local spectrum and acceleration, while online adaptation primarily improves convergence robustness rather than paired finite-window local contraction rates. This robustness–rate separation reproduces across three real-audio evaluations, including a 504-file MUSAN study.

Related work. Griffin–Lim belongs to the projection-based phase-retrieval lineage of Gerchberg–Saxton and Fienup (Gerchberg and Saxton, 1972; Fienup, 1982; Griffin and Lim, 1984). FGLA, AGLA, and NAGLA develop increasingly general inertial accelerations, with complementary global or local convergence guarantees (Perraudin et al., 2013; Nenov et al., 2024; Mezza et al., 2026; Cicognani et al., 2026); broader phase-retrieval theory studies alternating minimization, projection methods, identifiability, and stability (Netrapalli et al., 2015; Marchesini et al., 2016; Luke and Martins, 2020; Thao et al., 2020; 2021; Grohs and Liehr, 2022; Bartusel, 2023; Alaifari and Yang, 2025). In this paper, our focus is instance-level local design: we expose the spectrum governing a regular Griffin–Lim fixed point and optimize the inertial dynamics against it.

The geometric foundation is the local theory of alternating projections between smooth manifolds, where convergence is governed by intersection geometry and principal angles (Lewis and Malick, 2008; Lewis et al., 2009). Here, this geometry identifies the complete transverse Griffin–Lim spectrum, reducing AGLA design to scalar second-order modal recurrences with a closed-form minimax solution. This differs from general inertial analyses that establish convergence for prescribed parameters through descent, KL geometry, or safeguards (Ochs et al., 2014; Li et al., 2017).

Finally, our adaptive analysis connects to Rayleigh–Ritz and subspace spectral approximation (Saad, 2011; Nakatsukasa and Tropp, 2024; Haas and Nakatsukasa, 2026), but with trial spaces generated by the optimization trajectory itself. Consequently, passive secants reveal only trajectory-active spectral information, not arbitrary unseen directions. Under explicit local conditions, Ritz estimates recover the active dominant edge needed by our conservative online design, yielding conditional recovery of its geometry-designed asymptotic rate.

2 Problem setup and regularity

Let $T : \mathbb{C}^L \to \mathbb{C}^M$ have full column rank and let $s \in \mathbb{R}^M_+$. We view $\mathbb{C}^M$ as a real Hilbert space with inner product

$$\langle x, y\rangle_{\mathbb{R}} = \Re\langle x, y\rangle_{\mathbb{C}},$$

and define the signal-consistency subspace and magnitude manifold by

$$\mathcal{U} = \mathrm{range}(T), \qquad \mathcal{M}_s = \{c \in \mathbb{C}^M : |c_i| = s_i\}.$$

Since T has full column rank, the orthogonal projector onto $\mathcal{U}$ is $P_{\mathcal{U}} = TT^{\dagger}$. The Griffin–Lim map and the phase-retrieval objective restricted to $\mathcal{U}$ are

$$G = P_{\mathcal{U}} P_{\mathcal{M}_s}, \qquad f(c) = \frac{1}{2} d^2_{\mathcal{M}_s}(c),$$

so GLA is simply $c_{n+1} = G(c_n)$. We study the local dynamics near a feasible point $c^\star \in \mathcal{U} \cap \mathcal{M}_s$, modulo the unavoidable global-phase orbit

$$\mathcal{O}^\star = \{e^{i\phi} c^\star : \phi \in \mathbb{R}\}.$$

Assumption 2.1 (Regular feasible solution)**.** Assume $s_i > 0$ for every i. Near $c^\star$, suppose that the feasible intersection consists only of the global-phase orbit $\mathcal{O}^\star$, and that

$$\mathcal{U} \cap T_{c^\star} \mathcal{M}_s = \operatorname{span}_{\mathbb{R}}\{ic^\star\}.$$

The first condition gives local isolation modulo phase, while the second excludes transverse tangencies. These are separate hypotheses: the tangent condition alone does not imply local isolation.

The positivity of s makes $\mathcal{M}_s$ a smooth product of circles near $c^\star$ and $P_{\mathcal{M}_s}$ smooth there. Indeed, for one coordinate,

$$p_s(z) = s\frac{z}{|z|}$$

has derivative, at $|z| = s > 0$,

$$Dp_s(z)[h] = h - u\Re(\bar{u}h), \qquad u = \frac{z}{|z|},$$

which is precisely the real-orthogonal projector onto the tangent line of the magnitude circle at z. Hence, writing

$$V = T_{c^\star} \mathcal{M}_s,$$

we obtain

$$DP_{\mathcal{M}_s}(c^\star) = P_V.$$

Since $P_{\mathcal{U}}$ is linear, the Griffin–Lim Jacobian at the feasible point is therefore

$$DG(c^\star) = P_{\mathcal{U}} P_V.$$

Thus the local dynamics are governed directly by the relative geometry of the two tangent spaces $\mathcal{U}$ and V, which is the starting point of the spectral analysis below.

3 Geometry determines the local spectrum

Let $V = T_{c^\star} \mathcal{M}_s$ and define the transverse slice

$$\mathcal{N}_\star = \{h \in \mathcal{U} : \langle h, ic^\star \rangle_{\mathbb{R}} = 0\}.$$

Since $DP_{\mathcal{M}_s}(c^\star) = P_V$, the local Griffin–Lim dynamics reduce to the geometry of the two subspaces $\mathcal{U}$ and V.

Theorem 3.1 (Geometry–spectrum correspondence)**.** *Under Assumption 2.1,*

$$DG(c^\star)|_{\mathcal{U}} = J_\star := P_{\mathcal{U}} P_V|_{\mathcal{U}}.$$

The operator $J_\star$ is self-adjoint positive semidefinite on $\mathcal{U}$, and its nonzero eigenvalues are

$$\lambda_j = \cos^2 \theta_j,$$

where θ_j are the principal angles between $\mathcal{U}$ and V. Moreover,

$$\ker(I - J_\star) = \mathcal{U} \cap V = \operatorname{span}_{\mathbb{R}}\{ic^\star\}.$$

Hence, $\mathcal{N}_\star$ is $J_\star$-invariant and $\|J_\star|_{\mathcal{N}_\star}\| < 1$. When $\mathcal{N}_\star \neq \{0\}$, define

$$\lambda_- := \lambda_{\min}(J_\star|_{\mathcal{N}_\star}), \qquad \lambda_+ := \lambda_{\max}(J_\star|_{\mathcal{N}_\star}),$$

so that $0 \le \lambda_- \le \lambda_+ < 1$.

For $x, y \in \mathcal{U}$,

$$\langle x, J_\star y\rangle_{\mathbb{R}} = \langle P_V x, P_V y\rangle_{\mathbb{R}}, \qquad \langle x, J_\star x\rangle_{\mathbb{R}} = \|P_V x\|^2,$$

giving self-adjointness and positivity; the remaining claims follow from principal-angle geometry and Assumption 2.1. Full proofs are in Appendix A.

Corollary 3.2 (Transverse curvature)**.** *For $f(c) = \frac{1}{2} d^2_{\mathcal{M}_s}(c)$ restricted to $\mathcal{U}$,*

$$\nabla(f|_{\mathcal{U}})(c) = c - G(c), \qquad \nabla^2(f|_{\mathcal{U}})(c^\star) = I - J_\star.$$

Thus, the extreme transverse Hessian eigenvalues are $1 - \lambda_+$ and $1 - \lambda_-$, with condition number

$$\kappa_{\mathrm{geom}} = \frac{1 - \lambda_-}{1 - \lambda_+}.$$

Moreover, for every $\varepsilon \in (0, 1 - \lambda_+)$, sufficiently near $\mathcal{O}^\star$,

$$\frac{1 - \lambda_+ - \varepsilon}{2} \operatorname{dist}(c, \mathcal{O}^\star)^2 \le f(c) - f(c^\star) \le \frac{1 - \lambda_- + \varepsilon}{2} \operatorname{dist}(c, \mathcal{O}^\star)^2. \tag{3.1}$$

A principal angle θ therefore determines both the GLA eigenvalue $\lambda = \cos^2\theta$ and the transverse curvature $1 - \lambda = \sin^2\theta$. Small angles produce weak curvature and slow GLA contraction, while a narrow cluster of angles produces a narrow spectral interval $[\lambda_-, \lambda_+]$, potentially allowing much stronger minimax acceleration even when λ_+ is close to one. After removing the sole neutral phase mode, $1 - \lambda_+ > 0$ is the transverse curvature margin. Thus, local acceleration is governed by the instance-dependent interval $[\lambda_-, \lambda_+]$, which we optimize against next.

4 From modal dynamics to exact minimax acceleration

Assume the nondegenerate transverse regime $0 \le \lambda_- < \lambda_+ < 1$. it is worth noting that AGLA iterates

$$\begin{aligned} t_n &= (1-\gamma) d_{n-1} + \gamma G(c_{n-1}), \\ c_n &= t_n + \alpha(t_n - t_{n-1}), \qquad d_n = t_n + \beta(t_n - t_{n-1}). \end{aligned} \tag{4.1}$$

For a transverse eigenmode $J_\star v = \lambda v$, its linearization obeys

$$z_n = A_\lambda z_{n-1} - C_\lambda z_{n-2}, \tag{4.2}$$

where

$$A_\lambda = (1-\gamma)(1+\beta) + \gamma\lambda(1+\alpha), \qquad C_\lambda = (1-\gamma)\beta + \gamma\lambda\alpha. \tag{4.3}$$

It turns out that $p_\lambda(r) = r^2 - A_\lambda r + C_\lambda$, with the companion matrix

$$M_\lambda = \begin{pmatrix} A_\lambda & -C_\lambda \\ 1 & 0 \end{pmatrix}.$$

At the phase mode,

$$p_1(r) = (r-1)(r - C_1), \qquad C_1 = (1-\gamma)\beta + \gamma\alpha,$$

so the unit root represents global-phase symmetry. Orbital stability therefore requires Schur stability of every transverse M_λ together with $|C_1| < 1$.

The local robust-design problem is

$$R(\alpha, \beta, \gamma; \lambda_-, \lambda_+) = \sup_{\lambda \in [\lambda_-, \lambda_+]} \rho(M_\lambda), \qquad \alpha, \beta \ge 0, \quad \gamma > 0. \tag{4.4}$$

For a target radius $q \in (0, 1)$, the exact radius-q Jury conditions are

$$|C_\lambda| \le q^2, \qquad q^2 - qA_\lambda + C_\lambda \ge 0, \qquad q^2 + qA_\lambda + C_\lambda \ge 0. \tag{4.5}$$

After splitting the absolute value, these inequalities are affine in λ. It implies that the feasibility over the full spectral interval is equivalent to feasibility at its two endpoints, reducing the continuum minimax problem to finite constraints.

Theorem 4.1 (Exact geometry-optimal local acceleration)**.** *Let*

$$D = \lambda_+ - \lambda_- > 0, \qquad B = 3\lambda_+ + \lambda_- - \lambda_+^2 - 3\lambda_+\lambda_-.$$

The exact minimax transverse root factor is

$$q^\star(\lambda_-, \lambda_+) = \frac{B - \sqrt{B^2 - 4\lambda_+ D^2}}{2D}. \tag{4.6}$$

It is attained by

$$\beta^\star = 0, \qquad \gamma^\star = \frac{(1-q^\star)^2}{1-\lambda_+}, \qquad \alpha^\star = \frac{q^{\star 2}(1-\lambda_+)}{\lambda_+(1-q^\star)^2}. \tag{4.7}$$

At the endpoints,

$$p_{\lambda_+}(r) = (r-q^\star)^2, \qquad p_{\lambda_-}(r) = (r+q^\star)\left(r + \frac{\lambda_-}{\lambda_+}q^\star\right).$$

Moreover,

$$C_1^\star = \frac{q^{\star 2}}{\lambda_+} < q^\star < 1,$$

so the secondary phase root is orbitally stable and does not enlarge the optimized transverse factor. Hence, after quotienting global phase, $q^\star(\lambda_-, \lambda_+)$ is also the asymptotic root factor of the full linearized two-step state at the optimizer.

The minimax lower bound follows directly from Equation (4.5). Writing $s_\lambda = A_\lambda - C_\lambda = 1 - \gamma(1-\lambda)$ shows that $\gamma \le 1$ implies

$$q \ge 1 - \sqrt{1-\lambda_+},$$

so any improvement over optimal FGLA requires over-relaxation $\gamma > 1$. In this regime, the endpoint constraints imply

$$Dq^2 - Bq + \lambda_+ D \le 0,$$

whose smaller root is exactly Equation (4.6). The parameters in Equation (4.7) attain this bound through the displayed endpoint factorizations; endpoint reduction then certifies every interior mode. Thus, the closed form matches a necessary lower bound rather than merely providing a tuned sufficient rate. Full derivations are in Appendix C.

Corollary 4.2 (Optimal FGLA and strict hierarchy)**.** *Within the FGLA subfamily $\gamma = 1$, the exact minimax factor and optimal momentum are*

$$q_{\mathrm{F}}^\star = 1 - \sqrt{1-\lambda_+}, \qquad \alpha_{\mathrm{F}}^\star = \frac{1-\sqrt{1-\lambda_+}}{1+\sqrt{1-\lambda_+}}. \tag{4.8}$$

For every $0 \le \lambda_- < \lambda_+ < 1$,

$$q^\star(\lambda_-, \lambda_+) < q_{\mathrm{F}}^\star < \lambda_+. \tag{4.9}$$

Hence, geometry-optimal AGLA strictly improves the optimal local linearized minimax factor of FGLA, and FGLA strictly improves that of GLA.

Why the lower spectral edge matters. Unlike optimal FGLA, whose factor depends only on λ_+, geometry-optimal AGLA exploits the full interval $[\lambda_-, \lambda_+]$. The conservative one-edge design is

$$q^\star(0, \lambda_+) = \frac{3 - \lambda_+ - \sqrt{(1-\lambda_+)(9-\lambda_+)}}{2}, \tag{4.10}$$

while for fixed $\lambda_+ < 1$ and $D = \lambda_+ - \lambda_- \downarrow 0$,

$$q^\star(\lambda_-, \lambda_+) = \frac{D}{4(1-\lambda_+)} + O(D^2). \tag{4.11}$$

As a result, instances with the same dominant edge can admit very different optimal local acceleration: clustered transverse spectra can yield a much smaller minimax factor. The optimizer has $\beta^\star = 0$, showing that the second inertial channel cannot improve the worst-case asymptotic factor once the local spectral interval is fixed, although it may still affect nonlinear transients or basin behavior. These conclusions are local and pointwise for $\lambda_+ < 1$; they neither imply global convergence nor remain uniform as $\lambda_+ \uparrow 1$, where $\gamma^\star$ may become large.

5 WHAT CAN A PASSIVE TRAJECTORY LEARN?

The geometry-optimal design requires spectral information that an online method must infer from its trajectory. Define the secants

$$s_n = c_n - c_{n-1}, \qquad y_n = G(c_n) - G(c_{n-1}).$$

Locally,

$$y_n = J_\star s_n + r_n, \qquad \frac{\|r_n\|}{\|s_n\|} \to 0, \tag{5.1}$$

so the first question is which parts of the spectrum are identifiable from passive secant observations.

Proposition 5.1 (Finite passive non-identifiability)**.** *Let $\mathcal{K}_N = \operatorname{span}\{s_1, \dots, s_N\}$. If $\mathcal{K}_N$ is a proper subspace of the transverse space, then within the class of self-adjoint PSD contractions there exist distinct operators that agree on $\mathcal{K}_N$ but have different spectra on $\mathcal{K}_N^\perp$. Hence, the finite passive first-order observations do not determine the complete transverse spectrum without additional structure.*

The construction is immediate: for a unit $u \in \mathcal{K}_N^\perp$, extend the same PSD contraction on $\mathcal{K}_N$ with two distinct eigenvalues along u. The proposition is a statement about the local self-adjoint PSD-contraction model class; it does not assert that every such completion is realizable by a phase-retrieval projector product.

We therefore estimate the spectrum represented by the trajectory itself. For memory p, let

$$S_n = [s_{n-p+1}, \dots, s_n], \qquad Y_n = [y_{n-p+1}, \dots, y_n],$$

and define

$$H_n = \Re(S_n^* S_n), \qquad B_n = \tfrac{1}{2}\Re(S_n^* Y_n + Y_n^* S_n). \tag{5.2}$$

Theorem 5.2 (Trajectory-active Ritz consistency)**.** *Suppose a global-phase representative is fixed so that $c_n \to c^\star$, $H_n \succ 0$ eventually, and*

$$\varepsilon_n := \|(Y_n - J_\star S_n) H_n^{-1/2}\| \to 0. \tag{5.3}$$

Assume also that the secant subspaces $\mathcal{V}_n := \operatorname{range}(S_n)$ converge in projector norm to a p-dimensional $J_\star$-invariant subspace $\mathcal{V}$. Then, the ordered generalized Ritz values of $B_n v = \mu H_n v$ converge to the ordered eigenvalues of $J_\star|_{\mathcal{V}}$. In particular, if $\mathcal{V}$ contains a dominant transverse eigendirection, then

$$\widehat{\lambda}_{+,n} \to \lambda_+. \tag{5.4}$$

The normalization in Equation (5.3) is essential because the secants vanish near the solution. Indeed, with $Q_n = S_n H_n^{-1/2}$,

$$H_n^{-1/2} B_n H_n^{-1/2} = \Re(Q_n^* J_\star Q_n) + E_n, \qquad \|E_n\| \le \varepsilon_n \to 0.$$

Projector convergence then identifies the limiting compression with $J_\star|_{\mathcal{V}}$, and standard eigenvalue perturbation gives the claim. Thus, recovery of λ_+ requires its eigendirection to be represented in the limiting secant subspace.

The information boundary is therefore explicit: passive secants can recover trajectory-active spectral information, but cannot in general reveal unexcited directions. In particular, the lower edge may disappear rapidly from the observed trajectory, motivating the conservative online interval $[0, \widehat{\lambda}_{+,n}]$. The theorem does not provide persistent excitation; if H_n becomes ill-conditioned, the implementation rejects the spectral update and uses the safeguarded fallback. Full proofs and a sufficient normalized-secant conditioning condition are given in Appendix D.

6 ONLINE GEOMETRY-ADAPTIVE RATE RECOVERY

Let $\Theta(0, \lambda)$ denote the one-edge minimax design from Theorem 4.1. Given an estimate $\widehat{\lambda}_{+,n}$, Online uses the conservative interval $[0, \widehat{\lambda}_{+,n}]$, with

$$q_0(\lambda) = q^\star(0, \lambda) = \frac{3 - \lambda - \sqrt{(1-\lambda)(9-\lambda)}}{2}, \tag{6.1}$$

and

$$\Theta(0,\lambda) = (\alpha_0(\lambda), 0, \gamma_0(\lambda)), \qquad \gamma_0(\lambda) = \frac{(1-q_0)^2}{1-\lambda}, \qquad \alpha_0(\lambda) = \frac{q_0^2(1-\lambda)}{\lambda(1-q_0)^2}. \tag{6.2}$$

Continuity of Θ on $(0,1)$ implies that consistent edge estimates yield convergent inertial parameters.

Because these parameters vary with n, the linearized two-step dynamics are nonautonomous. In phase-aligned coordinates,

$$\mathbb{A}_n = \begin{pmatrix} \mathsf{A}_n & -\mathsf{C}_n \\ I & 0 \end{pmatrix},$$

where

$$\mathsf{A}_n = (1-\gamma_n)(1+\beta_{n-1})I + \gamma_n(1+\alpha_{n-1})J_\star, \qquad \mathsf{C}_n = (1-\gamma_n)\beta_{n-1}I + \gamma_n\alpha_{n-1}J_\star.$$

As a consequence, $\widehat{\lambda}_{+,n} \to \lambda_+$ implies $\mathbb{A}_n \to \mathbb{A}_\star$, the state operator synthesized from $[0, \lambda_+]$.

Let $X_n = (t_n, t_{n-1})$ and $\mathcal{X}^\star = \{(u,u) : u \in \mathcal{O}^\star\}$.

Theorem 6.1 (Conditional online asymptotic rate recovery)**.** *Assume $\widehat{\lambda}_{+,n} \to \lambda_+ \in (0,1)$ and that, sufficiently late, the realized two-step state enters a sufficiently small regular neighborhood of $\mathcal{X}^\star$. Then, the nonautonomous geometry-adaptive dynamics converge orbitally and*

$$\limsup_{n\to\infty} \operatorname{dist}(X_n, \mathcal{X}^\star)^{1/n} \le q^\star(0,\lambda_+) < 1 - \sqrt{1-\lambda_+}.$$

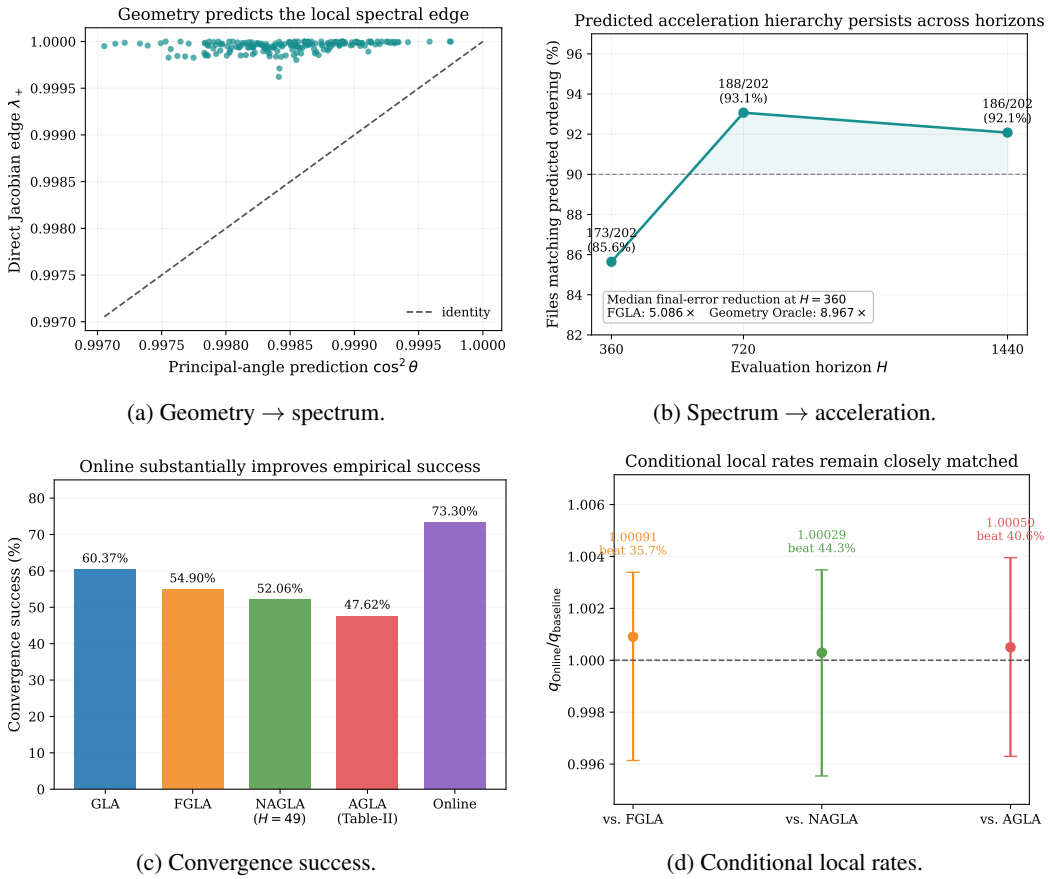
After quotienting the global-phase mode, the limiting state operator has spectral radius at most $q^\star(0,\lambda_+) < 1$. An adapted norm, parameter convergence, and C^1 smoothness then provide a local contraction margin for the nonautonomous dynamics and yield the stated root rate; see Appendix E.

The result is deliberately conditional. Theorem 5.2 gives spectral consistency along a convergent, sufficiently excited trajectory, whereas Theorem 6.1 assumes edge consistency and local entry to recover the designed rate. Together they characterize the estimator and optimizer components, but not unconditional convergence of their coupled dynamics. Because passive trajectories need not reveal λ_-, Online targets $[0, \widehat{\lambda}_+]$: its limiting factor $q^\star(0,\lambda_+)$ is generally slower than the two-edge oracle optimum $q^\star(\lambda_-,\lambda_+)$ but remains strictly better than the optimal FGLA factor. These statements are pointwise for $\lambda_+ < 1$; as $\lambda_+ \uparrow 1$, the transverse curvature gap closes and the local guarantees need not be uniform.

7 Experiments

We evaluate the theory on real-audio STFT phase retrieval through four complementary questions: whether principal-angle geometry predicts the local Griffin–Lim spectrum, whether this spectrum predicts the acceleration observed in local dynamics, whether trajectory-active spectral information can be exploited effectively online, and which components of the online mechanism are associated with the resulting robustness gain. All evaluation protocols are frozen across populations. Online uses Ritz memory $p = 3$, update period $K = 8$, zero additional margin, a running envelope, and GLA fallback, without retuning. Throughout, we distinguish finite-window fitted contraction factors and empirical convergence success from the asymptotic root rates characterized by the local theory. Full protocols, diagnostics, failure accounting, and replication results are provided in Appendix F.

Geometry predicts the spectrum, and the spectrum predicts acceleration. Principal-angle geometry accurately predicts the dominant transverse Jacobian edge: across the frozen 202-file population, the median and maximum relative discrepancies are 1.4385×10^{-3} and 2.8971×10^{-3}, respectively (Figure 1(a)). The resulting spectrum also predicts the observed acceleration hierarchy (Figure 1(b)): the GLA/FGLA/Geometry-Oracle ordering holds for 173/202 files at $H = 360$, 188/202 at $H = 720$, and 186/202 at $H = 1440$. At $H = 360$, FGLA and the Geometry Oracle reduce median final error relative to GLA by $5.086\times$ and $8.967\times$, respectively. Thus, the experiments support both links of the local theory: principal-angle geometry predicts the transverse spectrum, and the spectrum, in turn, predicts the acceleration observed in local dynamics.



(a) Geometry → spectrum.

(b) Spectrum → acceleration.

(c) Convergence success.

(d) Conditional local rates.

Figure 1: **Experimental evidence for the geometry–spectrum–acceleration picture.** **(a)** Principal-angle geometry predicts the dominant transverse Jacobian edge on the frozen 202-file population with median relative discrepancy 1.44×10^{-3}. **(b)** The predicted GLA/FGLA/Geometry-Oracle ordering holds for 173/202, 188/202, and 186/202 files at $H = 360, 720, 1440$, respectively; at $H = 360$, median final-error reductions over GLA are $5.086\times$ for FGLA and $8.967\times$ for the Geometry Oracle. **(c)** On the disjoint 4,752-run evaluation, Online attains 73.30% convergence success, versus 60.37% for GLA, 54.90% for FGLA, 52.06% for NAGLA, and 47.62% for AGLA. **(d)** Among jointly successful runs, median $q_{\text{Online}}/q_{\text{baseline}}$ is 1.00091, 1.00030, and 1.00050 against FGLA, NAGLA, and AGLA; bars show the 10–90% ranges. Together, the results support the geometry–spectrum–acceleration mechanism while revealing a distinct empirical effect of online adaptation: higher convergence success without a systematic conditional local-rate advantage.

Online adaptation improves convergence success, not conditional local rate. A consistent robustness–rate separation emerges across evaluation populations. Online achieves 74.94%, 73.30%, and 74.86% convergence success on the original, disjoint held-out, and larger MUSAN-25 evaluations, compared with 54.87%, 54.90%, and 55.04% for oracle-tuned FGLA. In contrast, conditional local rates are essentially matched: among jointly successful runs, median $q_{\text{Online}}/q_{\text{FGLA}}$ is 1.00093, 1.00091, and 1.00098, with Online attaining the smaller fitted factor in only 34.47%, 35.74%, and 33.55% of pairs. Thus, the reproducible benefit of online adaptation is higher convergence success, not systematically faster conditional local contraction.

The robustness gain persists against contemporary accelerations. On the disjoint 4,752-run evaluation, frozen Online achieves 73.30% convergence success, compared with 60.37% for GLA, 54.90% for optimal fixed FGLA, 52.06% for NAGLA with $H = 49$ (Cicognani et al., 2026), and 47.62% for the published Table-II AGLA configuration $(\alpha, \beta, \gamma) = (1.05, 1.35, 1.25)$ (Nenov et al., 2024); see Figure 1(c). Thus, the robustness advantage is not specific to comparison with classical GLA or fixed-momentum FGLA, but persists against the tested recent accelerated variants. Online

exceeds NAGLA and AGLA by 21.23 and 25.67 percentage points, with source-file cluster-bootstrap 95% intervals $[19.64, 22.57]$ and $[24.01, 27.06]$, respectively. The intervals remain well separated from zero even when dependence among runs from the same source file is accounted for through clustering.

This success advantage is not explained by faster conditional local contraction. Among jointly successful runs (Figure 1(d)), median $q_{\mathrm{Online}}/q_{\mathrm{baseline}}$ is 1.00091, 1.00030, and 1.00050 against FGLA, NAGLA, and AGLA, respectively. Instead, Online converts many baseline failures into successes: it rescues 889, 1025, and 1235 failures of FGLA, NAGLA, and AGLA while losing only 15, 16, and 15 corresponding baseline successes. This asymmetry identifies the main empirical effect of adaptation: Online succeeds on substantially more runs while essentially matching the conditional local rates of the accelerated baselines once both methods converge. NAGLA and AGLA use their published configurations rather than MUSAN-specific retuning; the comparison therefore applies to these literature configurations rather than all possible tunings.

Mechanism ablation. A frozen common-population ablation probes which components of Online account for the observed robustness gain by separating trajectory-adaptive spectral control from fixed one-edge control and from the envelope/fallback safeguards. Full Online improves convergence success over FGLA by 26.12 percentage points, while raw trajectory-adaptive spectral control exceeds the fixed one-edge comparator by 39.09 points. The latter comparison is particularly informative: both methods use the same one-edge design principle, but only the adaptive variant updates its spectral information from the observed trajectory. Envelope and fallback changes are smaller and interacting, so their effects are not interpreted additively. Thus, under this protocol, the dominant empirical gain is associated with trajectory-adaptive spectral estimation and control rather than the safeguards alone. This is a mechanism association, not a universal causal decomposition. Full ablations and protocol provenance are reported in Appendix F.

8 Discussion and Conclusion

This work identifies local geometry as the organizing principle for Griffin–Lim acceleration. At a regular feasible orbit, principal angles determine the transverse Griffin–Lim spectrum, and this same spectrum governs both local curvature and inertial dynamics. Acceleration therefore becomes an explicit instance-dependent spectral design problem. For the three-parameter AGLA family, we solve this problem exactly: the geometry-optimal parameters attain the minimax transverse root factor and strictly improve the optimal local linearized factors of FGLA and GLA on every nondegenerate spectral interval. These guarantees concern local linearized asymptotic rates, rather than global convergence.

A second contribution is to separate *available acceleration* from *realizable online acceleration*. Full knowledge of the transverse spectral interval yields the two-edge oracle optimum, whereas passive secants reveal only trajectory-active spectral information. This information boundary motivates the conservative one-edge design: under explicit excitation, consistency, and local-entry conditions, Ritz estimation recovers the dominant transverse edge and conditionally realizes the corresponding geometry-designed asymptotic rate. Closing the remaining gap to the full two-edge oracle requires additional spectral information that passive trajectories need not reveal.

The experiments expose a complementary robustness–rate separation. Principal-angle geometry accurately predicts the local spectrum and its acceleration consequences, while online adaptation substantially increases convergence success without systematically improving conditional local contraction. This pattern reproduces across real-audio evaluations and persists against the tested NAGLA and AGLA configurations. The ablation further associates the dominant robustness gain with trajectory-adaptive spectral estimation and control. Thus, the main empirical value of adaptation is not simply a better asymptotic contraction factor: it is the ability to translate trajectory-active geometric information into substantially more robust convergence under the tested protocols.

An end-to-end theory connecting spectral estimation, safeguards, basin entry, and convergence remains open. Multi-edge online estimation could further narrow the gap to the full interval-aware oracle, while a theory of adaptive basin enlargement could help explain the robustness gains observed here. More broadly, the geometry–spectrum–information viewpoint suggests a route to instance-adaptive acceleration for alternating-projection and smooth fixed-point methods beyond Griffin–Lim.

REPRODUCIBILITY STATEMENT

All assumptions used by the theoretical results are stated in the main text, and complete proofs are provided in Appendices A–E. Appendix F records the frozen experimental protocols, population accounting, validity rules, rate-fitting conventions, and robustness diagnostics. The anonymized supplementary materials provide the frozen aggregate ledgers and analysis/figure scripts needed to audit the reported numerical summaries and regenerate the main evidence plots; they are not presented as a standalone reimplementation of the underlying Griffin–Lim solver. The anonymous reproducibility repository is available at `https://anonymous.4open.science/r/geometry-accelerated-griffin-lim-1725/`.

AI USE STATEMENT

Generative AI tools were used as research assistants for information searching, proof auditing, code implementation and review, and manuscript editing. The authors reviewed and verified the mathematical claims, proofs, experimental protocols, code paths used for reported results, numerical outputs, citations, and final wording, and take full responsibility for the submission.

REFERENCES

Rima Alaifari and Yunan Yang. Multiwindow approaches for direct and stable stft phase retrieval. *SIAM Journal on Applied Mathematics*, 85(5):2376–2398, 2025. doi: 10.1137/24M1701332.

David Bartusel. Injectivity conditions for stft phase retrieval on $\mathbb{Z}$, $\mathbb{Z}_d$ and $\mathbb{R}^d$. *Journal of Fourier Analysis and Applications*, 29:53, 2023. doi: 10.1007/s00041-023-10026-2.

Massimo Cicognani, Alberto Bernardini, Alessandro Ilic Mezza, and Roberto Leone Cicognani. Gradient descent in phase retrieval. *Annali dell'Università di Ferrara*, 72, 2026. doi: 10.1007/s11565-026-00753-x.

James R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1982. doi: 10.1364/AO.21.002758.

R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35(2):237–246, 1972.

Daniel W. Griffin and Jae S. Lim. Signal estimation from modified short-time fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984. doi: 10.1109/TASSP.1984.1164317.

Philipp Grohs and Lukas Liehr. On foundational discretization barriers in stft phase retrieval. *Journal of Fourier Analysis and Applications*, 28:39, 2022. doi: 10.1007/s00041-022-09935-5.

Irina-Beatrice Haas and Yuji Nakatsukasa. Sharp error bounds for approximate eigenvalues and singular values from subspace methods. *SIAM Journal on Matrix Analysis and Applications*, 47 (1):308–329, 2026. doi: 10.1137/25M1764554.

Adrian S. Lewis and Jérôme Malick. Alternating projections on manifolds. *Mathematics of Operations Research*, 33(1):216–234, 2008. doi: 10.1287/moor.1070.0291.

Adrian S. Lewis, D. Russell Luke, and Jérôme Malick. Local linear convergence for alternating and averaged nonconvex projections. *Foundations of Computational Mathematics*, 9(4):485–513, 2009. doi: 10.1007/s10208-008-9036-y.

Qunwei Li, Yi Zhou, Yingbin Liang, and Pramod K. Varshney. Convergence analysis of proximal gradient with momentum for nonconvex optimization. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2111–2119. PMLR, 2017.

D. Russell Luke and Anna-Lena Martins. Convergence analysis of iterative algorithms for phase retrieval. In Tim Salditt, Alexander Egner, and D. Russell Luke, editors, *Nanoscale Photonic Imaging*, volume 134 of *Topics in Applied Physics*, pages 583–601. Springer, Cham, 2020. doi: 10.1007/978-3-030-34413-9_23.

Stefano Marchesini, Yu-Chao Tu, and Hau-Tieng Wu. Alternating projection, ptychographic imaging and phase synchronization. *Applied and Computational Harmonic Analysis*, 41(3):815–851, 2016. doi: 10.1016/j.acha.2015.06.005.

Alessandro Ilic Mezza, Massimo Cicognani, Roberto Leone Cicognani, and Alberto Bernardini. On the convergence of the fast griffin–lim algorithm. *IEEE Signal Processing Letters*, 33:2006–2010, 2026. doi: 10.1109/LSP.2026.3693202.

Yuji Nakatsukasa and Joel A. Tropp. Fast and accurate randomized algorithms for linear systems and eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 45(2):1183–1214, 2024. doi: 10.1137/23M1565413.

Rossen Nenov, Dang-Khoa Nguyen, Peter Balazs, and Radu Ioan Boţ. Accelerated griffin–lim algorithm: A fast and provably converging numerical method for phase retrieval. *IEEE Transactions on Signal Processing*, 72:190–202, 2024. doi: 10.1109/TSP.2023.3339232.

Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015. doi: 10.1109/TSP.2015.2448516.

Peter Ochs, Yunjin Chen, Thomas Brox, and Thomas Pock. ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419, 2014. doi: 10.1137/130942954.

Nathanaël Perraudin, Peter Balazs, and Peter L. Søndergaard. A fast griffin–lim algorithm. In *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, 2013. doi: 10.1109/WASPAA.2013.6701851.

Yousef Saad. *Numerical Methods for Large Eigenvalue Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA, revised edition, 2011. ISBN 9781611970722. doi: 10.1137/1.9781611970739.

David Snyder, Guoguo Chen, and Daniel Povey. MUSAN: A Music, Speech, and Noise Corpus, 2015. arXiv:1510.08484v1.

Nguyen Hieu Thao, David Russell Luke, Oleg Soloviev, and Michel Verhaegen. Phase retrieval with sparse phase constraint. *SIAM Journal on Mathematics of Data Science*, 2(1):246–263, 2020. doi: 10.1137/19M1266800.

Nguyen Hieu Thao, Oleg Soloviev, and Michel Verhaegen. Convex combination of alternating projection and douglas–rachford operators for phase retrieval. *Advances in Computational Mathematics*, 47(3):33, 2021. doi: 10.1007/s10444-021-09861-y.

A Projection Geometry and Principal Angles

This appendix proves Theorem 3.1 and corollary 3.2. All orthogonality and adjoints are taken with respect to the underlying real inner product

$$\langle x, y\rangle_{\mathbb{R}} := \Re\langle x, y\rangle_{\mathbb{C}}.$$

A.1 Derivative of the magnitude projection

For $s > 0$, define

$$p_s : \mathbb{C} \setminus \{0\} \to \mathbb{C}, \qquad p_s(z) := s\frac{z}{|z|}.$$

We identify $\mathbb{C}$ with $\mathbb{R}^2$ when differentiating.

Lemma A.1 (Scalar projection derivative)**.** *Let $z \neq 0$, set $u = z/|z|$, and let $h \in \mathbb{C}$. Then,*

$$Dp_s(z)[h] = \frac{s}{|z|}\left(h - u\,\Re(\bar{u}h)\right). \tag{A.1}$$

If $|z| = s$, then $Dp_s(z)$ is the real-orthogonal projector onto

$$T_z\{w \in \mathbb{C} : |w| = s\} = iu\mathbb{R}.$$

Proof. For $r(z) = |z|$,

$$Dr(z)[h] = \frac{\Re(\bar{z}h)}{|z|} = \Re(\bar{u}h).$$

Since $p_s(z) = sz/r(z)$,

$$Dp_s(z)[h] = \frac{s}{|z|}\left(h - u\Re(\bar{u}h)\right).$$

At $|z| = s$, every $h \in \mathbb{C}$ decomposes as

$$h = u\Re(\bar{u}h) + iu\,\Im(\bar{u}h),$$

so the displayed derivative is precisely the real-orthogonal projection onto $iu\mathbb{R}$. □

A.2 Product magnitude manifold

By Assumption 2.1, $s_i > 0$ for every coordinate, hence

$$\mathcal{M}_s = \prod_{i=1}^{M} s_i S^1 = \{c \in \mathbb{C}^M : |c_i| = s_i,\ i = 1, \ldots, M\}$$

is a smooth real product manifold. Writing

$$u_i^\star := \frac{c_i^\star}{|c_i^\star|},$$

its tangent space at $c^\star$ is

$$V = T_{c^\star}\mathcal{M}_s = \left\{h \in \mathbb{C}^M : \Re(\overline{u_i^\star}h_i) = 0\ \forall i\right\} = \prod_{i=1}^{M} iu_i^\star\mathbb{R}. \tag{A.2}$$

Lemma A.2 (Derivative of the magnitude projection)**.** *Under Assumption 2.1, $P_{\mathcal{M}_s}$ is single-valued and smooth in a neighborhood of $c^\star$, and*

$$DP_{\mathcal{M}_s}(c^\star) = P_V.$$

Proof. Since every $c_i^\star \neq 0$, in a sufficiently small neighborhood of $c^\star$ no coordinate vanishes, and the nearest-point projection is componentwise:

$$[P_{\mathcal{M}_s}(c)]_i = s_i\frac{c_i}{|c_i|}.$$

Applying Lemma A.1 coordinatewise gives

$$[DP_{\mathcal{M}_s}(c^\star)h]_i = h_i - u_i^\star\Re(\overline{u_i^\star}h_i),$$

which is exactly $P_V h$ by (A.2). □

A.3 JACOBIAN AND PRINCIPAL-ANGLE SPECTRUM

Since $G = P_{\mathcal{U}} P_{\mathcal{M}_s}$ and $P_{\mathcal{U}}$ is linear,

$$J_\star = DG(c^\star)|_{\mathcal{U}} = P_{\mathcal{U}} P_V|_{\mathcal{U}}. \tag{A.3}$$

Lemma A.3 (Self-adjointness, positivity, and spectrum)**.** *The operator $J_\star : \mathcal{U} \to \mathcal{U}$ is self-adjoint and positive semidefinite. For every $x, y \in \mathcal{U}$,*

$$\langle x, J_\star y \rangle_{\mathbb{R}} = \langle P_V x, P_V y \rangle_{\mathbb{R}}\,, \tag{A.4}$$

and therefore,

$$\langle x, J_\star x \rangle_{\mathbb{R}} = \|P_V x\|^2. \tag{A.5}$$

Its nonzero eigenvalues are $\cos^2 \theta_j$, where θ_j are the principal angles between the real subspaces $\mathcal{U}$ and V.

Proof. For $x, y \in \mathcal{U}$,

$$\langle x, J_\star y \rangle_{\mathbb{R}} = \langle x, P_{\mathcal{U}} P_V y \rangle_{\mathbb{R}} = \langle x, P_V y \rangle_{\mathbb{R}} = \langle P_V x, P_V y \rangle_{\mathbb{R}}\,.$$

The right-hand side is symmetric in x and y, proving self-adjointness, while (A.5) gives positive semidefiniteness.

Now let

$$A := P_V|_{\mathcal{U}} : \mathcal{U} \to V.$$

Its real adjoint is $A^* = P_{\mathcal{U}}|_V$, so

$$A^* A = P_{\mathcal{U}} P_V|_{\mathcal{U}} = J_\star.$$

The singular values of $P_V|_{\mathcal{U}}$ are the cosines of the principal angles between $\mathcal{U}$ and V. Hence the nonzero eigenvalues of $J_\star = A^* A$ are $\cos^2 \theta_j$. □

A.4 GLOBAL PHASE AND TRANSVERSE DECOMPOSITION

Lemma A.4 (Neutral phase direction)**.** *The vector $ic^\star$ belongs to $\mathcal{U} \cap V$ and satisfies*

$$J_\star(ic^\star) = ic^\star.$$

Under Assumption 2.1,

$$\ker(I - J_\star) = \mathcal{U} \cap V = \operatorname{span}_{\mathbb{R}}\{ic^\star\}.$$

Consequently,

$$\mathcal{N}_\star = (ic^\star)^\perp \cap \mathcal{U}$$

is $J_\star$-invariant and

$$\|J_\star|_{\mathcal{N}_\star}\| < 1.$$

Proof. Because $\mathcal{U}$ is complex linear and $c^\star \in \mathcal{U}$, we have $ic^\star \in \mathcal{U}$. Moreover,

$$ic_i^\star = is_i u_i^\star \in iu_i^\star \mathbb{R},$$

so $ic^\star \in V$. Hence, (A.3) gives $J_\star(ic^\star) = ic^\star$.

If $x \in \mathcal{U}$ and $J_\star x = x$, then by (A.5),

$$\|x\|^2 = \langle x, J_\star x \rangle_{\mathbb{R}} = \|P_V x\|^2.$$

Equality in the norm contraction of the orthogonal projector P_V implies $x \in V$. Thus, $\ker(I - J_\star) \subseteq \mathcal{U} \cap V$, while the reverse inclusion follows immediately from (A.3). Therefore,

$$\ker(I - J_\star) = \mathcal{U} \cap V = \operatorname{span}_{\mathbb{R}}\{ic^\star\}$$

by Assumption 2.1.

Self-adjointness makes the orthogonal complement of this eigenspace invariant. Moreover, (A.5) implies $\operatorname{spec}(J_\star) \subset [0, 1]$. Since no eigenvalue 1 remains on $\mathcal{N}_\star$, finite dimensionality gives

$$\|J_\star|_{\mathcal{N}_\star}\| < 1.$$

□

Proof of Theorem 3.1. Equation (A.3) gives the Jacobian identity. Lemma A.3 gives self-adjointness, positivity, and the principal-angle spectrum, while Lemma A.4 gives the phase and transverse decomposition. Thus, whenever $\mathcal{N}_\star \neq \{0\}$,

$$\lambda_- := \lambda_{\min}(J_\star|_{\mathcal{N}_\star}), \qquad \lambda_+ := \lambda_{\max}(J_\star|_{\mathcal{N}_\star})$$

satisfy

$$0 \leq \lambda_- \leq \lambda_+ < 1.$$

The geometry theorem itself requires no spectral nondegeneracy; the acceleration results in Section 4 separately assume $0 \leq \lambda_- < \lambda_+ < 1$. □

A.5 Gradient, Hessian, and quadratic growth modulo phase

Lemma A.5 (Restricted gradient and Hessian)**.** *In a neighborhood of $c^\star$, for $c \in \mathcal{U}$,*

$$\nabla(f|_{\mathcal{U}})(c) = c - G(c). \tag{A.6}$$

Moreover,

$$\nabla^2(f|_{\mathcal{U}})(c^\star) = I - J_\star. \tag{A.7}$$

Proof. Near $c^\star$, $P_{\mathcal{M}_s}$ is single-valued and smooth. Hence

$$\nabla\left(\frac{1}{2}d^2_{\mathcal{M}_s}\right)(c) = c - P_{\mathcal{M}_s}(c).$$

Restricting to the linear subspace $\mathcal{U}$ and using $c \in \mathcal{U}$ gives

$$\nabla(f|_{\mathcal{U}})(c) = P_{\mathcal{U}}(c - P_{\mathcal{M}_s}(c)) = c - G(c).$$

Differentiating at $c^\star$ yields (A.7). □

Lemma A.6 (Phase alignment)**.** *There exists a neighborhood $\mathcal{W}$ of $\mathcal{O}^\star$ in $\mathcal{U}$ such that every $c \in \mathcal{W}$ admits a phase $\phi(c)$ for which*

$$\widetilde{c} := e^{-i\phi(c)}c = c^\star + h, \qquad h \in \mathcal{N}_\star,$$

and

$$\|h\| = \operatorname{dist}(c, \mathcal{O}^\star). \tag{A.8}$$

Moreover, $f(\widetilde{c}) = f(c)$.

Proof. The orbit

$$\mathcal{O}^\star = \{e^{i\phi}c^\star : \phi \in \mathbb{R}\}$$

is a smooth embedded circle with tangent $ie^{i\phi}c^\star$ at $e^{i\phi}c^\star$. In a sufficiently small tubular neighborhood, every c has a unique nearest point $e^{i\phi(c)}c^\star$ on the orbit. Hence,

$$c - e^{i\phi(c)}c^\star \perp_{\mathbb{R}} ie^{i\phi(c)}c^\star.$$

Multiplication by $e^{-i\phi(c)}$ preserves the real inner product, so

$$h := e^{-i\phi(c)}c - c^\star \perp_{\mathbb{R}} ic^\star.$$

Because $c, c^\star \in \mathcal{U}, \mathcal{U}$ is complex linear, and therefore invariant under multiplication by $e^{-i\phi(c)}$, we also have $h \in \mathcal{U}$. Consequently,

$$h \in (ic^\star)^\perp \cap \mathcal{U} = \mathcal{N}_\star.$$

Unitary invariance gives

$$\|h\| = \|c - e^{i\phi(c)}c^\star\| = \operatorname{dist}(c, \mathcal{O}^\star).$$

Finally, $\mathcal{M}_s$ is invariant under global phase, so $d_{\mathcal{M}_s}(e^{-i\phi}c) = d_{\mathcal{M}_s}(c)$ and hence, $f(e^{-i\phi}c) = f(c)$. □

Proof of Corollary 3.2. The gradient and Hessian identities follow from Lemma A.5. By Theorem 3.1, for every $h \in \mathcal{N}_\star$,

$$(1-\lambda_+)\|h\|^2 \le \langle h, (I-J_\star)h\rangle_{\mathbb{R}} \le (1-\lambda_-)\|h\|^2.$$

Thus, the extreme transverse Hessian eigenvalues are $1-\lambda_+$ and $1-\lambda_-$, and

$$\kappa_{\mathrm{geom}} = \frac{1-\lambda_-}{1-\lambda_+}.$$

Since $f|_{\mathcal{U}}$ is C^2 near $c^\star$, $f(c^\star)=0$, and $\nabla(f|_{\mathcal{U}})(c^\star)=0$, Taylor expansion along $\mathcal{N}_\star$ gives

$$f(c^\star+h) - f(c^\star) = \frac{1}{2}\langle h, (I-J_\star)h\rangle_{\mathbb{R}} + o(\|h\|^2). \tag{A.9}$$

Fix $\varepsilon \in (0, 1-\lambda_+)$. After shrinking the neighborhood,

$$|o(\|h\|^2)| \le \frac{\varepsilon}{2}\|h\|^2,$$

and therefore,

$$\frac{1-\lambda_+-\varepsilon}{2}\|h\|^2 \le f(c^\star+h) - f(c^\star) \le \frac{1-\lambda_-+\varepsilon}{2}\|h\|^2.$$

Applying Lemma A.6 replaces $\|h\|$ by $\mathrm{dist}(c, \mathcal{O}^\star)$ while leaving f invariant, proving (3.1). □

B Modal Reduction and Exact Schur Stability

This appendix derives the modal linearization and the exact Schur conditions used in Section 4. All linearizations are taken on the real Hilbert space $\mathcal{U}$ from Section 2. By Theorem 3.1, $J_\star$ is self-adjoint and therefore admits a real-orthonormal eigenbasis.

B.1 Two-step linearization

Lemma B.1 (Linearization of the inertial family)**.** *Let $e_n := t_n - c^\star$ and suppose $(e_{n-1}, e_{n-2}) \to (0,0)$. Then,*

$$\begin{aligned} e_n = &\Big[(1-\gamma)(1+\beta)I + \gamma(1+\alpha)J_\star\Big]e_{n-1} \\ &- \Big[(1-\gamma)\beta I + \gamma\alpha J_\star\Big]e_{n-2} + o(\|(e_{n-1}, e_{n-2})\|)\,. \end{aligned} \tag{B.1}$$

Proof. From (4.1),

$$c_{n-1} - c^\star = (1+\alpha)e_{n-1} - \alpha e_{n-2}, \qquad d_{n-1} - c^\star = (1+\beta)e_{n-1} - \beta e_{n-2}.$$

Since $G(c^\star) = c^\star$ and $DG(c^\star)|_{\mathcal{U}} = J_\star$,

$$G(c_{n-1}) - c^\star = J_\star(c_{n-1} - c^\star) + o(\|c_{n-1} - c^\star\|).$$

The map $(e_{n-1}, e_{n-2}) \mapsto c_{n-1} - c^\star$ is linear and bounded, so the remainder is $o(\|(e_{n-1}, e_{n-2})\|)$. Substitution into the t_n update yields (B.1). □

B.2 Reduction to scalar geometric modes

Lemma B.2 (Scalar modal polynomial)**.** *Let $J_\star v = \lambda v$. On the two-state invariant subspace generated by $(v,0)$ and $(0,v)$, the linear part of* (B.1) *is*

$$M_\lambda = \begin{pmatrix} A_\lambda & -C_\lambda \\ 1 & 0 \end{pmatrix}, \tag{B.2}$$

where

$$A_\lambda = (1-\gamma)(1+\beta) + \gamma\lambda(1+\alpha), \qquad C_\lambda = (1-\gamma)\beta + \gamma\lambda\alpha. \tag{B.3}$$

Its characteristic polynomial is

$$p_\lambda(r) = r^2 - A_\lambda r + C_\lambda. \tag{B.4}$$

Proof. Substituting $e_{n-j} = z_{n-j}v$ into the linear part of (B.1) gives

$$z_n = A_\lambda z_{n-1} - C_\lambda z_{n-2}.$$

Thus, $(z_n, z_{n-1})^\top = M_\lambda(z_{n-1}, z_{n-2})^\top$, and direct evaluation of $\det(rI - M_\lambda)$ gives (B.4). □

B.3 Exact quadratic Schur criterion

Lemma B.3 (Schur criterion for a monic real quadratic)**.** *Let*

$$p(r) = r^2 - Ar + C, \qquad A, C \in \mathbb{R}.$$

Both roots lie strictly inside the unit disk if and only if

$$|C| < 1, \qquad 1 - A + C > 0, \qquad 1 + A + C > 0. \tag{B.5}$$

Proof. Apply the Cayley transform

$$z = \frac{r-1}{r+1}, \qquad r = \frac{1+z}{1-z},$$

which maps the open unit disk to the open left half-plane. Multiplying $p((1+z)/(1-z))$ by $(1-z)^2$ gives

$$(1 - A + C) + 2(1 - C)z + (1 + A + C)z^2.$$

A real quadratic $a_2 z^2 + a_1 z + a_0$ with $a_2 > 0$ is Hurwitz if and only if $a_1 > 0$ and $a_0 > 0$. Hence Schur stability is equivalent to

$$1 - C > 0, \qquad 1 - A + C > 0, \qquad 1 + A + C > 0.$$

The last two inequalities imply $C > -1$, so these conditions are equivalent to (B.5). □

B.4 Transverse and phase stability

Proposition B.4 (Exact modal Schur stability)**.** *For a transverse eigenvalue $\lambda < 1$ and $\gamma > 0$, M_λ is Schur stable if and only if*

$$|(1 - \gamma)\beta + \gamma\lambda\alpha| < 1 \tag{B.6}$$

and

$$1 + (1 - \gamma)(1 + 2\beta) + \gamma\lambda(1 + 2\alpha) > 0. \tag{B.7}$$

At the phase eigenvalue $\lambda = 1$,

$$p_1(r) = (r - 1)(r - C_1), \qquad C_1 = (1 - \gamma)\beta + \gamma\alpha. \tag{B.8}$$

Proof. By Lemma B.3, the transverse conditions are

$$|C_\lambda| < 1, \qquad 1 - A_\lambda + C_\lambda > 0, \qquad 1 + A_\lambda + C_\lambda > 0.$$

Using (B.3),

$$1 - A_\lambda + C_\lambda = \gamma(1 - \lambda) > 0,$$

so the middle condition is automatic for $\gamma > 0$ and $\lambda < 1$. The remaining plus condition becomes

$$1 + A_\lambda + C_\lambda = 1 + (1 - \gamma)(1 + 2\beta) + \gamma\lambda(1 + 2\alpha),$$

giving (B.6)–(B.7).

At $\lambda = 1$,

$$A_1 = (1 - \gamma)(1 + \beta) + \gamma(1 + \alpha) = 1 + C_1,$$

and hence,

$$p_1(r) = r^2 - (1 + C_1)r + C_1 = (r - 1)(r - C_1).$$

□

Corollary B.5 (Orbital Schur stability of the linearization)**.** *The two-step linearized dynamics are Schur stable modulo global phase if and only if every transverse block M_λ is Schur stable for*

$$\lambda \in \operatorname{spec}(J_\star|_{\mathcal{N}_\star})$$

and

$$|C_1| < 1.$$

After quotienting the phase unit root, the spectral radius of the full linearized state is

$$\rho_{\text{quot}} = \max\left\{|C_1|, \max_{\lambda \in \operatorname{spec}(J_\star|_{\mathcal{N}_\star})} \rho(M_\lambda)\right\}. \tag{B.9}$$

Proof. Choose a real-orthonormal eigenbasis of $J_\star$ adapted to

$$\mathcal{U} = \mathrm{span}_{\mathbb{R}}\{ic^\star\} \oplus \mathcal{N}_\star.$$

In the induced two-step coordinates, the linearized state operator decomposes into the companion blocks M_λ. Every transverse component decays exactly when its block is Schur stable. At $\lambda = 1$, (B.8) gives the neutral phase root 1 and the secondary root C_1. Quotienting the symmetry removes the former, leaving (B.9). □

Transverse versus quotient-state design. The minimax problem in Section 4 optimizes the transverse factor

$$R(\alpha, \beta, \gamma; \lambda_-, \lambda_+) = \sup_{\lambda \in [\lambda_-, \lambda_+]} \rho(M_\lambda). \tag{B.10}$$

This is distinct from the full quotient-state factor (B.9). For the optimizer in Theorem 4.1,

$$C_1^\star = \frac{(q^\star)^2}{\lambda_+}.$$

Since Corollary 4.2 gives $q^\star < \lambda_+ < 1$,

$$0 \le C_1^\star < q^\star < 1.$$

Thus, the secondary phase root neither violates orbital stability nor dominates the optimized transverse factor.

B.5 Radius-q Schur conditions

The minimax proof requires the same criterion on a disk of arbitrary radius.

Lemma B.6 (Roots in a disk of radius q)**.** *Let $q > 0$ and $p(r) = r^2 - Ar + C$ with $A, C \in \mathbb{R}$. Both roots satisfy $|r| < q$ if and only if*

$$|C| < q^2, \tag{B.11}$$

$$q^2 - qA + C > 0, \tag{B.12}$$

$$q^2 + qA + C > 0. \tag{B.13}$$

Both roots satisfy $|r| \le q$ if and only if

$$|C| \le q^2, \qquad q^2 - qA + C \ge 0, \qquad q^2 + qA + C \ge 0. \tag{B.14}$$

Proof. Set $r = qu$. Dividing $p(qu)$ by q^2 gives

$$\widetilde{p}(u) = u^2 - \frac{A}{q}u + \frac{C}{q^2}.$$

The strict statement follows immediately from Lemma B.3.

For the closed disk, necessity follows by approximating any polynomial whose roots lie in $|u| \le 1$ by polynomials obtained by radially contracting both roots by a factor $1 - \delta$, and then letting $\delta \downarrow 0$.

Conversely, suppose the three non-strict inequalities hold. For any $\eta \in (0, 1)$ define

$$A_\eta := \eta A, \qquad C_\eta := \eta^2 C.$$

As a result,

$$|C_\eta| \le \eta^2 q^2 < q^2,$$

and

$$q^2 - qA_\eta + C_\eta = (1-\eta)q^2 + \eta(q^2 - qA + C) + \eta(1-\eta)(q^2 - C) > 0,$$

while similarly

$$q^2 + qA_\eta + C_\eta = (1-\eta)q^2 + \eta(q^2 + qA + C) + \eta(1-\eta)(q^2 - C) > 0.$$

It turns out that the roots of $r^2 - A_\eta r + C_\eta$ lie strictly inside $|r| < q$. As $\eta \uparrow 1$, the coefficients converge to (A, C), and continuity of polynomial roots implies that the roots of p lie in $|r| \le q$. □

C Proof of the Exact Minimax Design

This appendix proves Theorem 4.1 and corollary 4.2 and derives the strict rate hierarchy (4.9), the dominant-edge-only formula (4.10), and the spectral-width asymptotic (4.11).

Throughout, write

$$m := \lambda_-, \qquad L := \lambda_+, \qquad D := L - m > 0,$$

so that

$$0 \le m < L < 1,$$

and define

$$B := 3L + m - L^2 - 3Lm. \tag{C.1}$$

For convenience, let

$$s_\lambda := A_\lambda - C_\lambda. \tag{C.2}$$

Using the modal coefficients in (4.3), the inertial terms cancel:

$$s_\lambda = 1 - \gamma(1 - \lambda). \tag{C.3}$$

This cancellation is the key algebraic structure behind the closed-form minimax solution.

C.1 Robust radius-q feasibility reduces to endpoints

Lemma C.1 (Endpoint reduction)**.** *Fix $q > 0$ and parameters $\alpha, \beta \ge 0$ and $\gamma > 0$. All roots of p_λ lie in the closed disk $|r| \le q$ for every $\lambda \in [m, L]$ if and only if the radius-q conditions*

$$|C_\lambda| \le q^2, \tag{C.4}$$

$$q^2 - qA_\lambda + C_\lambda \ge 0, \tag{C.5}$$

$$q^2 + qA_\lambda + C_\lambda \ge 0 \tag{C.6}$$

hold at both endpoints $\lambda = m$ and $\lambda = L$.

Proof. By Lemma B.6, (C.4)–(C.6) are necessary and sufficient at each fixed λ. Since A_λ and C_λ are affine in λ, the absolute-value condition is equivalent to the two affine inequalities

$$C_\lambda \le q^2, \qquad -C_\lambda \le q^2,$$

while the left-hand sides of (C.5) and (C.6) are affine as well. An affine function satisfies a one-sided inequality throughout a compact interval if and only if it does so at both endpoints. Hence, all these affine inequalities need only be checked at m and L. □

C.2 A universal upper-edge bound and exclusion of $\gamma \le 1$

Lemma C.2 (Upper-edge root-radius bound)**.** *Suppose every root of p_L lies in $|r| \le q$ with $0 < q < 1$. Then*

$$s_L \le 2q - q^2. \tag{C.7}$$

Consequently, if $0 < \gamma \le 1$, then

$$q \ge q_{\mathrm{F}} := 1 - \sqrt{1 - L}. \tag{C.8}$$

Proof. From (C.5) at $\lambda = L$ and $A_L = s_L + C_L$,

$$q^2 - qs_L + (1 - q)C_L \ge 0.$$

Because $q < 1$ and (C.4) gives $C_L \le q^2$,

$$qs_L \le q^2 + (1 - q)C_L \le q^2 + (1 - q)q^2 = q^2(2 - q).$$

Dividing by $q > 0$ yields (C.7).

If $0 < \gamma \le 1$, then (C.3) gives

$$s_L = 1 - \gamma(1 - L) \ge L.$$

Therefore,

$$L \le 2q - q^2.$$

Since $q \mapsto 2q - q^2$ is strictly increasing on $[0, 1]$, solving the corresponding equality gives

$$q \ge 1 - \sqrt{1 - L} = q_{\mathrm{F}}.$$

□

Consequence for acceleration. Lemma C.2 shows that every design with $0 < \gamma \leq 1$ has worst-case transverse factor at least

$$q_{\mathrm{F}} = 1 - \sqrt{1 - L}.$$

Consequently, any member of the three-parameter family that strictly improves upon the optimal FGLA factor must use genuine over-relaxation,

$$\gamma > 1.$$

This separates the minimax proof into two regimes: $\gamma \leq 1$ is already lower-bounded by q_{F}, whereas the strictly better regime $\gamma > 1$ admits the sharper interval-dependent bound derived next.

C.3 Lower bound for every $\gamma > 1$ design

Assume $\gamma > 1$, and define

$$c := (1 - \gamma)\beta \leq 0, \qquad d := \gamma\alpha \geq 0. \tag{C.9}$$

Then,

$$C_\lambda = c + d\lambda. \tag{C.10}$$

Lemma C.3 (Necessary quadratic inequality)**.** *Suppose $\alpha, \beta \geq 0$, $\gamma > 1$, and*

$$\sup_{\lambda \in [m, L]} \rho(M_\lambda) \leq q$$

for some $0 < q < 1$. Then, necessarily

$$P(q) := Dq^2 - Bq + LD \leq 0. \tag{C.11}$$

Proof. At the upper endpoint, (C.4) gives

$$C_L \leq q^2.$$

Since $c \leq 0$ and $m/L \in [0, 1)$,

$$C_m = c + dm \tag{C.12}$$

$$= \frac{m}{L}(c + dL) + \left(1 - \frac{m}{L}\right) c \tag{C.13}$$

$$\leq \frac{m}{L} C_L \leq \frac{m}{L} q^2. \tag{C.14}$$

By Lemma C.2,

$$s_L \leq 2q - q^2.$$

Together with

$$s_L = 1 - \gamma(1 - L),$$

this implies

$$\gamma \geq \frac{(1 - q)^2}{1 - L}. \tag{C.15}$$

Hence,

$$s_m = 1 - \gamma(1 - m) \leq 1 - \frac{(1 - m)(1 - q)^2}{1 - L}. \tag{C.16}$$

Now apply the lower-endpoint plus-Jury inequality (C.6). Since $A_m = s_m + C_m$,

$$0 \leq q^2 + qA_m + C_m = q^2 + qs_m + (q + 1)C_m.$$

Using (C.14) and (C.16) gives the necessary condition

$$0 \leq q^2 + q\left[1 - \frac{(1 - m)(1 - q)^2}{1 - L}\right] + (q + 1)\frac{m}{L} q^2. \tag{C.17}$$

Direct simplification yields

$$q^2 + q\left[1 - \frac{(1 - m)(1 - q)^2}{1 - L}\right] + (q + 1)\frac{m}{L} q^2 = -\frac{q}{L(1 - L)} P(q). \tag{C.18}$$

Since $q > 0$, $L > 0$, and $1 - L > 0$, (C.17) is equivalent to $P(q) \leq 0$. □

C.4 Location of the minimax root

Lemma C.4 (Location of the minimax root)**.** *Define*

$$q_\star := \frac{B - \sqrt{B^2 - 4LD^2}}{2D}. \tag{C.19}$$

Then

$$0 < q_\star < q_{\mathrm{F}} < L < 1. \tag{C.20}$$

In particular, the discriminant is strictly positive and $q_\star$ is the smaller positive root of P.

Proof. First, $B > 0$. Indeed, B is affine in m, and on $m \in [0, L]$ its endpoint values are

$$B|_{m=0} = L(3 - L) > 0, \qquad B|_{m=L} = 4L(1 - L) > 0.$$

Thus, $B > 0$ throughout the relevant range.

Moreover,

$$P(0) = LD > 0.$$

Let

$$q_{\mathrm{F}} = 1 - \sqrt{1 - L}.$$

Direct substitution gives

$$P(q_{\mathrm{F}}) = -(1 - L)(L + 3m)q_{\mathrm{F}} < 0. \tag{C.21}$$

Therefore, P has a real root in $(0, q_{\mathrm{F}})$, so its discriminant is strictly positive. Since P has leading coefficient $D > 0$, root product

$$\frac{LD}{D} = L > 0$$

and root sum

$$\frac{B}{D} > 0,$$

both roots are positive. The root in $(0, q_{\mathrm{F}})$ is therefore the smaller root and equals (C.19). Hence

$$0 < q_\star < q_{\mathrm{F}}.$$

Finally,

$$q_{\mathrm{F}} < L \iff 1 - \sqrt{1 - L} < L \iff \sqrt{1 - L} > 1 - L,$$

which is strict for $L \in (0, 1)$. This proves (C.20). □

Proposition C.5 (Global minimax lower bound)**.** *For every*

$$\alpha, \beta \geq 0, \qquad \gamma > 0,$$

the transverse worst-case factor satisfies

$$R(\alpha, \beta, \gamma; m, L) \geq q_\star. \tag{C.22}$$

Proof. Write

$$R := R(\alpha, \beta, \gamma; m, L).$$

If $R \geq q_{\mathrm{F}}$, then $R > q_\star$ by Lemma C.4. Suppose instead that $R < q_{\mathrm{F}}$. Choose any $q \in [R, q_{\mathrm{F}})$. Every modal root on $[m, L]$ then lies in $|r| \leq q$. By Lemma C.2, this is impossible for $\gamma \leq 1$, so necessarily $\gamma > 1$. Hence, Lemma C.3 applies and gives

$$P(q) \leq 0.$$

Because $q_\star$ is the smaller root of the upward-opening quadratic P,

$$P(q) > 0 \qquad \text{for } 0 \leq q < q_\star.$$

Therefore, $q \geq q_\star$. Letting $q \downarrow R$ yields

$$R \geq q_\star.$$

As a consequence, every admissible parameter triple satisfies

$$R(\alpha, \beta, \gamma; m, L) \geq q_\star.$$

□

C.5 ATTAINMENT AND ENDPOINT EQUIOSCILLATION

Proposition C.6 (Attainment). *Let $q = q_\star$ and define*

$$\beta^\star = 0, \qquad \gamma^\star = \frac{(1-q)^2}{1-L}, \qquad \alpha^\star = \frac{q^2(1-L)}{L(1-q)^2}. \tag{C.23}$$

Then,

$$\sup_{\lambda\in[m,L]} \rho(M_\lambda) = q. \tag{C.24}$$

Proof. Since $\beta^\star = 0$,

$$C_\lambda^\star = \gamma^\star\alpha^\star\lambda = \frac{q^2}{L}\lambda. \tag{C.25}$$

Moreover,

$$s_\lambda^\star = 1 - \gamma^\star(1-\lambda). \tag{C.26}$$

At $\lambda = L$,

$$C_L^\star = q^2, \qquad s_L^\star = 1-(1-q)^2 = 2q - q^2,$$

and therefore,

$$A_L^\star = 2q.$$

It implies that

$$p_L(r) = (r-q)^2. \tag{C.27}$$

At $\lambda = m$,

$$C_m^\star = \frac{m}{L}q^2.$$

Because $P(q) = 0$, direct substitution into $A_m^\star = s_m^\star + C_m^\star$ gives

$$A_m^\star = -q\left(1 + \frac{m}{L}\right). \tag{C.28}$$

Equivalently, after multiplication by $L(1-L)$, this identity reduces exactly to

$$Dq^2 - Bq + LD = 0.$$

Hence,

$$p_m(r) = r^2 + q\left(1+\frac{m}{L}\right)r + \frac{m}{L}q^2 \tag{C.29}$$

$$= (r+q)\left(r + \frac{m}{L}q\right). \tag{C.30}$$

The endpoint factorizations show more than feasibility. At the upper edge,

$$p_L(r) = (r-q)^2,$$

so the radius constraint is active at $+q$, while at the lower edge,

$$p_m(r) = (r+q)\left(r+\frac{m}{L}q\right),$$

so the complementary constraint is active at $-q$. For completeness, the remaining endpoint radius-q inequalities are satisfied. At both endpoints,

$$|C_\lambda^\star| \le q^2.$$

For the minus condition,

$$q^2 - qA_L^\star + C_L^\star = 0, \qquad q^2 - qA_m^\star + C_m^\star = 2q^2\left(1+\frac{m}{L}\right) > 0,$$

whereas for the plus condition,

$$q^2 + qA_L^\star + C_L^\star = 4q^2 > 0, \qquad q^2 + qA_m^\star + C_m^\star = 0.$$

By Lemma C.1, every modal root for $\lambda \in [m, L]$ lies in $|r| \le q$. Since both endpoint polynomials contain a root of modulus exactly q,

$$\sup_{\lambda\in[m,L]} \rho(M_\lambda) = q.$$

Thus, the endpoint factorization is an attainment certificate for the independently derived lower bound $q_\star$, rather than an ansatz used to guess the optimum. □

Proof of Theorem 4.1. The closed-form factor in (4.6) is (C.19) after substituting

$$m = \lambda_-, \qquad L = \lambda_+.$$

The global lower bound follows from Proposition C.5, while Proposition C.6 constructs an admissible parameter triple attaining it. Therefore

$$\inf_{\alpha,\beta\geq 0,\ \gamma>0} R(\alpha,\beta,\gamma;\lambda_-,\lambda_+) = q^\star(\lambda_-,\lambda_+).$$

The optimizer is exactly (4.7), and the endpoint factorizations above give the equioscillation identities stated in Theorem 4.1.

Finally,

$$C_1^\star = (1-\gamma^\star)\beta^\star + \gamma^\star\alpha^\star = \frac{q_\star^2}{L}.$$

Since Lemma C.4 gives $q_\star < L < 1$,

$$0 \leq C_1^\star = \frac{q_\star^2}{L} < q_\star < 1.$$

Hence, the secondary phase root is strictly stable and does not enlarge the optimized transverse factor. Combining this with Corollary B.5, the exact minimax value $q_\star$ is also the asymptotic root factor of the full linearized two-step state after quotienting the global-phase symmetry. □

C.6 Optimal FGLA and the strict hierarchy

Proposition C.7 (Exact optimal FGLA factor)**.** *Within the FGLA subfamily $\gamma = 1$ with $\alpha \geq 0$ (the parameter β is then immaterial), the minimax factor over $\lambda \in [m, L]$ is*

$$q_{\mathrm{F}} = 1 - \sqrt{1-L}, \tag{C.31}$$

attained by

$$\alpha_{\mathrm{F}}^\star = \frac{1-\sqrt{1-L}}{1+\sqrt{1-L}}. \tag{C.32}$$

Proof. For $\gamma = 1$,

$$s_\lambda = A_\lambda - C_\lambda = \lambda.$$

If the worst-case FGLA factor is at least one, it is already strictly larger than $q_{\mathrm{F}} < 1$. Otherwise, Lemma C.2 at $\lambda = L$ gives

$$q \geq q_{\mathrm{F}}.$$

Thus, q_{F} is a universal lower bound for FGLA. For attainment, set $q = q_{\mathrm{F}}$. Since

$$L = 2q - q^2 = q(2-q),$$

choose

$$\alpha_{\mathrm{F}}^\star = \frac{q}{2-q} = \frac{q^2}{L},$$

which is equivalent to (C.32). Then,

$$C_\lambda = \frac{q^2}{L}\lambda, \qquad A_\lambda = \lambda + C_\lambda.$$

At $\lambda = L$,

$$C_L = q^2, \qquad A_L = 2q, \qquad p_L(r) = (r-q)^2.$$

At $\lambda = 0$,

$$A_0 = C_0 = 0,$$

so the closed radius-q Jury conditions hold. Endpoint reduction therefore establishes feasibility on the larger interval $[0, L]$, and hence on the required interval $[m, L]$. Thus q_{F} is attained for every $0 \leq m < L$. □

Proof of Corollary 4.2. Proposition C.7 gives exactly the factor and parameter stated in (4.8). Moreover,

$$q^\star(m, L) < q_{\mathrm{F}} < L$$

by Lemma C.4. Since GLA has modal factor λ and hence worst-case transverse factor L, replacing m, L by λ_-, λ_+ gives (4.9). □

C.7 Dominant-edge-only specialization

Setting $m = 0$ gives $D = L$ and

$$B = L(3 - L).$$

Substitution into (C.19) yields

$$q^\star(0, L) = \frac{3 - L - \sqrt{(3 - L)^2 - 4L}}{2} \tag{C.33}$$

$$= \frac{3 - L - \sqrt{(1 - L)(9 - L)}}{2}. \tag{C.34}$$

This is exactly (4.10). Its strict improvement over optimal FGLA follows from (4.9).

C.8 Spectral-width limit

Fix $L \in (0, 1)$ and let

$$D = L - m \downarrow 0.$$

Since $m = L - D$,

$$B = 4L(1 - L) + (3L - 1)D. \tag{C.35}$$

Rationalizing the smaller-root formula gives

$$q_\star = \frac{2LD}{B + \sqrt{B^2 - 4LD^2}}. \tag{C.36}$$

For fixed $L < 1$, (C.35) implies

$$B + \sqrt{B^2 - 4LD^2} = 8L(1 - L) + O(D).$$

Therefore,

$$q_\star = \frac{D}{4(1 - L)} + O(D^2), \qquad D \downarrow 0. \tag{C.37}$$

This proves (4.11). In particular,

$$q_\star \to 0 \qquad \text{as } D \downarrow 0$$

for every fixed $L \in (0, 1)$. The limit is pointwise in $L < 1$: it is not uniform as $L \uparrow 1$, since the leading coefficient $1/[4(1 - L)]$ diverges. Thus, the result quantifies the local acceleration available from a tightly clustered transverse spectrum without implying uniform acceleration near phase-degenerate geometry.

D Passive Online Geometry Estimation

This appendix proves the passive non-identifiability and trajectory-active Ritz-consistency results used in Section 5. Throughout, complex ambient spaces are regarded as real Hilbert spaces with

$$\langle x, y \rangle_{\mathbb{R}} := \Re \langle x, y \rangle_{\mathbb{C}}.$$

Accordingly, secant coefficient spaces are real: whenever complex matrix notation is used below, Gram matrices and self-adjoint compressions are represented through their real parts.

D.1 Secant linearization near a regular solution

Lemma D.1 (Asymptotically exact secant action)**.** *Let G be continuously differentiable on a neighborhood $\mathcal{N}$ of $c^\star$, with*

$$DG(c^\star)|_{\mathcal{U}} = J_\star.$$

If $a_n, b_n \in \mathcal{N}$, $a_n, b_n \to c^\star$, and $s_n := a_n - b_n \neq 0$, then

$$G(a_n) - G(b_n) = J_\star s_n + r_n, \qquad \frac{\|r_n\|}{\|s_n\|} \to 0. \tag{D.1}$$

Proof. After shrinking $\mathcal{N}$ if necessary, assume it is convex. The fundamental theorem of calculus on the underlying finite-dimensional real Hilbert space gives

$$G(a_n) - G(b_n) = \int_0^1 DG\big(b_n + t(a_n - b_n)\big)s_n \, dt.$$

Hence,

$$\|G(a_n) - G(b_n) - J_\star s_n\| \le \sup_{0\le t\le 1} \big\|DG(b_n + ts_n) - J_\star\big\| \, \|s_n\|.$$

Because $a_n, b_n \to c^\star$, the segment $b_n + [0,1]s_n$ converges uniformly to $c^\star$. Continuity of DG therefore makes the displayed supremum tend to zero. Dividing by $\|s_n\|$ proves (D.1). □

In the phase-retrieval setting, Assumption 2.1 and Section A imply that $P_{\mathcal{M}_s}$, and hence

$$G = P_{\mathcal{U}} P_{\mathcal{M}_s},$$

is smooth in a sufficiently small neighborhood of $c^\star$ in which no coordinate vanishes. Under the representative-convergence assumption of Theorem 5.2, $c_n \to c^\star$. Applying Lemma D.1 with

$$a_n = c_n, \qquad b_n = c_{n-1},$$

gives

$$y_n = J_\star s_n + r_n, \qquad \frac{\|r_n\|}{\|s_n\|} \to 0, \tag{D.2}$$

which is (5.1).

D.2 Finite-data non-identifiability

Proof of Proposition 5.1. Let $\mathcal{K} \subsetneq \mathcal{H}$ be the observed subspace and choose a unit vector

$$u \in \mathcal{K}^\perp.$$

Fix any self-adjoint PSD contraction $A : \mathcal{K} \to \mathcal{K}$ and extend it by zero to $\mathcal{K}^\perp$. For distinct $a, b \in [0,1]$, define

$$J_1 := A + a\, u \otimes u, \qquad J_2 := A + b\, u \otimes u, \tag{D.3}$$

where

$$(u \otimes u)x := \langle x, u\rangle_{\mathbb{R}}\, u.$$

Both J_1 and J_2 are self-adjoint PSD contractions, and for every $s \in \mathcal{K}$,

$$J_1 s = As = J_2 s.$$

Thus, all first-order observations on $\mathcal{K}$ are identical under the two operators.

However, u is an eigenvector with eigenvalue a for J_1 and b for J_2. Hence, the spectral data on $\mathcal{K}^\perp$ can change while every observation on $\mathcal{K}$ remains unchanged. Finite first-order information restricted to a proper observed subspace therefore cannot identify the complete spectrum within the class of self-adjoint PSD contractions. □

D.3 Generalized Ritz compression

Recall

$$S_n = [s_{n-p+1}, \dots, s_n], \qquad Y_n = [y_{n-p+1}, \dots, y_n],$$

and define the real coefficient-space matrices

$$H_n := \Re(S_n^* S_n), \qquad B_n := \frac{1}{2}\Re(S_n^* Y_n + Y_n^* S_n)\,. \tag{D.4}$$

Let

$$R_n := Y_n - J_\star S_n. \tag{D.5}$$

Because $J_\star$ is self-adjoint with respect to the underlying real inner product,

$$B_n = \Re(S_n^* J_\star S_n) + \frac{1}{2}\Re(S_n^* R_n + R_n^* S_n)\,. \tag{D.6}$$

Assume $H_n \succ 0$ and define

$$Q_n := S_n H_n^{-1/2}. \tag{D.7}$$

Then, as a map from $\mathbb{R}^p$ into the real Hilbert space $\mathcal{U}$, Q_n is an isometry:

$$\Re(Q_n^* Q_n) = I_p, \qquad \operatorname{range}(Q_n) = \operatorname{range}(S_n) =: \mathcal{V}_n. \tag{D.8}$$

Define the exact real Rayleigh–Ritz compression

$$\mathcal{J}_n := \Re(Q_n^* J_\star Q_n) \in \mathbb{R}^{p\times p}. \tag{D.9}$$

Since $J_\star$ is real-self-adjoint, $\mathcal{J}_n$ is real symmetric. Multiplying (D.6) on both sides by $H_n^{-1/2}$ gives

$$H_n^{-1/2} B_n H_n^{-1/2} = \mathcal{J}_n + E_n, \tag{D.10}$$

where

$$E_n = \frac{1}{2}\Re\Big[Q_n^* R_n H_n^{-1/2} + (R_n H_n^{-1/2})^* Q_n\Big]. \tag{D.11}$$

Since Q_n is an isometry on the real coefficient space,

$$\|E_n\| \le \|R_n H_n^{-1/2}\| = \varepsilon_n. \tag{D.12}$$

Thus, the generalized Ritz matrix is an ε_n-perturbation of the exact Rayleigh–Ritz compression of $J_\star$ onto the observed secant subspace.

D.4 Subspace convergence and Ritz consistency

Lemma D.2 (Aligned orthonormal bases for convergent subspaces)**.** *Let $\mathcal{V}_n$ and $\mathcal{V}$ be p-dimensional subspaces of a finite-dimensional real Hilbert space and suppose*

$$\|P_{\mathcal{V}_n} - P_{\mathcal{V}}\| \to 0.$$

For any orthonormal basis map

$$Q : \mathbb{R}^p \to \mathcal{V},$$

there exist orthonormal basis maps

$$\widetilde{Q}_n : \mathbb{R}^p \to \mathcal{V}_n$$

such that

$$\|\widetilde{Q}_n - Q\| \to 0. \tag{D.13}$$

Moreover, every other orthonormal basis map Q_n of $\mathcal{V}_n$ satisfies

$$Q_n = \widetilde{Q}_n O_n$$

for some $O_n \in O(p)$.

Proof. Since

$$P_{\mathcal{V}_n} Q \to Q$$

in operator norm, $P_{\mathcal{V}_n} Q$ has full column rank for all sufficiently large n. Define its real polar factor by

$$\widetilde{Q}_n := P_{\mathcal{V}_n} Q \left(Q_{\mathbb{R}}^* P_{\mathcal{V}_n} Q\right)^{-1/2},$$

where $Q_{\mathbb{R}}^*$ denotes the adjoint for the underlying real Hilbert structure. Equivalently, in complex coordinates,

$$Q_{\mathbb{R}}^* P_{\mathcal{V}_n} Q = \Re(Q^* P_{\mathcal{V}_n} Q).$$

Then, $\widetilde{Q}_n$ has real-orthonormal columns spanning $\mathcal{V}_n$. Since

$$Q_{\mathbb{R}}^* P_{\mathcal{V}_n} Q \to I_p,$$

continuity of the inverse square root yields

$$\widetilde{Q}_n \to Q.$$

Finally, two real-orthonormal bases of the same p-dimensional real subspace differ by right multiplication by an orthogonal matrix. □

Lemma D.3 (Convergence of exact Ritz spectra)**.** *Under the assumptions of Lemma D.2, let $J_\star$ be fixed and self-adjoint. If $\mathcal{V}$ is $J_\star$-invariant, then the ordered eigenvalues of*

$$\mathcal{J}_n = \Re(Q_n^* J_\star Q_n)$$

converge to the ordered eigenvalues of $J_\star|_{\mathcal{V}}$.

Proof. Choose an orthonormal basis map Q for $\mathcal{V}$ and aligned bases $\widetilde{Q}_n$ from Lemma D.2. Since

$$Q_n = \widetilde{Q}_n O_n, \qquad O_n \in O(p),$$

the real symmetric matrices

$$\Re(Q_n^* J_\star Q_n) \quad \text{and} \quad \Re(\widetilde{Q}_n^* J_\star \widetilde{Q}_n)$$

are orthogonally similar and therefore have identical eigenvalues.

Let

$$\Delta_n := \widetilde{Q}_n - Q.$$

Then

$$\begin{aligned} &\Re(\widetilde{Q}_n^* J_\star \widetilde{Q}_n) - \Re(Q^* J_\star Q) \\ &\quad = \Re(\Delta_n^* J_\star Q + Q^* J_\star \Delta_n + \Delta_n^* J_\star \Delta_n)\,, \end{aligned} \tag{D.14}$$

so

$$\left\| \Re(\widetilde{Q}_n^* J_\star \widetilde{Q}_n) - \Re(Q^* J_\star Q) \right\| \le 2\|J_\star\|\|\Delta_n\| + \|J_\star\|\|\Delta_n\|^2 \to 0. \tag{D.15}$$

Since $\mathcal{V}$ is $J_\star$-invariant,

$$\Re(Q^* J_\star Q)$$

is the matrix representation of $J_\star|_{\mathcal{V}}$ in the real orthonormal basis Q. Weyl's inequality therefore gives convergence of all ordered eigenvalues. □

Proof of Theorem 5.2. The generalized eigenvalues of (B_n, H_n) are the ordinary eigenvalues of the real symmetric matrix

$$\widehat{J}_n := H_n^{-1/2} B_n H_n^{-1/2}. \tag{D.16}$$

By (D.10) and (D.12),

$$\|\widehat{J}_n - \mathcal{J}_n\| \le \varepsilon_n \to 0.$$

By Lemma D.3, the ordered eigenvalues of $\mathcal{J}_n$ converge to those of $J_\star|_{\mathcal{V}}$. A second application of Weyl's inequality therefore gives, for $j = 1, \dots, p$,

$$|\widehat{\mu}_{j,n} - \mu_j| \le \varepsilon_n + |\lambda_j(\mathcal{J}_n) - \mu_j| \longrightarrow 0. \tag{D.17}$$

If the limiting active subspace contains a dominant transverse eigendirection, equivalently

$$\lambda_{\max}(J_\star|_{\mathcal{V}}) = \lambda_+, \tag{D.18}$$

then the largest generalized Ritz value satisfies

$$\widehat{\lambda}_{+,n} = \widehat{\mu}_{p,n} \longrightarrow \lambda_+,$$

which proves (5.4). □

D.5 A SUFFICIENT NORMALIZED-SECANT CONDITION

The consistency theorem assumes the scale-free residual condition

$$\|(Y_n - J_\star S_n) H_n^{-1/2}\| \to 0. \tag{D.19}$$

This assumption is deliberately stronger than the columnwise secant expansion (D.2): as the raw secants vanish, near-linear dependence can amplify their residuals after normalization.

The following condition is sufficient.

Lemma D.4 (Uniformly conditioned normalized secants imply vanishing Ritz residual)**.** *Assume every column of S_n is nonzero and write*

$$S_n = \bar{S}_n D_n, \qquad D_n = \operatorname{diag}\big(\|s_{n-p+1}\|, \dots, \|s_n\|\big), \tag{D.20}$$

so that every column of $\bar{S}_n$ has unit norm. Suppose

$$\sigma_{\min}(\bar{S}_n) \geq \sigma_0 > 0 \tag{D.21}$$

eventually, where the singular value is taken for the corresponding real-linear map. If

$$\max_{j=n-p+1,\dots,n} \frac{\|y_j - J_\star s_j\|}{\|s_j\|} \longrightarrow 0, \tag{D.22}$$

then

$$\|(Y_n - J_\star S_n) H_n^{-1/2}\| \to 0.$$

Proof. Let

$$R_n := Y_n - J_\star S_n, \qquad \bar{R}_n := R_n D_n^{-1}.$$

Since p is fixed, (D.22) implies

$$\|\bar{R}_n\| \to 0.$$

Interpret $S_n : \mathbb{R}^p \to \mathcal{U}$ as the corresponding injective real-linear map. Its real Gram matrix is H_n, so

$$Q_n = S_n H_n^{-1/2}$$

is its polar factor. Hence,

$$S_n = Q_n H_n^{1/2}, \qquad S_n^\dagger = H_n^{-1/2} Q_n^{*_{\mathbb{R}}},$$

and therefore

$$H_n^{-1/2} = S_n^\dagger Q_n. \tag{D.23}$$

Because

$$S_n = \bar{S}_n D_n$$

with invertible diagonal D_n and full column rank,

$$S_n^\dagger = D_n^{-1} \bar{S}_n^\dagger.$$

Consequently,

$$R_n H_n^{-1/2} = R_n S_n^\dagger Q_n = \bar{R}_n \bar{S}_n^\dagger Q_n.$$

Since Q_n is a real isometry,

$$\|R_n H_n^{-1/2}\| \leq \|\bar{R}_n\| \, \|\bar{S}_n^\dagger\| \leq \sigma_0^{-1} \|\bar{R}_n\| \to 0.$$

□

E Proof of Online Asymptotic Rate Recovery

This appendix proves Theorem 6.1. Two points are essential. First, the online iteration is genuinely nonautonomous: t_n uses the current relaxation γ_n, whereas c_{n-1} and d_{n-1} contain the previous inertial parameters. Second, the solution is a global-phase orbit rather than an isolated fixed point. We therefore analyze the induced two-step dynamics on a local quotient by the diagonal phase action, removing the neutral phase root while retaining the secondary phase-state root.

E.1 Continuity of the conservative parameter map

For $\lambda \in (0,1)$, define

$$q_0(\lambda) := q^\star(0,\lambda) = \frac{3-\lambda-\sqrt{(1-\lambda)(9-\lambda)}}{2}. \tag{E.1}$$

The conservative geometry-optimal parameters are

$$\beta_0(\lambda) = 0, \qquad \gamma_0(\lambda) = \frac{(1-q_0(\lambda))^2}{1-\lambda}, \qquad \alpha_0(\lambda) = \frac{q_0(\lambda)^2(1-\lambda)}{\lambda(1-q_0(\lambda))^2}. \tag{E.2}$$

Lemma E.1 (Continuity of the online parameter map)**.** *The map*

$$\Theta(0,\lambda) := (\alpha_0(\lambda), 0, \gamma_0(\lambda))$$

is continuous on $(0,1)$*. Hence, if*

$$\widehat{\lambda}_{+,n} \to \lambda_+ \in (0,1),$$

then

$$\vartheta_n := \Theta(0, \widehat{\lambda}_{+,n}) \longrightarrow \vartheta_\star := \Theta(0,\lambda_+). \tag{E.3}$$

In particular, the parameter sequence is eventually contained in a compact subset of the local parameter domain.

Proof. The radicand

$$(1-\lambda)(9-\lambda)$$

is strictly positive on $(0,1)$, so q_0 is continuous there. Moreover, the strict hierarchy (4.9), specialized to $[0,\lambda]$, gives

$$0 < q_0(\lambda) < \lambda < 1.$$

Hence, the denominators in (E.2) are nonzero throughout $(0,1)$. Continuity of Θ follows, and (E.3) follows by composition. □

E.2 Variable-parameter linearization

Let

$$t_n = (1-\gamma_n)d_{n-1} + \gamma_n G(c_{n-1}), \tag{E.4}$$
$$c_n = t_n + \alpha_n(t_n - t_{n-1}), \tag{E.5}$$
$$d_n = t_n + \beta_n(t_n - t_{n-1}). \tag{E.6}$$

For a representative $c^\star \in \mathcal{O}^\star$, set

$$e_n := t_n - c^\star.$$

Then,

$$c_{n-1} - c^\star = (1+\alpha_{n-1})e_{n-1} - \alpha_{n-1}e_{n-2}, \tag{E.7}$$
$$d_{n-1} - c^\star = (1+\beta_{n-1})e_{n-1} - \beta_{n-1}e_{n-2}. \tag{E.8}$$

Therefore the first-order two-step operator is

$$\mathbb{A}_n = \begin{pmatrix} \mathsf{A}_n & -\mathsf{C}_n \\ I & 0 \end{pmatrix}, \tag{E.9}$$

where

$$\mathsf{A}_n = (1-\gamma_n)(1+\beta_{n-1})I + \gamma_n(1+\alpha_{n-1})J_\star, \tag{E.10}$$
$$\mathsf{C}_n = (1-\gamma_n)\beta_{n-1}I + \gamma_n\alpha_{n-1}J_\star. \tag{E.11}$$

The index shift in α_{n-1} and β_{n-1} is essential: the current update uses γ_n, but the stored states c_{n-1} and d_{n-1} were formed using the previous inertial parameters.

Let

$$\zeta_\star := (ic^\star, ic^\star) \tag{E.12}$$

be the diagonal phase tangent at $(c^\star, c^\star)$.

Lemma E.2 (Nonautonomous quotient expansion)**.** *Under Assumption 2.1 and* (E.3)*, there exist* N_{par} *and a smooth local quotient chart around* $\mathcal{X}^\star$ *such that, for every* $n \geq N_{\mathrm{par}}$*,*

$$\overline{E}_n = \overline{\mathbb{A}}_n \overline{E}_{n-1} + \overline{R}_n(\overline{E}_{n-1}), \tag{E.13}$$

where $\overline{\mathbb{A}}_n$ *is induced by* $\mathbb{A}_n$ *on*

$$(\mathcal{U} \times \mathcal{U})/\operatorname{span}_{\mathbb{R}}\{\zeta_\star\},$$

and

$$\sup_{n \geq N_{\mathrm{par}}} \frac{\|\overline{R}_n(\overline{E})\|}{\|\overline{E}\|} \longrightarrow 0 \qquad \text{as } \|\overline{E}\| \to 0. \tag{E.14}$$

Moreover,

$$\|\overline{\mathbb{A}}_n - \overline{\mathbb{A}}_\star\| \to 0, \tag{E.15}$$

where $\overline{\mathbb{A}}_\star$ *is the quotient derivative associated with*

$$\vartheta_\star = \Theta(0, \lambda_+).$$

Proof. Let F_ϑ denote the two-step state map associated with a fixed parameter triple $\vartheta = (\alpha, \beta, \gamma)$. Every F_ϑ is global-phase equivariant and leaves the diagonal solution orbit $\mathcal{X}^\star$ pointwise fixed. Its derivative at $(c^\star, c^\star)$, evaluated at the parameters used at iteration n, is precisely (E.9). Since

$$J_\star(ic^\star) = ic^\star,$$

direct substitution gives

$$\mathbb{A}_n \zeta_\star = \zeta_\star.$$

Hence, the orbit tangent is invariant and $\mathbb{A}_n$ induces a quotient operator $\overline{\mathbb{A}}_n$. Choose a smooth local slice for the diagonal S^1 action near $\mathcal{X}^\star$. Phase equivariance induces a smooth quotient map on this slice whose derivative at the orbit class is $\overline{\mathbb{A}}_n$.

By Lemma E.1, the parameters eventually lie in a compact neighborhood of $\vartheta_\star$. The state map depends smoothly on the parameters and, by regularity, G is C^1 on a sufficiently small neighborhood of the solution orbit. Uniform continuity of the state derivative on the resulting compact state–parameter neighborhood gives the uniform first-order estimate (E.14).

Finally,

$$\alpha_{n-1} \to \alpha_\star, \qquad \beta_{n-1} \to \beta_\star, \qquad \gamma_n \to \gamma_\star,$$

so

$$\mathbb{A}_n \to \mathbb{A}_\star.$$

All these operators preserve the same orbit tangent, hence passage to the fixed quotient is continuous and gives (E.15). □

E.3 Spectrum of the limiting quotient operator

The diagonal phase action is

$$e^{i\phi} \cdot (t, t^-) = (e^{i\phi} t, e^{i\phi} t^-),$$

with tangent $\zeta_\star$ at $(c^\star, c^\star)$. Define

$$\overline{\mathcal{H}} := (\mathcal{U} \times \mathcal{U})/\operatorname{span}_{\mathbb{R}}\{\zeta_\star\}.$$

Lemma E.3 (Limiting quotient spectral radius)**.** *For*

$$\vartheta_\star = \Theta(0, \lambda_+),$$

the limiting quotient operator satisfies

$$\rho(\overline{\mathbb{A}}_\star) \leq q_0(\lambda_+) = q^\star(0, \lambda_+) < 1. \tag{E.16}$$

Every transverse eigenvalue

$$\lambda \in \operatorname{spec}(J_\star|_{\mathcal{N}_\star}) \subset [0, \lambda_+]$$

contributes a companion block whose roots have modulus at most $q_0(\lambda_+)$*. The phase block has roots*

$$1 \qquad \text{and} \qquad C_1^\star = \frac{q_0(\lambda_+)^2}{\lambda_+},$$

with

$$0 < C_1^\star < q_0(\lambda_+) < 1. \tag{E.17}$$

The quotient removes precisely the neutral root 1*.*

Proof. The transverse bound is the exact minimax guarantee of Theorem 4.1 applied to the conservative interval $[0, \lambda_+]$. For the phase eigenvalue $\lambda = 1$, Proposition B.4 gives

$$p_1(r) = (r-1)(r - C_1^\star).$$

At the conservative optimizer,

$$\beta^\star = 0, \qquad C_1^\star = \gamma^\star \alpha^\star = \frac{q_0(\lambda_+)^2}{\lambda_+}.$$

Since

$$0 < q_0(\lambda_+) < \lambda_+ < 1,$$

we have

$$0 < C_1^\star < q_0(\lambda_+) < 1.$$

The diagonal phase tangent is the eigenvector associated with the unit root. Quotienting by its span removes that neutral direction, while the secondary phase-state root and all transverse roots remain. Therefore

$$\rho(\overline{\mathbb{A}}_\star) \le q_0(\lambda_+) < 1.$$

□

E.4 Adapted-norm contraction

Lemma E.4 (Spectral radius and an adapted norm)**.** *Let A be a linear map on a finite-dimensional real vector space. For every $r > \rho(A)$ there exists a norm $\|\cdot\|_\dagger$, equivalent to the Euclidean norm, such that*

$$\|A\|_\dagger < r.$$

Proof. Complexify the space and put the complexified operator into Jordan form. Within each Jordan block, rescale the generalized eigenvectors geometrically so that the contribution of each superdiagonal entry is arbitrarily small. Since every eigenvalue has modulus strictly below r, the induced max norm can be chosen so that every block has operator norm strictly below r. Restricting this norm to the real invariant subspace gives the desired equivalent real norm. □

Fix $\varepsilon > 0$ with

$$q_0(\lambda_+) + \varepsilon < 1.$$

By Lemmas E.3 and E.4, choose a quotient norm $\|\cdot\|_{\dagger,\varepsilon}$ satisfying

$$\|\overline{\mathbb{A}}_\star\|_{\dagger,\varepsilon} < q_0(\lambda_+) + \frac{\varepsilon}{3}.$$

Because all norms are equivalent in finite dimension, (E.15) implies that, for all sufficiently large n,

$$\|\overline{\mathbb{A}}_n\|_{\dagger,\varepsilon} \le q_0(\lambda_+) + \frac{2\varepsilon}{3}.$$

Likewise, the uniform remainder estimate implies that, after shrinking the quotient neighborhood if necessary,

$$\|\overline{R}_n(\overline{E})\|_{\dagger,\varepsilon} \le \frac{\varepsilon}{3}\|\overline{E}\|_{\dagger,\varepsilon}$$

uniformly for all sufficiently large n. Hence, within that neighborhood,

$$\|\overline{E}_n\|_{\dagger,\varepsilon} \le \big(q_0(\lambda_+) + \varepsilon\big)\|\overline{E}_{n-1}\|_{\dagger,\varepsilon}. \tag{E.18}$$

E.5 Quotient distance and proof of the theorem

Lemma E.5 (Local quotient-distance equivalence)**.** *There exist a neighborhood $\mathcal{W}_0$ of $\mathcal{X}^\star$ and constants $0 < c_1 \le c_2 < \infty$ such that every $X \in \mathcal{W}_0$ admits a smooth local phase alignment with quotient coordinate $\overline{E}(X)$ satisfying*

$$c_1\|\overline{E}(X)\|_\dagger \le \operatorname{dist}(X, \mathcal{X}^\star) \le c_2\|\overline{E}(X)\|_\dagger \tag{E.19}$$

for any fixed equivalent quotient norm $\|\cdot\|_\dagger$.

Proof. The diagonal S^1 action is smooth and free at the nonzero state $(c^\star, c^\star)$. Hence a sufficiently small tubular neighborhood of the orbit admits a smooth slice transverse to $\zeta_\star$. The slice provides a smooth local quotient chart with invertible transverse derivative.

After shrinking the neighborhood, the Euclidean norm of the slice coordinate is therefore locally bi-Lipschitz equivalent to Euclidean distance from the orbit. Any fixed quotient norm is equivalent to the Euclidean norm in finite dimension, yielding (E.19). □

Proof of Theorem 6.1. Write

$$q_0 := q_0(\lambda_+) = q^\star(0, \lambda_+).$$

We separate local convergence from recovery of the sharp asymptotic factor.

Step 1: local orbital convergence. Choose the fixed margin

$$\varepsilon_0 := \frac{1 - q_0}{2} > 0.$$

Then

$$q_0 + \varepsilon_0 = \frac{1 + q_0}{2} < 1.$$

Applying the adapted-norm argument above with ε_0 gives a time N_0 and a sufficiently small regular quotient neighborhood $\mathcal{W}$ such that, for every $n \geq N_0$,

$$\|\overline{E}_n\|_{\dagger,0} \leq (q_0 + \varepsilon_0)\|\overline{E}_{n-1}\|_{\dagger,0} \tag{E.20}$$

whenever $X_{n-1} \in \mathcal{W}$.

Shrink $\mathcal{W}$ if necessary so that the corresponding quotient ball is forward invariant under this contraction. By the sufficiently-late local-entry assumption in Theorem 6.1, the realized trajectory enters $\mathcal{W}$ at some

$$N_{\text{in}} \geq N_0.$$

It then remains in $\mathcal{W}$, and iterating (E.20) yields

$$\overline{E}_n \to 0.$$

By Lemma E.5,

$$\operatorname{dist}(X_n, \mathcal{X}^\star) \to 0.$$

Thus, the nonautonomous dynamics converge orbitally.

Step 2: recovery of the limiting root factor. Let $\varepsilon > 0$ be arbitrary with

$$q_0 + \varepsilon < 1.$$

The adapted-norm construction gives a corresponding norm $\|\cdot\|_{\dagger,\varepsilon}$, a time N_ε, and a local neighborhood $\mathcal{W}_\varepsilon$ such that

$$\|\overline{E}_n\|_{\dagger,\varepsilon} \leq (q_0 + \varepsilon)\|\overline{E}_{n-1}\|_{\dagger,\varepsilon}$$

for $n \geq N_\varepsilon$ whenever the state lies in $\mathcal{W}_\varepsilon$.

Step 1 has already established convergence of this same realized trajectory to $\mathcal{X}^\star$. Hence there exists

$$K_\varepsilon \geq N_\varepsilon$$

after which the trajectory remains inside $\mathcal{W}_\varepsilon$. Therefore

$$\|\overline{E}_n\|_{\dagger,\varepsilon} \leq C_\varepsilon (q_0 + \varepsilon)^{\,n - K_\varepsilon}$$

for some finite C_ε. Local quotient-distance equivalence then gives

$$\limsup_{n\to\infty} \operatorname{dist}(X_n, \mathcal{X}^\star)^{1/n} \leq q_0 + \varepsilon.$$

Because $\varepsilon > 0$ is arbitrary,

$$\boxed{\limsup_{n\to\infty} \operatorname{dist}(X_n, \mathcal{X}^\star)^{1/n} \leq q_0(\lambda_+) = q^\star(0, \lambda_+).} \tag{E.21}$$

Finally, (4.9) gives

$$q_0(\lambda_+) < 1 - \sqrt{1 - \lambda_+},$$

which is the strict comparison with optimally tuned FGLA. □

Why no estimator convergence rate is required. The proof uses only

$$\vartheta_n \to \vartheta_\star,$$

not a quantitative rate for this convergence. Slow stabilization of $\widehat{\lambda}_{+,n}$ can therefore postpone the finite iteration at which the asymptotic regime becomes visible, but it does not alter the limiting root-rate bound once dominant-edge consistency and sufficiently late local entry hold.

Logical scope. The result is deliberately conditional. It does not prove that the adaptive trajectory reaches the regular local neighborhood, nor that passive Ritz estimation becomes consistent before such entry. Rather, conditional on dominant-edge consistency and sufficiently late local entry, it shows that the resulting nonautonomous iteration inherits the conservative geometry-designed asymptotic root factor.

F Experimental Details and Audited Evidence

This appendix provides the frozen protocols and audited evidence supporting the empirical claims in Section 7. Populations are reported separately, failed or irregular cases remain in the accounting, and finite-window contraction estimates are distinguished from the asymptotic root factors of the local theory.

F.1 Reproducibility and reporting contract

All experiments are executed from version-controlled command-line scripts and produce immutable JSON/CSV ledgers recording configurations, seeds, and raw outputs. Numerical tolerances, fitting windows, success criteria, estimator checks, and fallback rules are frozen in configuration files; failed, irregular, rejected, and numerically invalid cases remain explicitly recorded. Publication figures are generated programmatically from these ledgers. Throughout, convergence success and fitted contraction rate are distinct finite-horizon quantities and are not identified with the asymptotic root factor of the local linearization.

F.2 Geometry–spectrum and local-acceleration validation

For each regular instance, we independently compute the transverse spectrum of $J_\star$, its principal-angle prediction, the endpoints $[\lambda_-, \lambda_+]$, and the corresponding GLA, optimal-FGLA, and Geometry-Oracle factors. Phase-mode multiplicity and residuals are audited separately.

Contraction factors are estimated on frozen windows without post-hoc window selection. The theoretical local-linearization ordering

$$q^\star(\lambda_-, \lambda_+) < q_{\mathrm{F}}^\star < \lambda_+ \tag{F.1}$$

is asymptotic; finite-window violations are therefore retained unless an independent diagnostic identifies an implementation error.

On the frozen 202-file theory-validation population, principal-angle geometry predicts the dominant transverse Jacobian edge with median relative discrepancy 1.4385×10^{-3} and maximum 2.8971×10^{-3}. At $H = 360$, the predicted GLA/FGLA/Geometry-Oracle ordering holds for 173/202 files, while FGLA and the Geometry Oracle reduce median final error relative to GLA by $5.086\times$ and $8.967\times$, respectively.

Retaining and extending all 29 ordering violations yields agreement for 188/202 files at $H = 720$ and 186/202 at $H = 1440$. Thus the predicted hierarchy becomes more visible at $H = 720$ and remains high at $H = 1440$, without supporting a monotonic-horizon interpretation.

F.3 Frozen online protocol

Online synthesizes parameters from the conservative interval $[0, \widehat{\lambda}_{+,n}]$ through $\Theta(0, \widehat{\lambda}_{+,n})$, using the frozen settings

$$p = 3, \qquad K = 8, \qquad \Delta = 0, \tag{F.2}$$

where p is secant memory, K the update period, and Δ the additional spectral margin. A fixed running-envelope rule is applied throughout, with GLA fallback whenever estimation or parameter synthesis fails its numerical admissibility checks. No safeguard is adapted to the signal, evaluation population, or observed performance.

F.4 Recorded-audio robustness protocol

Real-audio evaluations use MUSAN recordings (Snyder et al., 2015) with deterministic preprocessing: mono conversion, resampling when required, mean-centering, and frozen excerpt selection. No file is manually modified based on algorithmic performance.

We use two frozen STFT profiles,

$$\text{Gaussian: } n_{\text{fft}} = 256,\ \text{hop} = 32, \qquad \text{Hann: } n_{\text{fft}} = 512,\ \text{hop} = 128,$$

with six initialization settings and two random seeds per valid file–profile case.

Success is recomputed from stored primitive quantities and requires

$$e_{\text{final}} \le 0.1\, e_{\text{initial}}, \qquad r_{\text{mag}}^{\text{final}} \le 10^{-5}. \tag{F.3}$$

We report the resulting frequency strictly as empirical convergence success under the frozen protocol, not as a theoretical basin size or global convergence probability. Near-degenerate cases remain in the evaluation and are not removed based on finite-horizon performance.

F.5 Original and disjoint held-out robustness evaluations

We evaluate the frozen protocol first on the original 201-file manifest and then, without retuning, on a disjoint 202-file MUSAN manifest with zero filename overlap. After regularity and numerical screening, the two populations yield 390 and 396 valid file–STFT-profile cases, corresponding to 4,680 and 4,752 runs, respectively. Success labels recomputed from (F.3) agree with the stored outcomes.

Online improves convergence success over oracle-tuned FGLA by 20.07 percentage points on the original population and 18.39 points on the disjoint held-out population. Source-file cluster-bootstrap 95% intervals are $[18.6, 21.5]$ and $[17.0, 19.8]$ percentage points, respectively.

Among jointly successful runs, Online attains the smaller fitted contraction factor in only 34.47% and 35.74% of pairs, with median $q_{\text{Online}}/q_{\text{FGLA}}$ equal to 1.00093 and 1.00091, respectively. No safeguard failure occurs in either evaluation. Thus the disjoint held-out population reproduces the original robustness–rate separation: substantially higher convergence success without an average paired finite-window local-rate advantage.

F.6 Contemporary-baseline closure

We compare Online with literature-configured NAGLA and AGLA on the same disjoint 4,752-run population, using identical run keys, initializations, seeds, success criteria, and rate-fitting procedures. NAGLA uses

$$\alpha_i = \frac{i + H - 2}{i + H + 1}, \qquad H = 49,$$

following Cicognani et al. (2026), while AGLA uses the published Table-II parameters

$$(\alpha, \beta, \gamma) = (1.05, 1.35, 1.25)$$

of Nenov et al. (2024). Neither baseline is retuned on MUSAN.

Online achieves 73.30% convergence success, compared with 54.90% for optimal FGLA, 52.06% for NAGLA, and 47.62% for AGLA. The paired Online gains over NAGLA and AGLA are 21.23 and 25.67 percentage points, with source-file cluster-bootstrap 95% intervals $[19.64, 22.57]$ and $[24.01, 27.06]$, respectively.

The success advantage does not translate into faster conditional local contraction. Among jointly successful runs, median $q_{\text{Online}}/q_{\text{baseline}}$ is 1.0003 against NAGLA and 1.0005 against AGLA,

Table 1: **Real-audio robustness across three evaluations.** Evaluations are reported separately; MUSAN-25 is a larger-scale replication, not an additional disjoint held-out population.

	Original	Disjoint	MUSAN-25
Manifest files	201	202	504
Basin runs	4,680	4,752	11,856
Online success	74.94%	73.30%	74.86%
FGLA success	54.87%	54.90%	55.04%
Difference	20.07 pp	18.39 pp	19.82 pp
Paired beat fraction	34.47%	35.74%	33.55%
Median $q_{\mathrm{Online}}/q_{\mathrm{FGLA}}$	1.00093	1.00091	1.00098
Safeguard failures	0/4,680	0/4,752	0/11,856

with Online attaining the smaller fitted factor in 44.3% and 40.6% of pairs, respectively. Thus the comparison again exhibits the robustness–rate separation observed with FGLA.

These results apply to the stated literature-transferred configurations under the frozen protocol; they do not claim superiority over all possible NAGLA or AGLA tunings.

F.7 Larger-scale MUSAN-25 replication

We further evaluate the frozen protocol on a stratified 25% MUSAN subset (seed 42) containing 504 unique WAV files. This is a larger-scale replication, not an additional independent held-out partition; overlap with the earlier manifests is not excluded.

Across two STFT profiles, 988 of 1,008 file–profile cases are valid, with 17 numerical geometry failures and three irregular cases retained in the accounting. On valid cases, the dominant-edge discrepancy has median 9.13×10^{-4} and maximum 2.00×10^{-3}.

The resulting 11,856 initialization runs yield 74.86% Online success versus 55.04% for FGLA, a 19.82 percentage-point difference. Among 6,489 jointly successful runs, Online has the smaller fitted contraction factor in 33.55% of cases, with

$$\operatorname{median} \frac{q_{\mathrm{Online}}}{q_{\mathrm{FGLA}}} = 1.00098, \qquad Q_{0.90}\left(\frac{q_{\mathrm{Online}}}{q_{\mathrm{FGLA}}}\right) = 1.00354.$$

No safeguard failure occurs in the 11,856 runs. Thus the larger-scale evaluation reproduces the robustness–rate separation: Online substantially increases convergence success while conditional finite-window contraction rates remain closely matched.

F.8 Mechanism ablation

We use a separate frozen common-population ablation to determine which components of Online are associated with the observed robustness gain. The study compares six variants under identical run keys and evaluation rules: oracle-tuned FGLA, a fixed one-edge oracle, Raw adaptive spectral control, the running-envelope variant, the fallback variant, and Full Online. The corresponding empirical convergence-success rates are

FGLA oracle	30.17%,
fixed one-edge oracle	16.52%,
Raw adaptive	55.61%,
Envelope	50.46%,
Fallback	55.65%,
Full Online	56.29%.

This ablation branch has its own frozen population and protocol, so these absolute percentages are not directly comparable with those of the original, disjoint held-out, or MUSAN-25 robustness evaluations.

Two contrasts are most informative. First, Full Online exceeds FGLA by 26.12 percentage points under this common protocol. Second, Raw adaptive spectral control exceeds the fixed one-edge

comparator by 39.20 points. The latter comparison isolates the role of trajectory adaptation more directly: both variants use the same conservative one-edge design principle, but only Raw adaptive updates the spectral information from the observed trajectory. The much larger success rate of Raw adaptive is therefore consistent with the hypothesis that trajectory-dependent spectral estimation and control are central to the robustness gain.

The safeguard variants provide a different picture. Adding the running envelope or fallback produces substantially smaller changes than the Raw-adaptive versus fixed-one-edge contrast, and the effects are interacting rather than additive. We therefore do not assign an independent causal contribution to each safeguard. Under this frozen ablation protocol, the evidence supports the more limited conclusion that the dominant empirical robustness gain is *associated with* trajectory-adaptive spectral estimation and control, while the envelope and fallback act as secondary safeguards.

As a forensic implementation check, the ablation branch additionally passes 18/18 trajectory-equivalence checks, 12/12 no-oracle-leakage checks, 18,960/18,960 deterministic reruns, 37/37 automated tests, and 12/12 resume-consistency checks. These audits support the implementation provenance of the ablation but are not themselves evidence of a causal decomposition.