RESIDUAL-BUDGETED INEXACT CONTOUR MOMENTS FOR NONLINEAR EIGENVALUE PROBLEMS

YUQI LIU*, YAOYAO ZHAN†, JOSE E. ROMAN‡, AND MEIYUE SHAO§,¶

Abstract. Contour-integral eigensolvers for analytic nonlinear eigenvalue problems require the solution of a linear system at each quadrature node and for every probing direction. Infinite GMRES amortizes this cost across nodes; however, the contour-NEP implementation of Liu, Roman, and Shao prescribes basis lengths a priori and advocates highly accurate node solves as the quadrature is refined. We develop an accuracy-control framework tailored to centered Beyn extraction. For block analytic NEPs, the residual-to-moment relation yields a computable bound, conditional on certified upper bounds for the resolvent norm, together with a dual-weighted identity, a direction-sensitive Keldysh subspace estimate, and a structured perturbation enclosure for the reduced Beyn matrix. The associated weights encode the quadrature rule, the local resolvent, and the extraction objective, and are therefore not equivalent to a maximum-residual criterion over the parameter domain. A nested trapezoidal difference provides a practical scale for the algebraic error, while an additional singular-subspace guard prevents premature termination when the extraction is sensitive. Expansion-point clusters incorporate the coupling among nodes that share an infinite-GMRES basis. Experiments on a nonnormal analytic family and two NLEVP problems confirm the residual bounds numerically. A paired-seed sweep over 13 nonnormality levels shows that fixed prefix work becomes unreliable as eigenvector conditioning grows, whereas residual matching keeps the observed solve-induced subspace error small. A 7×8 stopping-parameter sweep exposes the loaded-string transition from plateau accuracy to premature stopping, and a contribution decomposition shows that one of eight clusters carries 83.8–90.9 percent of the evaluated bound. In retrospective prefix replay, the nominal safeguarded criterion preserves the accuracy plateau attained by fixed iteration counts while reducing the counted Arnoldi-prefix work in every tested configuration.

Key words. nonlinear eigenvalue problem, contour integral, Beyn method, infinite GMRES, inexact linear solve, a posteriori error estimate

AMS subject classifications. 65F15, 65F10, 65H17, 65D30

1. Introduction.

Let $T : \mathcal{D} \to \mathbb{C}^{n\times n}$ be analytic on an open set $\mathcal{D} \subset \mathbb{C}$. The nonlinear eigenvalue problem (NEP) seeks $(\lambda, v) \in \mathcal{D} \times (\mathbb{C}^n \setminus \{0\})$ satisfying

$$T(\lambda)v = 0. \tag{1}$$

Many applications require all eigenvalues within a prescribed bounded region, rather than only those nearest a selected shift. Contour methods address this requirement by sampling the resolvent $T(z)^{-1}$ along the region boundary and compressing its meromorphic information into moments. Beyn's method uses only the first two moments to recover the enclosed simple eigenvalues from a small dense eigenproblem [2]. More general interpretations relate these algorithms to rational interpolation and system realization [3], while [10] situates them within the broader numerical analysis of NEPs.

The dominant computational cost arises from the node systems. Given quadrature nodes z_j and a probing block Z, one must solve

$$T(z_j)X_j = Z, \qquad j = 0, \ldots, N-1. \tag{2}$$

*School of Mathematical Sciences, Fudan University, Shanghai 200433, China.

†Shanghai University of Finance and Economics, Shanghai 200433, China.

‡D. Sistemes Informàtics i Computació, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain.

§School of Data Science, Fudan University, Shanghai 200433, China.

¶Shanghai Key Laboratory for Contemporary Applied Mathematics, Fudan University, Shanghai 200433, China.

Factoring every matrix $T(z_j)$ may be prohibitively expensive. Infinite GMRES replaces these factorizations with a Taylor linearization and a Krylov basis shared across multiple parameter values [16]. Preconditioning and inexact applications of the expansion-point inverse were analyzed in [6]. Liu, Roman, and Shao subsequently combined weighting, multiple expansion points, and a TOAR-type representation to deploy infinite GMRES effectively within contour NEP solvers [20]. Remark 6 of that work identifies an unresolved question: what node-solve accuracy is required to achieve a prescribed quadrature accuracy? In the absence of an accuracy-allocation rule, their implementation recommends solving to machine precision.

The requisite accuracy cannot be inferred from a single residual criterion. A small companion least-squares residual does not necessarily imply a small physical residual for (2). Although the maximum physical residual over the contour is more conservative, it neglects both cancellation and the unequal contributions of individual nodes to the moments. When several nodes share an Arnoldi basis, nodewise tolerances do not translate directly into computational work. The effect of the resulting moment perturbations also depends on the smallest retained singular value, the interaction between the probing block and the left eigenvectors, and the conditioning of the reduced extraction problem.

Several related algorithms address individual parts of this accuracy chain. For linear problems, IFEAST uses a resolvent-weighted sum of shifted-system residuals in its outer convergence analysis [9]; earlier FEAST studies likewise examined stopping criteria and inner-system accuracy [18, 23]. Flexible moment subspace iteration switches between inner and outer iterations, supports single- and multiple-moment strategies, and includes an adaptive quadrature variant [12]. Error analyses for Rayleigh–Ritz contour methods and for analytic or polynomial NEPs are given in [13, 5, 17]. Verified complex-moment computations for Hermitian generalized problems further account for quadrature, solve, and rounding errors [14, 15]. Recent rational Krylov methods adapt poles or shifts from residual information over a parameter domain [7, 4]. Thus, the present work does not claim priority for residual-aware or inexact contour eigensolvers in general. It instead addresses a specific unresolved setting: centered Beyn moments, the propagation of their perturbations to nonlinear eigendata, and work allocation across expansion-point clusters whose nodes share a basis.

The principal contributions are as follows.

1. We extend the resolvent-weighted inexact-solve relation underlying IFEAST to block analytic NEPs and centered Beyn moments. The resulting norm bound separates estimates based on certified resolvent information from practical indicators, while the dual formulation identifies problem-dependent accuracy goals.
2. We establish two complementary propagation results for moment errors. A graph representation gives a direction-sensitive angle bound for the Keldysh subspace. In parallel, an exact identity for the compressed operator $K_L = (M_0 L)^\dagger (\overline{M}_1 L)$ yields a structured, residual-goal-weighted Bauer–Fike enclosure for the Beyn eigenvalues and accommodates oversampled probing blocks.
3. We combine a certified strip-analyticity bound, when available, with a nested-rule balancing test that reuses the fine-grid node solutions. We then formulate tolerance allocation under a logarithmic work model and adapt the allocation to expansion-point clusters sharing infinite-GMRES bases.
4. We provide a reproducible reference implementation that evaluates physical, rather than companion, residuals at every checkpoint. A paired-seed nonnormality study, a two-parameter stopping sweep, and a node–cluster contribu-

tion decomposition distinguish the safe algebraic bounds from the modeling assumption implicit in a one-level nested quadrature indicator and quantify the retrospective prefix-work reductions obtained with safeguarded stopping.

A small final eigenpair residual does not prove that the contour extraction has captured every enclosed eigenvalue. Completeness continues to depend on the contour-moment rank, the quadrature resolution, and the regularity assumptions underlying Keldysh extraction. Similarly, an estimated resolvent norm produces an estimator rather than a certificate. These distinctions are maintained throughout the analysis and experiments.

2. Beyn moments and centered scaling. Assume that T is regular, that $\overline{\Omega} \subset \mathcal{D}$, and that $\Gamma = \partial\Omega$ is a positively oriented smooth Jordan curve on which $T(z)$ is nonsingular. The main analysis considers a region Ω containing exactly r distinct simple eigenvalues $\lambda_1, \ldots, \lambda_r$. We further assume $r \leq n$ and linear independence of the corresponding normalized right eigenvectors. For a probing block $Z \in \mathbb{C}^{n\times s}$, define

$$M_0 = \frac{1}{2\pi \mathrm{i}} \int_\Gamma T(z)^{-1} Z \, dz, \qquad M_1 = \frac{1}{2\pi \mathrm{i}} \int_\Gamma z T(z)^{-1} Z \, dz. \tag{3}$$

If $s \geq r$, $\operatorname{rank}(M_0) = r$, and the probing block satisfies the nondegeneracy condition specified below, the rank-r singular value decomposition $M_0 = U\Sigma V^*$ defines the reduced matrix

$$K_B = U^* M_1 V \Sigma^{-1}. \tag{4}$$

This reduced matrix has eigenvalues $\lambda_1, \ldots, \lambda_r$, and the corresponding eigenvectors of T belong to $\operatorname{range}(U)$ [2].

A contour centered far from the origin can render M_1 poorly scaled relative to M_0. Introduce

$$\mu(z) = \frac{z - c}{s_\lambda}, \qquad s_\lambda > 0, \tag{5}$$

where c denotes the contour center and s_λ is a characteristic radius. We therefore replace M_1 with the centered, dimensionless moment

$$\overline{M}_1 = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mu(z) T(z)^{-1} Z \, dz = \frac{M_1 - cM_0}{s_\lambda}. \tag{6}$$

Beyn extraction applied to $(M_0, \overline{M}_1)$ returns $\mu_i = (\lambda_i - c)/s_\lambda$, after which the original eigenvalues are recovered as $\lambda_i = c + s_\lambda \mu_i$. This affine transformation leaves the target subspace unchanged while keeping the two moment blocks comparably scaled for elongated contours or contours far from the origin.

Let Γ be parameterized by a 2π-periodic C^1 function ϕ. The N-point trapezoidal rule is defined by

$$z_j = \phi(\theta_j), \quad \theta_j = \frac{2\pi j}{N}, \quad \omega_j = \frac{\phi'(\theta_j)}{\mathrm{i}N}. \tag{7}$$

Setting $X_j = T(z_j)^{-1} Z$ and $\mu_j = \mu(z_j)$ gives the exact nodal moments

$$M_{0,N} = \sum_{j=0}^{N-1} \omega_j X_j, \qquad \overline{M}_{1,N} = \sum_{j=0}^{N-1} \omega_j \mu_j X_j. \tag{8}$$

We assemble these moments horizontally as $\mathcal{M}_N = [M_{0,N}, \overline{M}_{1,N}]$. This representation gives

$$\mathcal{M}_N = \sum_{j=0}^{N-1} \omega_j X_j H(\mu_j), \qquad H(\mu) = [I_s, \mu I_s], \quad \|H(\mu)\|_2 = \sqrt{1+|\mu|^2}. \tag{9}$$

3. Physical residuals and moment error. Suppose that an iterative solver produces approximations $\widehat{X}_j$, and define the corresponding physical residual by

$$R_j = T(z_j)\widehat{X}_j - Z. \tag{10}$$

Although the sign convention in (10) does not affect the norm bounds, fixing it yields the exact identity below.

THEOREM 1. *Assume that every $T(z_j)$ is nonsingular. Then the solve-induced error in the centered moment block satisfies*

$$E_N^{\mathrm{lin}} := \widehat{\mathcal{M}}_N - \mathcal{M}_N = \sum_{j=0}^{N-1} \omega_j T(z_j)^{-1} R_j H(\mu_j). \tag{11}$$

Consequently,

$$\left\|E_N^{\mathrm{lin}}\right\|_2 \le \sum_{j=0}^{N-1} |\omega_j| \sqrt{1+|\mu_j|^2}\, \left\|T(z_j)^{-1} R_j\right\|_2 \tag{12}$$

$$\le \sum_{j=0}^{N-1} |\omega_j| \sqrt{1+|\mu_j|^2}\, \Gamma_j \left\|R_j\right\|_2 =: \eta_N, \tag{13}$$

whenever $\Gamma_j \ge \left\|T(z_j)^{-1}\right\|_2$.

Proof. Because $T(z_j)X_j = Z$, equation (10) implies $\widehat{X}_j - X_j = T(z_j)^{-1}R_j$. Substituting this relation into the two quadrature sums proves (11). Submultiplicativity and the triangle inequality yield (12), and the assumed resolvent bounds then establish (13). ∎

The correction bound (12) is particularly useful for validation because it separates the effect of the triangle inequality from that of the resolvent estimate. Computing each correction $T(z_j)^{-1}R_j$, however, is generally as costly as improving the original solution; we therefore refer to this quantity as an oracle bound in the experiments. By contrast, (13) requires only the physical residuals once reliable values of Γ_j are available. For small problems, we take $\Gamma_j = 1/\sigma_{\min}(T(z_j))$. At larger scales, possible alternatives include verified lower bounds for the smallest singular value or a preconditioner B_j satisfying

$$\|I - B_j T(z_j)\|_2 \le q_j < 1 \quad \Longrightarrow \quad \left\|T(z_j)^{-1}\right\|_2 \le \frac{\|B_j\|_2}{1-q_j}. \tag{14}$$

Randomized inverse probes may reduce the cost, but they do not furnish an upper bound without an accompanying probabilistic analysis and an explicitly specified failure probability.

The identity also admits a goal-oriented formulation. Equip matrix blocks with the Frobenius inner product $\langle A, B\rangle_F = \operatorname{trace}(A^*B)$, conjugate-linear in its first argument. Then, for $Y = [Y_0, Y_1]$,

$$\left\langle Y, E_N^{\mathrm{lin}}\right\rangle_F = \sum_{j=0}^{N-1} \langle G_j(Y), R_j\rangle_F\,, \qquad G_j(Y) = T(z_j)^{-*}\big(\overline{\omega_j}Y_0 + \overline{\omega_j\mu_j}Y_1\big). \tag{15}$$

This equality permits the estimation of a selected scalar functional using a small number of adjoint solves. It does not certify the full moment norm unless the supremum is taken over all unit dual directions.

3.1. Quadrature error and a nested balancing rule. Define

$$g(\theta) = \frac{\phi'(\theta)}{\mathrm{i}} T(\phi(\theta))^{-1} Z H(\mu(\phi(\theta))). \tag{16}$$

The continuous centered block is then the mean of g, with $\mathcal{M}_N$ as its trapezoidal approximation.

THEOREM 2. *Assume that g is analytic in $|\operatorname{Im}\theta| < \alpha$, continuous on $|\operatorname{Im}\theta| \le \alpha$, and satisfies $\sup_{|\operatorname{Im}\theta|\le\alpha} \|g(\theta)\|_2 \le G_\alpha$. Then*

$$\|\mathcal{M}_N - \mathcal{M}\|_2 \le q_N := \frac{2G_\alpha}{\exp(\alpha N) - 1}. \tag{17}$$

Consequently, $\left\|\widehat{\mathcal{M}}_N - \mathcal{M}\right\|_2 \le q_N + \eta_N$.

Proof. Let $\widehat{g}_k$ denote the matrix-valued Fourier coefficients of g. Shifting the contour to $\operatorname{Im}\theta = \pm\alpha$ gives $\|\widehat{g}_k\|_2 \le G_\alpha e^{-\alpha|k|}$. The N-point trapezoidal rule aliases exactly those coefficients indexed by $k = mN$, $m \ne 0$. Summation of the resulting geometric tail yields (17); see also [24]. ∎

The admissible strip narrows as the complexified contour approaches a pole of T^{-1}. A certified value of G_α is therefore informative but may be expensive to obtain. A nested difference provides a practical balancing quantity without requiring additional node solves. Suppose that N is even. The $N/2$ coarse nodes coincide with the even-indexed fine nodes, and their weights are $2\omega_{2j}$. Construct $\widehat{\mathcal{M}}_{N/2}$ from the same approximations $\widehat{X}_{2j}$ and define

$$\widehat{D}_N = \widehat{\mathcal{M}}_N - \widehat{\mathcal{M}}_{N/2}. \tag{18}$$

With $\Delta X_j = T(z_j)^{-1}R_j$, the solve-induced perturbation satisfies

$$\widehat{D}_N - D_N = \sum_{j\ \mathrm{odd}} \omega_j \Delta X_j H(\mu_j) - \sum_{j\ \mathrm{even}} \omega_j \Delta X_j H(\mu_j).$$

Its norm is thus bounded by the same quantity η_N, rather than by the sum of independent fine- and coarse-grid bounds. It follows that

$$\left|\left\|\widehat{D}_N\right\|_2 - \|D_N\|_2\right| \le \eta_N. \tag{19}$$

If $0 < \vartheta < 1$ and

$$\eta_N \le \vartheta \left\|\widehat{D}_N\right\|_2, \tag{20}$$

then $\eta_N / \|D_N\|_2 \le \vartheta/(1-\vartheta)$. This implication is rigorous. Interpreting $\|D_N\|$ as an estimate of the full quadrature tail requires the additional assumption that the trapezoidal errors have entered a stable geometric regime. If that assumption cannot be justified, one should instead use (17) or introduce a second refinement level.

4. Propagation to the extracted eigendata. The moment certificate quantifies the accuracy of the contour integrals. We now propagate this error to the extracted invariant subspace and the associated reduced eigenproblem.

4.1. Keldysh structure and a direction-sensitive angle bound. Normalize the right eigenvectors by $\|v_i\|_2 = 1$ and scale the left eigenvectors so that $w_i^* T'(\lambda_i) v_i = 1$. Assume that $V = [v_1, \dots, v_r]$ has full column rank. Keldysh's theorem [2, 10] provides a local resolvent expansion; contour integration then yields

$$M_0 = VC, \qquad \overline{M}_1 = V\Lambda_\mu C, \tag{21}$$

where $C = W^* Z \in \mathbb{C}^{r\times s}$, $W = [w_1, \dots, w_r]$, and $\Lambda_\mu = \operatorname{diag}(\mu_1, \dots, \mu_r)$. If C has full row rank, then $\operatorname{rank}(M_0) = r$ and $\operatorname{range}(M_0) = \operatorname{range}(V)$. To establish the graph and reduced-operator identities below, we initially take $s = r$ and assume that C is nonsingular. The analysis extends directly to a fixed compression $Z \mapsto ZG$, with $G \in \mathbb{C}^{s\times r}$, whenever CG is nonsingular. The following bound exposes two distinct sources of sensitivity:

$$\sigma_r(M_0) \ge \sigma_r(V)\sigma_r(W^* Z). \tag{22}$$

Thus, strong eigenvector nonorthogonality or a poorly aligned probing block can obscure the numerical rank of the moment matrix even when the node systems have been solved accurately.

THEOREM 3. *Let $A = M_0 \in \mathbb{C}^{n\times r}$ have full column rank. Define $P = AA^\dagger$ and set $\widehat{A} = A + E_0$. If $\|PE_0\|_2 < \sigma_r(A)$, then $P\widehat{A}$ has full column rank, and*

$$\tan\theta_{\max}(\operatorname{range}(A), \operatorname{range}(\widehat{A})) = \left\| (I-P)E_0 (P\widehat{A})^\dagger \right\|_2 \tag{23}$$

$$\le \frac{\|(I-P)E_0\|_2}{\sigma_r(A) - \|PE_0\|_2}. \tag{24}$$

Proof. Let U have orthonormal columns spanning $\operatorname{range}(A)$, and complete it with $U_\perp$ to form a unitary basis. Write $P\widehat{A} = UF$ and $(I-P)E_0 = U_\perp G$. By the hypothesis and Weyl's inequality, $\sigma_r(F) \ge \sigma_r(A) - \|PE_0\|_2 > 0$, so F is nonsingular. Consequently, $\widehat{A} = (U + U_\perp G F^{-1})F$, which represents $\operatorname{range}(\widehat{A})$ as a graph over $\operatorname{range}(A)$. The singular values of GF^{-1} are therefore the tangents of the principal angles. This characterization gives (23), while the lower bound on $\sigma_r(F)$ yields (24). ∎

The theorem separates perturbations parallel and orthogonal to the target subspace. At first order, the orthogonal component rotates the subspace, whereas the parallel component affects only the denominator. Write $Q_0 = M_{0,N} - M_0$ and $\widehat{M}_0 = \widehat{M}_{0,N}$. Suppose that $q_\perp \ge \|(I-P)Q_0\|_2$, $q_\parallel \ge \|PQ_0\|_2$, and $\|R_j\|_2 \le s_j \tau_j$. Define the directional amplification factors

$$\gamma_{\perp j} = \left\| (I-P)T(z_j)^{-1} \right\|_2, \qquad \gamma_{\parallel j} = \left\| PT(z_j)^{-1} \right\|_2, \tag{25}$$

If $B_\delta := \delta\sigma_r(M_0) - q_\perp - \delta q_\parallel > 0$ and the residual tolerances satisfy the strict linear budget

$$\sum_j |\omega_j| s_j (\gamma_{\perp j} + \delta\gamma_{\parallel j})\tau_j < B_\delta. \tag{26}$$

then (24) gives $\tan\theta_{\max}(\mathrm{range}(M_0),\mathrm{range}(\widehat{M}_0)) < \delta$. Because P is not known initially, this sharper budget serves primarily as an analytical and diagnostic device. A fully a posteriori alternative uses a total bound $\varepsilon_0 \geq \left\|\widehat{M}_0 - M_0\right\|_2$.

COROLLARY 4. *Let $M_0 \in \mathbb{C}^{n\times s}$ have rank $r \leq s$, and let the columns of $\widehat{U}_r$ be the leading r left singular vectors of $\widehat{M}_0$. If $\varepsilon_0 \geq \left\|\widehat{M}_0 - M_0\right\|_2$ and $\sigma_r(\widehat{M}_0) > 0$, then*

$$\sin\theta_{\max}(\mathrm{range}(M_0),\mathrm{range}(\widehat{U}_r)) \leq \min\left\{1, \frac{\varepsilon_0}{\sigma_r(\widehat{M}_0)}\right\}. \tag{27}$$

Proof. Let $P = M_0 M_0^{\dagger}$, and write the leading singular relation as $\widehat{M}_0\widehat{V}_r = \widehat{U}_r\widehat{\Sigma}_r$. Since $(I-P)M_0 = 0$,

$$(I-P)\widehat{U}_r = (I-P)(\widehat{M}_0 - M_0)\widehat{V}_r\widehat{\Sigma}_r^{-1}.$$

The norm of the left-hand side equals the largest sine of the principal angles, which proves the result. ∎

For a stopping decision, it is useful to distinguish the solve error from the quadrature error. Let $M_{0,N}$ and $\widehat{M}_{0,N}$ denote the exact-solve and inexact-solve nodal moments, respectively, and define $\widehat{g}_r = \sigma_r(\widehat{M}_{0,N}) - \sigma_{r+1}(\widehat{M}_{0,N})$. Wedin's theorem, combined with Weyl's inequality [22], gives the following conservative solve-only guard:

$$\sin\theta_{\max}(\mathcal{U}_r(M_{0,N}),\mathcal{U}_r(\widehat{M}_{0,N})) \leq \min\left\{1, \frac{2\eta_{0,N}}{\widehat{g}_r - 2\eta_{0,N}}\right\}, \qquad 2\eta_{0,N} < \widehat{g}_r, \tag{28}$$

where $\eta_{0,N} = \sum_j |\omega_j|\Gamma_j \left\|R_j\right\|_2$ and $\mathcal{U}_r(A)$ denotes the leading r-dimensional left singular subspace of A. Weyl's inequality provides the lower bound $\widehat{g}_r - \eta_{0,N}$ for the cross-gap required by Wedin's theorem. The factor of two and the smaller denominator displayed in (28) constitute a deliberately conservative simplification. Unlike (27), this guard quantifies only the rotation induced by the iterative solves; it does not certify the quadrature approximation to the continuous invariant subspace.

4.2. An exact reduced-operator identity. The next result exploits the algebraic structure of (21), thereby avoiding a generic pseudoinverse perturbation bound.

Set

$$\begin{aligned} E_0 &:= \widehat{M}_{0,N} - M_0 = Q_0 + S_0, \quad Q_0 := M_{0,N} - M_0, \quad S_0 := \sum_j \omega_j T(z_j)^{-1} R_j, \\ E_1 &:= \widehat{\overline{M}}_{1,N} - \overline{M}_1 = Q_1 + S_1, \quad Q_1 := \overline{M}_{1,N} - \overline{M}_1, \quad S_1 := \sum_j \omega_j \mu_j T(z_j)^{-1} R_j. \end{aligned} \tag{29}$$

For brevity, write $A = M_0$, $B = \overline{M}_1$, $\widehat{A} = \widehat{M}_{0,N}$, and $\widehat{B} = \widehat{\overline{M}}_{1,N}$.

LEMMA 5. *Let $s \geq r$, $A = VC$, and $B = V\Lambda_\mu C$, where both V and C have rank r. Let the columns of $L = \widehat{V}_r \in \mathbb{C}^{s\times r}$ be the leading r right singular vectors of $\widehat{A}$, and set $\widehat{\sigma} = \sigma_r(\widehat{A})$. If $\varepsilon_0 \geq \|E_0\|_2$ and $\varepsilon_0 < \widehat{\sigma}$, then $C_L := CL$ is nonsingular and*

$$K_L := (AL)^{\dagger}(BL) = C_L^{-1}\Lambda_\mu C_L, \qquad BL = (AL)K_L. \tag{30}$$

Moreover,

$$\widehat{K}_L := (\widehat{A}L)^\dagger(\widehat{B}L) = \widehat{\Sigma}_r^{-1}\widehat{U}_r^*\widehat{B}\widehat{V}_r \tag{31}$$

is similar to the rank-r Beyn matrix $\widehat{K}_B = \widehat{U}_r^*\widehat{B}\widehat{V}_r\widehat{\Sigma}_r^{-1}$.

Proof. Because $\widehat{A}L = \widehat{U}_r\widehat{\Sigma}_r$ and $\|E_0L\|_2 \le \varepsilon_0$, Weyl's inequality gives $\sigma_r(AL) \ge \widehat{\sigma} - \varepsilon_0 > 0$. Hence $AL = VCL$ has full column rank, which implies that CL is nonsingular; the Keldysh factorization then yields (30). The singular relation gives (31). Finally, $\widehat{K}_B = \widehat{\Sigma}_r\widehat{K}_L\widehat{\Sigma}_r^{-1}$. ∎

THEOREM 6. *Under the assumptions of Theorem 5,*

$$\widehat{K}_L - K_L = (\widehat{A}L)^\dagger(E_1L - E_0LK_L). \tag{32}$$

If also $\varepsilon_1 \ge \|E_1\|_2$, *then*

$$\left\|\widehat{K}_L - K_L\right\|_2 \le \frac{\varepsilon_1 + \varepsilon_0\left\|\widehat{K}_L\right\|_2}{\widehat{\sigma} - \varepsilon_0}. \tag{33}$$

Proof. The identities $BL = (AL)K_L$ and $(\widehat{A}L)^\dagger(\widehat{A}L) = I$ yield (32). On its right-hand side, substitute $K_L = \widehat{K}_L - (\widehat{K}_L - K_L)$ and take norms. Using $\left\|(\widehat{A}L)^\dagger\right\|_2 = 1/\widehat{\sigma}$ and moving the resulting perturbation term to the left gives (33). ∎

The compressed formulation retains a useful structural cancellation. Indeed, (29) gives

$$E_1L - E_0L\widehat{K}_L = Q_1L - Q_0L\widehat{K}_L + \sum_j \omega_j T(z_j)^{-1}R_jL(\mu_jI - \widehat{K}_L). \tag{34}$$

Suppose that $q_K \ge \left\|Q_1L - Q_0L\widehat{K}_L\right\|_2$, $q_0 \ge \|Q_0L\|_2$, $\|R_j\|_2 \le s_j\tau_j$, and $\Gamma_j \ge \left\|T(z_j)^{-1}\right\|_2$. Define

$$a_{0j} = |\omega_j|\Gamma_j s_j, \qquad b_j = |\omega_j|\Gamma_j s_j\left\|\mu_jI - \widehat{K}_L\right\|_2. \tag{35}$$

If $q_0 + \sum_j a_{0j}\tau_j < \widehat{\sigma}$, the argument used in the proof of Theorem 6 yields

$$\left\|\widehat{K}_L - K_L\right\|_2 \le \frac{q_K + \sum_j b_j\tau_j}{\widehat{\sigma} - q_0 - \sum_j a_{0j}\tau_j}. \tag{36}$$

The factor $\left\|\mu_jI - \widehat{K}_L\right\|$ preserves the coupling between the two moments, which is lost if E_0 and E_1 are bounded separately.

Since $K_L = C_L^{-1}\Lambda_\mu C_L$, the Bauer–Fike theorem [22], together with (36), places each computed eigenvalue $\widehat{\mu}$ in a union of disks:

$$\min_i |\widehat{\mu} - \mu_i| \le \mathrm{cond}_2(C_L)\left\|\widehat{K}_L - K_L\right\|_2. \tag{37}$$

These disks imply a one-to-one matching only when their radii are smaller than half the minimum eigenvalue separation. Given $\overline{\kappa} \ge \mathrm{cond}_2(C_L)$ and a target normalized radius δ_μ, define $B_\mu := \delta_\mu(\widehat{\sigma} - q_0) - \overline{\kappa}q_K$. If $B_\mu > 0$, the linear budget

$$\sum_j (\overline{\kappa}b_j + \delta_\mu a_{0j})\tau_j < B_\mu \tag{38}$$

is sufficient to enforce the target normalized radius. The corresponding radius in the original variable is $s_\lambda \delta_\mu$. Without reliable upper bounds on $\mathrm{cond}_2(C_L)$ and the quadrature terms, (36) remains a conditional reduced-operator bound rather than a verified eigenvalue enclosure.

5. Accuracy control with shared infinite-GMRES bases.

5.1. The residual that must be monitored. Infinite GMRES expands T in a Taylor series about an expansion point η and applies a companion linearization to the resulting degree-p polynomial,

$$P_p(z) = \sum_{k=0}^{p} (z-\eta)^k T_k, \qquad T_k = \frac{T^{(k)}(\eta)}{k!}. \tag{39}$$

For every node z assigned to η, a single Arnoldi relation induces a small shifted least-squares problem [16, 20]. Let g_j denote the corresponding companion residual and $\widehat{X}_j$ the recovered physical block. Because the moment certificate depends on the residual of the original NEP, substituting g_j for R_j is justified only after accounting for both the linearization map and the Taylor truncation. In particular,

$$\left\| T(z_j)\widehat{X}_j - Z \right\|_2 \le \left\| P_p(z_j)\widehat{X}_j - Z \right\|_2 + \left\| T(z_j) - P_p(z_j) \right\|_2 \left\| \widehat{X}_j \right\|_2. \tag{40}$$

Our implementation therefore evaluates R_j explicitly at every checkpoint. The additional matrix–block product is inexpensive relative to extending the basis and prevents the stopping test from underestimating the error near a boundary of Taylor convergence.

5.2. Node tolerances and cluster work. To motivate node-dependent stopping tolerances, first consider independent node solves. A generic linear error budget has the form

$$\sum_j a_j \tau_j \le B, \tag{41}$$

where a_j may represent the moment coefficient in (13), the subspace coefficient in (26), or the eigenvalue-goal coefficient in (38). Under a logarithmic work model, the following proposition characterizes the resulting nonuniform allocation.

PROPOSITION 7. *Assume $a_j > 0$, $0 < \tau_j^{\min} \le \tau_j^{\max}$, and $B \ge \sum_j a_j \tau_j^{\min}$. Let the work to reach tolerance τ_j be $C_j(\tau_j) = \beta_j \log(\tau_j^{\mathrm{ref}}/\tau_j) + c_j$ with $\beta_j > 0$. If $B \ge \sum_j a_j \tau_j^{\max}$, the unique minimizer is $\tau_j = \tau_j^{\max}$. Otherwise the budget is active and the unique primal minimizer is*

$$\tau_j(\nu) = \mathrm{clip}\left(\frac{\beta_j}{\nu a_j}, \tau_j^{\min}, \tau_j^{\max} \right), \tag{42}$$

where $\nu > 0$ is chosen so that $\sum_j a_j \tau_j(\nu) = B$. If no box constraint is active, then

$$\tau_j^* = \frac{\beta_j B}{a_j \sum_k \beta_k}. \tag{43}$$

Proof. After constants are removed, the objective is $-\sum_j \beta_j \log \tau_j$, which is strictly convex on the positive orthant. For an interior variable, the stationarity condition is $-\beta_j/\tau_j + \nu a_j = 0$. The complementary-slackness conditions clip the resulting value at the box constraints. Monotone bisection determines the multiplier, which is unique whenever at least one variable is interior. ∎

The proposition is exact under the stated work model; it is not a general GMRES complexity result. In particular, residual histories for nonnormal problems may depart substantially from exponential decay. At each checkpoint, $\widehat{K}_L$, $\widehat{\sigma}$, and the associated coefficients are held fixed. The resulting allocation governs only the next batch and is recomputed at the subsequent checkpoint. The proposition thus establishes conditional optimality for a frozen local model, without asserting global optimality of the adaptive schedule. Pilot iterations may be used to estimate β_j, whereas the final stopping decision is determined by the physical-residual bound.

The shared-basis structure of infinite GMRES introduces a different coupling. All nodes assigned to an expansion point η_t use the same basis and therefore inherit its length m_t. For a cluster $\mathcal{J}_t$, define its contribution to the error bound by

$$\Phi_t(m) = \sum_{j\in\mathcal{J}_t} |\omega_j|\sqrt{1+|\mu_j|^2}\Gamma_j \left\|R_j(m)\right\|_2 . \tag{44}$$

If the residual histories are fixed and fully recorded, selecting the lengths m_t yields the discrete resource-allocation problem

$$\min_{m_t\in\{0,\dots,m_{\max}\}} \sum_t C_t(m_t) \quad \text{subject to} \quad q_N + \sum_t \Phi_t(m_t) \le B. \tag{45}$$

Dynamic programming can solve (45) offline. An online scheduler, however, has no access to future residual trajectories. We estimate the marginal decrease in Φ_t from recent history, select one cluster, extend its basis by a short batch, and then reevaluate every physical residual in that cluster. Provided that each Γ_j is a certified upper bound, an inaccurate prediction can reduce efficiency but cannot invalidate the subsequently recomputed bound. With floating-point resolvent estimates, the same quantity serves only as a numerical scheduling surrogate.

The fallback in Algorithm 1 is particularly important near poles and for poorly placed expansion points. Residual-driven pole or shift selection is established in rational Krylov methods, which offer more sophisticated treatments of the parameter domain [7, 4]. The score used here has a different objective: it quantifies each cluster's contribution to the contour moments and to the Beyn extraction target. Region-partitioning experiments in [21] further demonstrate that inaccurate infinite-GMRES solves may miss eigenvalues near difficult contour segments. This evidence motivates explicit monitoring of the physical residual together with a local fallback.

6. Numerical experiments. The experiments are designed to address five questions. First, do the evaluated residual bounds dominate the observed solve-induced moment error? Second, at what stage does the reference-grid discrepancy become dominant? Third, can a singular-subspace safeguard prevent premature termination by the one-level nested test? Fourth, how does increasing nonnormality change the relation between a fixed basis length, the physical residual, and the extracted subspace? Fifth, how sensitive is the stopping decision to its two thresholds, and how concentrated are the node and cluster contributions that drive work allocation?

6.1. Implementation and protocol. The reproducibility package provides a compact Python implementation of the companion Arnoldi relation for infinite GMRES. The code constructs the truncated block linearization implicitly, applies two-pass modified Gram–Schmidt, and solves the shifted least-squares problems at selected basis lengths. It deliberately omits the TOAR compression and some weighting optimizations described in [20]; its purpose is to provide an accuracy-oriented reference

Algorithm 1 Residual-budgeted centered Beyn moments

Choose an even N, probing block Z, moment scale (c, s_λ), expansion points $\{\eta_t\}$, and a subspace tolerance $\delta_{\mathcal{U}}$.
Assign each node z_j to a cluster $\mathcal{J}_t$ and initialize one infinite-GMRES basis per cluster and probing direction.
Extend every basis to a short pilot length; recover $\widehat{X}_j$ from the small shifted least-squares problems.
Evaluate the physical residuals $R_j = T(z_j)\widehat{X}_j - Z$ and assemble $\widehat{\mathcal{M}}_N$ and the reused-node difference $\widehat{D}_N$.
Set the solve-only subspace guard to $+\infty$ whenever $2\eta_{0,N} \geq \widehat{g}_r$; otherwise evaluate (28).
while $\eta_N > \vartheta \left\|\widehat{D}_N\right\|_2$ or the bound in (28) exceeds $\delta_{\mathcal{U}}$ **do**
  Select the cluster with the largest predicted bound decrease per unit work.
  Extend its basis by a short batch; recompute its physical residuals, η_N, $\widehat{\mathcal{M}}_N$, $\widehat{D}_N$, the singular gap, and the subspace guard.
  if the cluster stagnates or a resolvent estimate exceeds a safeguard **then**
    Add a local expansion point, apply an independent correction solve, or use a direct-solve fallback for the marked nodes.
  end if
end while
Compute a rank-revealing SVD of $\widehat{M}_{0,N}$, solve the centered Beyn problem, and map $\widehat{\mu}_i$ to $\widehat{\lambda}_i = c + s_\lambda \widehat{\mu}_i$.
Check final NEP residuals. If they fail while rank and quadrature indicators are stable, continue the influential bases or invoke a local fallback; refine N or the contour only when rank or the quadrature tail remains unresolved.

implementation. Every reported node residual is recomputed using the original matrix function $T(z_j)$. For deterministic checkpoint comparisons, each basis is first constructed to the longest requested length and subsequently replayed using shorter prefixes. Consequently, the reported work equals the sum of retained Arnoldi-prefix iterations and should not be interpreted as measured wall-clock acceleration from an online scheduler.

For validation, we set $\Gamma_j = 1/\widehat{\sigma}_{\min}(T(z_j))$, with the singular value evaluated in ordinary double precision. The resulting quantity is a numerical oracle bound rather than an outward-rounded certificate. A verified application of (13) would require a certified lower bound for $\sigma_{\min}$. We report this resolvent bound alongside the correction oracle (12). The reference moments are computed by direct solves on a 256-node trapezoidal grid, while an independent 256-versus-512 comparison measures the remaining reference-grid difference. For error reporting, eigenvalues are paired by a minimum-cost assignment; the pairing is empirical and is not deduced from Bauer–Fike. We measure pointwise accuracy using the matrix-normalized NEP residual

$$\rho(\widehat{\lambda}, \widehat{v}) = \frac{\left\|T(\widehat{\lambda})\widehat{v}\right\|_2}{\left\|T(\widehat{\lambda})\right\|_2 \|\widehat{v}\|_2}. \tag{46}$$

For the base experiments, random probing blocks are generated from seed 20260722 and then orthonormalized. The experiments were run on Windows with Python 3.9.13,

TABLE 1
Configurations used in the numerical experiments. The moment scale is $s_\lambda = a$.

problem	n	rank	block size	N	(c, a, b)	expansion points
synthetic delay	40	6	6	32	$(0, 0.95, 0.72)$	1
Hadeler	200	13	18	16,32,64	$(0.01, 1.14\cdot10^{-3}, 1.15\cdot10^{-4})$	1
loaded string	200	10	14	64	$(1000, 850, 85)$	8

NumPy 1.21.5, and SciPy 1.9.1. The archive records the environment metadata in `summary.json`; separate CSV files contain timings, residual distributions, moment errors, subspace angles, and scaled matched-eigenvalue errors $|\widehat{\lambda} - \lambda|/\max\{1, |\lambda|\}$. The nonnormality sweep uses 13 values of $\mathrm{cond}_2(S)$, five paired seed pairs (problem: 20260722–20260726; probe: 20270722–20270726), and 17 prefixes. The stopping study replays every combination of seven balance fractions and eight subspace tolerances against four archived histories, without interpolation. The contribution study independently rebuilds the loaded-string bases and records every node and cluster term in (44). As a reference-grid sensitivity check, doubling the number of nodes from 256 to 512 changes the stacked centered moments by $4.94\cdot10^{-16}$ for Hadeler and $8.57\cdot10^{-10}$ for loaded string. The corresponding maximum scaled eigenvalue changes are $5.38\cdot10^{-17}$ and $6.81\cdot10^{-15}$.

6.2. A nonnormal analytic family. We first construct an entire NEP whose target roots are prescribed. Let S be nonsingular, set $\Lambda = \mathrm{diag}(\lambda_i)$, and choose small positive parameters α_i. In eigenvector coordinates, the scalar factors are

$$d_i(z) = z - \lambda_i + \alpha_i\big(e^{-\tau(z-\lambda_i)} - 1\big), \qquad \tau = 0.3. \tag{47}$$

The matrix function $T(z) = S\,\mathrm{diag}(d_i(z))S^{-1}$ therefore has every prescribed λ_i as an exact eigenvalue, although the exponential terms also introduce nonprincipal roots. We use Rouché's theorem to certify the number of enclosed eigenvalues. On the contour, $0.72 \le |z| \le 0.95$; each target satisfies $|\lambda_i| \le 0.55$, each prescribed exterior value satisfies $|\lambda_i| \ge 1.25$, and $\alpha_i \le 0.03125$. For a target factor,

$$|\alpha_i(e^{-\tau(z-\lambda_i)} - 1)| \le 0.03125\{e^{0.3(0.95+0.55)} + 1\} < 0.081 < 0.17 \le |z - \lambda_i|,$$

whereas for an exterior factor, the analogous upper bound with 1.65 in place of 0.55 is below $0.100 < 0.30 \le |z - \lambda_i|$. Each scalar factor consequently has the same enclosed zero count as $z - \lambda_i$, and exactly six eigenvalues lie inside the ellipse. We consider $\mathrm{cond}_2(S)$ equal to 1, 10^3, and 10^6.

Figure 1 compares the observed solve-induced moment error with the correction oracle and reports the effectivity of the numerically evaluated resolvent bound. At every recorded checkpoint, both bounds exceed the observed error; the data archive reports their minimum empirical effectivities. The 10^6 case also exposes a limitation of residual-only assessment. Severe eigenvector nonnormality delays convergence of the extracted subspace even though the underlying algebraic relations remain valid.

To separate conditioning effects from a single random realization, we next use 13 logarithmically spaced values of $\mathrm{cond}_2(S)$ from 1 to 10^6 and five paired seeds. Within each seed, the target and exterior eigenvalues, the two unitary factors of S, the nonlinear coefficients, and the probing block are fixed; only the prescribed singular values of S change. We compare a fixed prefix $m = 24$ with the computed checkpoint whose maximum node relative residual is closest to 10^{-2}. The latter is a diagnostic residual-matched comparison rather than a new stopping rule.

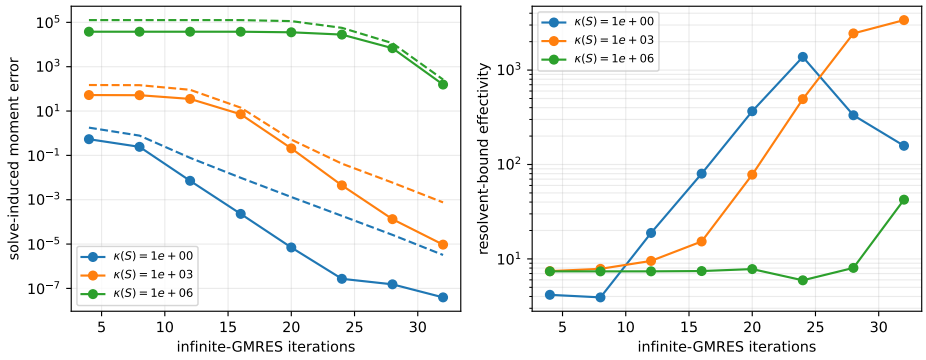

FIG. 1. *Synthetic delay NEP. Left: solve-induced centered-moment error (solid with markers) and correction oracle (dashed). Right: effectivity of the numerically evaluated resolvent bound. In these runs, the evaluated bound dominates the observed error, while its sharpness and the sensitivity of the extraction vary with eigenvector conditioning.*

Figure 2 reveals a sharp distinction between fixed work and fixed residual quality. At $m = 24$, the median maximum node residual grows from $4.25 \cdot 10^{-4}$ at $\mathrm{cond}_2(S) = 1$ to 1.04 at 10^6, while the median solve-induced subspace-angle sine grows from $5.30 \cdot 10^{-6}$ to $8.49 \cdot 10^{-1}$. Reaching the 10^{-2} residual scale instead raises the median selected prefix from $m = 18$ to $m = 36$; at $\mathrm{cond}_2(S) = 10^6$, the median solve-induced angle is then $3.02 \cdot 10^{-4}$. Thus, a basis length calibrated on a mildly nonnormal instance does not transfer across the family. Across all 1105 computed prefixes, the evaluated resolvent-bound and correction-oracle effectivities remain at least 4.28 and 1.57, respectively. The large effectivities in the residual-matched panel also emphasize that validity and sharpness are separate questions.

6.3. NLEVP benchmarks and the reference-grid plateau. The first benchmark is the Hadeler problem from NLEVP [1], defined by

$$T(\lambda) = (e^{\lambda} - 1)B + \lambda^2 A_2 - 100I. \tag{48}$$

For $n = 200$, the selected ellipse contains 13 eigenvalues. We place a single expansion point at c and scale the Taylor variable by four times the largest node distance to improve convergence of the companion Krylov iteration.

The second benchmark is the loaded vibrating string problem, defined by

$$T(\lambda) = A - \lambda B + \frac{\lambda}{\lambda - 1} C, \tag{49}$$

where C has rank one. The contour encloses 10 eigenvalues and excludes the pole at 1. Eight expansion points are placed on an ellipse whose radii are 0.82 times those of the contour. For each cluster, the Taylor-variable scale is twice the largest distance to one of its nodes.

Figures 3 and 4 present the prefix work–precision curves. For Hadeler, extending the basis initially reduces the total moment error, after which each curve reaches an N-dependent plateau. At $N = 64$, the solve-induced component continues to decrease even after the total error relative to the 256-node reference has stabilized. The eigenvalue error displays the same transition. The plateau is more pronounced for loaded string: beyond a modest basis length, additional Arnoldi iterations change

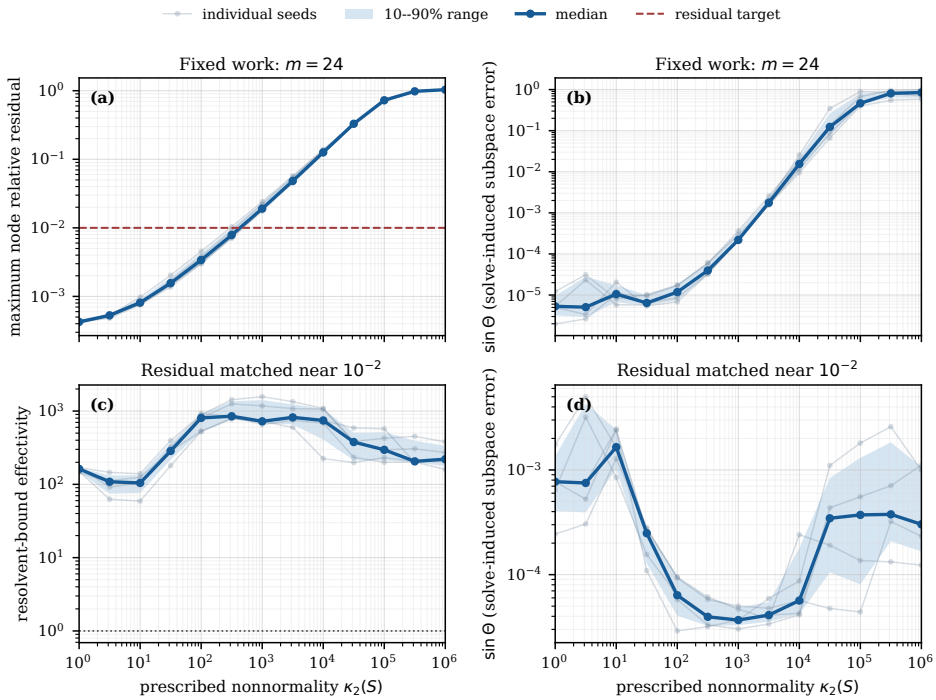

FIG. 2. *Paired-seed nonnormality sweep for the synthetic delay NEP (five seeds per condition). Top: maximum node residual and solve-induced subspace error at fixed prefix $m = 24$. Bottom: resolvent-bound effectivity and subspace error at the computed prefix whose maximum node residual is closest to 10^{-2}. Thick curves denote medians, shaded regions the 10th–90th percentile range, and faint curves individual seeds.*

TABLE 2

Moment-only and singular-subspace-safeguarded stopping. Error is the maximum scaled matched-eigenvalue error; the baseline is the longest recorded prefix.

problem	N	m_{mom}	save (%)	$\mathrm{error}_{\mathrm{mom}}$	m_{safe}	save (%)	$\mathrm{error}_{\mathrm{safe}}$	$\mathrm{error}_{\mathrm{long}}$
hadeler	16	28	41.7	9.95×10^{-6}	40	16.7	1.12×10^{-5}	1.12×10^{-5}
hadeler	32	28	41.7	1.19×10^{-7}	40	16.7	6.03×10^{-8}	6.03×10^{-8}
hadeler	64	32	33.3	1.22×10^{-10}	40	16.7	4.67×10^{-12}	4.69×10^{-12}
loaded string	64	12	70.0	1.01×10^{-4}	24	40.0	4.46×10^{-8}	4.42×10^{-8}

neither the reference-grid moment discrepancy nor the matched eigenvalues at the displayed scale. The independent 256-versus-512 comparison provides evidence that the selected reference grid is numerically converged; it does not constitute a verified bound on the continuous quadrature error.

Table 2 distinguishes the two stopping decisions. The moment-only criterion applies (20) with $\vartheta = 0.2$ and reduces counted prefix work by 33–70 percent. This reduction alone, however, does not ensure accurate eigendata. For loaded string, the criterion selects $m = 12$, at which point the maximum scaled eigenvalue error increases from $4.42 \cdot 10^{-8}$ to $1.01 \cdot 10^{-4}$. The one-level nested difference, 5.97, exceeds the 256-node reference-grid discrepancy, 0.247, by a factor of about 24. Although this observation is consistent with (19), it indicates that the geometric-tail model has not

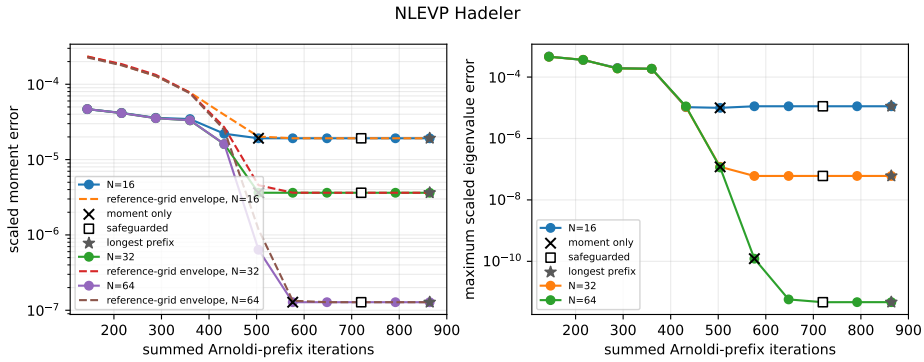

FIG. 3. *Hadeler prefix work–precision results. The dashed curves combine the correction oracle with the 256-node reference-grid discrepancy. Crosses, open squares, and stars identify the moment-only, safeguarded, and longest-prefix choices, respectively. Once the solid curve reaches its N-dependent plateau, additional infinite-GMRES iterations do not improve the extracted eigenvalues at the displayed scale.*

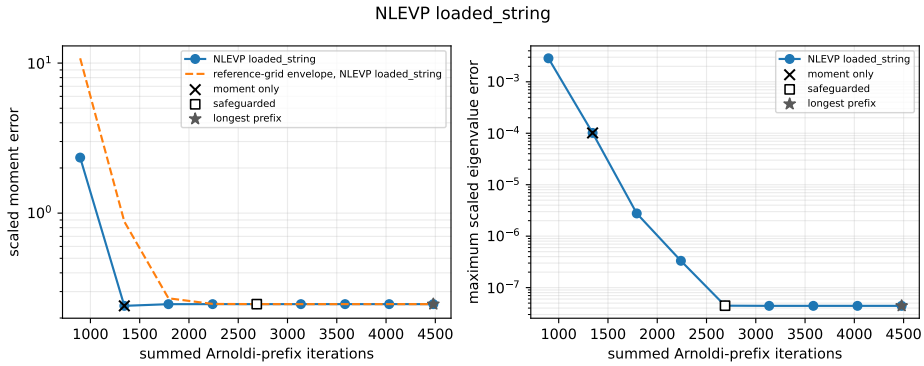

FIG. 4. *Loaded-string prefix work–precision results with eight expansion-point clusters. Crosses, open squares, and stars identify the moment-only, safeguarded, and longest-prefix choices. Centering the moment removes a factor proportional to the contour center from the stopping weights.*

yet stabilized.

The safeguarded criterion additionally requires (28) to be no greater than $\delta_{\mathcal{U}} = 10^{-4}$. For loaded string, it selects $m = 24$, reduces counted prefix work by 40 percent, and yields a maximum scaled eigenvalue error of $4.46 \cdot 10^{-8}$. For each of the three Hadeler quadratures, it selects $m = 40$, corresponding to a 16.7 percent reduction, and reproduces the longest-prefix errors to the displayed accuracy. The guard controls rotation of the leading singular subspace of the discrete moment matrix; in isolation, it is not an eigenvalue certificate. Its empirical performance supports the two-level decision for these tests, but certified eigendata accuracy still requires the goal-oriented bound (36).

6.4. Stopping sensitivity and contribution concentration. The nominal parameters should not be interpreted from one selected point alone. We therefore replay the two-level rule for $\vartheta \in \{0.01, 0.03, 0.1, 0.2, 0.3, 0.6, 1\}$ and $\delta_{\mathcal{U}} \in \{10^{-1}, 3 \cdot 10^{-2}, 10^{-2}, 3 \cdot 10^{-3}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$. This produces 224 decisions across the three Hadeler quadratures and loaded string, each at an actually computed prefix.

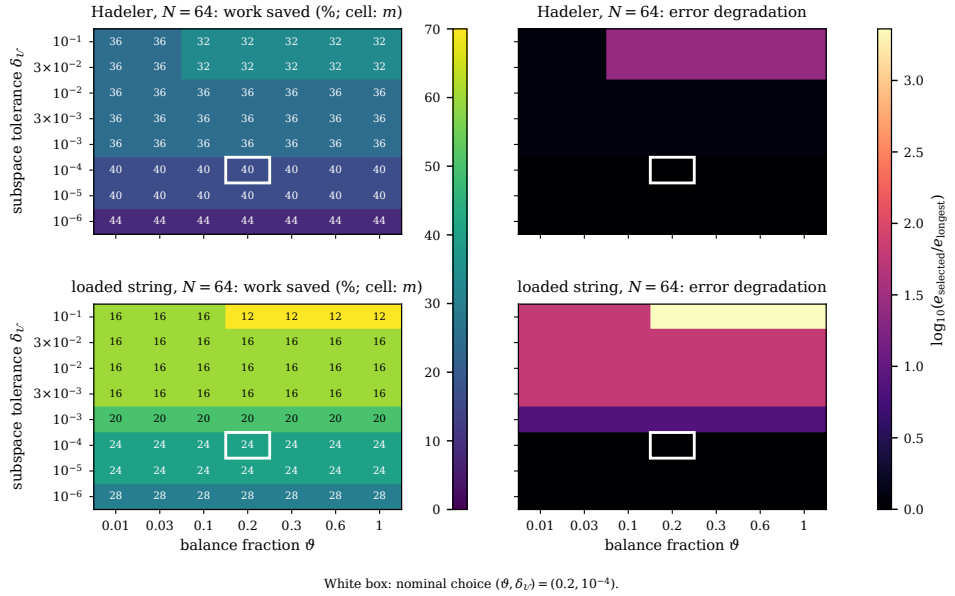

FIG. 5. *Sensitivity of two-level stopping to the balance fraction ϑ and subspace tolerance $\delta_{\mathcal{U}}$. Left panels show counted prefix-work reduction, with the selected basis length m printed in each cell. Right panels show $\log_{10}(e_{\text{selected}}/e_{\text{longest}})$ for the maximum scaled eigenvalue error. White boxes mark the nominal choice $(0.2, 10^{-4})$.*

Figure 5 shows the representative $N = 64$ cases. At the nominal point $(0.2, 10^{-4})$, Hadeler selects $m = 40$, saves 16.7 percent of counted work, and has an error ratio of 0.997 relative to the longest prefix; loaded string selects $m = 24$, saves 40 percent, and has an error ratio of 1.008. The loaded-string transition is sharp: with $\vartheta = 0.2$, relaxing $\delta_{\mathcal{U}}$ from 10^{-4} to 10^{-3}, 10^{-2}, and 10^{-1} changes the selected prefixes to $m = 20$, 16, and 12, but increases the eigenvalue error by factors of 7.50, 62.5, and 2294, respectively. This sweep does not establish universal thresholds; it demonstrates why both parameters must be reported and stress-tested.

The cluster-allocation premise can also be checked directly. For each loaded-string checkpoint $m \in \{12, 16, 20, 24, 40\}$, we recompute the node terms

$$c_j^{(m)} = |\omega_j| \sqrt{1 + |\mu_j|^2} \Gamma_j \, \|R_j(m)\|_2 \,, \qquad p_j^{(m)} = c_j^{(m)} \Big/ \sum_k c_k^{(m)}.$$

The fresh sums reproduce the archived resolvent bounds to a maximum relative difference of $4.73 \cdot 10^{-13}$. As Figure 6 shows, the distribution is highly concentrated rather than uniform. One expansion-point cluster carries 83.8–90.9 percent of the bound across the five prefixes; the eight largest node contributions carry 86.5–97.8 percent. The inverse-Herfindahl effective counts are only 4.02–6.40 nodes out of 64 and 1.20–1.41 clusters out of eight. These data do not prove that a greedy online scheduler is optimal, but they show that cluster-aware allocation acts on a substantial and persistent imbalance.

These experiments do not establish a wall-clock advantage over specialized direct solvers. The Hadeler matrix is dense, whereas loaded string is tridiagonal plus rank one; both problems possess structure that the generic iterative implementation leaves

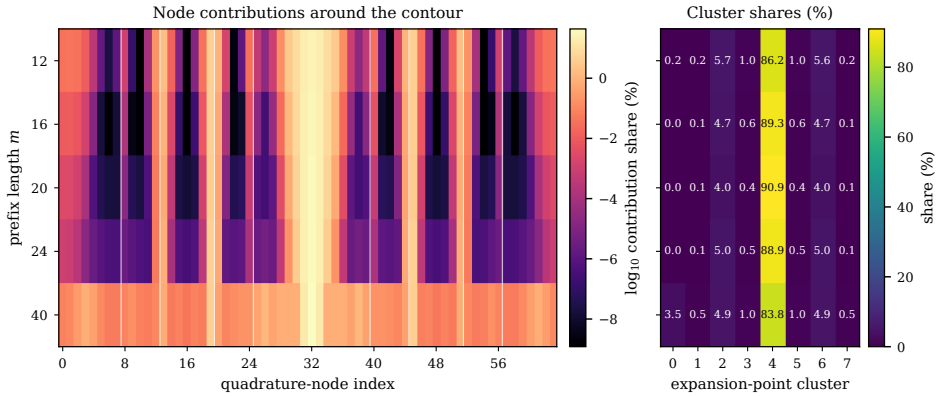

FIG. 6. *Loaded-string contributions to the evaluated residual bound. Left: logarithmic node shares around the 64-node contour; white separators indicate expansion-point clusters. Right: aggregate cluster shares in percent. Only a few nodes and one cluster dominate throughout the recorded prefixes.*

unexploited. The benchmarks instead serve a diagnostic purpose: they reveal oversolving and permit direct evaluation of the relevant singular values. A production-scale assessment should include large sparse PDE models, certified or statistically calibrated resolvent estimators, the weighted TOAR implementation of [20], and direct comparisons with modern rational Krylov solvers [7, 4].

7. Limitations and implications. The preceding analysis delineates four requirements that remain for a fully rigorous and broadly applicable implementation.

First, certification requires each Γ_j to be a valid upper bound. Floating-point singular values, condition estimators, and random inverse probes can guide work allocation, but multiplying such estimates by an empirical safety factor does not provide a verified guarantee. A practical implementation could adopt a hybrid strategy: inexpensive estimates would schedule the clusters, while outward-rounded singular-value or preconditioner bounds would validate the largest predicted contributors.

Second, the nested difference diagnoses the current discretization but does not furnish a universal tail bound. When an eigenvalue lies close to the contour, the observed error may remain preasymptotic. In that regime, one must invoke the analytic-strip bound in Theorem 2, introduce a second refinement level, or deform the contour. Deterioration in accuracy near a quadrature node is documented even for linear problems [11].

Third, a small value of (46) certifies a small pointwise unstructured perturbation of $T(\widehat{\lambda})$ for the reported vector. It does not certify either a small structured perturbation of the analytic matrix function or completeness of the spectrum enclosed by the contour. Those properties require separate evidence, such as a stable numerical rank, agreement under quadrature refinement, and an eigenvalue count or verified moment argument. Recent convergence analysis explains why iterative NLFEAST can refine beyond a fixed quadrature filter, whereas one-shot Beyn extraction generally encounters a quadrature floor [8, 19]. The present bounds quantify the algebraic work needed to approach that floor; they do not supply an outer refinement mechanism.

Fourth, the extraction theory assumes simple eigenvalues, a known target rank, linearly independent right eigenvectors, and a full-rank compressed probing interac-

tion C_L. Rank misidentification, multiple or defective eigenvalues, and extraction from higher moments introduce additional gap conditions. The residual identity in Theorem 1 remains valid for any finite collection of moments, with $\sqrt{1+|\mu|^2}$ replaced by $(\sum_{k=0}^{q-1}|\mu|^{2k})^{1/2}$. Extending the structured reduced-operator enclosure to defective eigenvalues requires a pseudospectral formulation in place of Bauer–Fike analysis.

8. Conclusions. This work argues that stopping criteria for inexact contour solves should be tied to their effect on the extracted quantity rather than to a uniform demand for machine precision. For centered Beyn moments, an exact identity expresses that effect in terms of the residuals of the original NEP. From this identity follow a norm bound, a dual-weighted estimate, a direction-sensitive subspace budget, and a structured perturbation enclosure for the reduced matrix. Once geometric convergence has been established, the nested rule supplies a practical scale for the algebraic error; before that regime, it remains diagnostic. Cluster-level contributions additionally account for the shared-basis cost structure of infinite GMRES.

The experiments identify both the useful regime and the failure mode of the proposed stopping logic. Once the solve contribution lies below a stable discretization scale, further Krylov iterations continue to reduce the node residuals without improving the computed eigendata. The paired-seed sweep additionally shows that fixed work is not an accuracy specification: at $m = 24$, increasing $\mathrm{cond}_2(S)$ from 1 to 10^6 changes the median solve-induced subspace-angle sine from $5.30 \cdot 10^{-6}$ to $8.49 \cdot 10^{-1}$. The loaded-string example shows that a one-level nested difference may declare the plateau prematurely, and the threshold grid quantifies how an overpermissive subspace tolerance converts additional work savings into orders-of-magnitude eigenvalue degradation. At the nominal settings, adding the solve-only singular-subspace guard preserves the accuracy of the longest-prefix calculation while avoiding 17–40 percent of counted prefix work. The contribution decomposition further shows that one loaded-string cluster carries more than 83 percent of the evaluated bound at every inspected prefix, supporting cluster-aware rather than uniform allocation. Establishing practical performance at scale will require inexpensive reliable resolvent bounds, an online cluster scheduler, and direct comparisons with state-of-the-art parameterized Krylov implementations.

Data and code availability. The package contains source, fixed seeds, unit tests, CSV outputs, figures, and LaTeX. The three drivers are `run_experiments.py`, `extended_synthetic_experiment.py`, and `run_additional_experiments.py`. The NLEVP formulas are implemented from the public collection [1]. All reported values, tables, and figures are generated from the archived CSV files.

REFERENCES

[1] T. Betcke, N. J. Higham, V. Mehrmann, C. Schröder, and F. Tisseur, *NLEVP: A collection of nonlinear eigenvalue problems*, ACM Transactions on Mathematical Software, 39 (2013), pp. 7:1–7:28, doi:10.1145/2427023.2427024.

[2] W.-J. Beyn, *An integral method for solving nonlinear eigenvalue problems*, Linear Algebra and its Applications, 436 (2012), pp. 3839–3863, doi:10.1016/j.laa.2011.03.030.

[3] M. C. Brennan, M. Embree, and S. Gugercin, *Contour integral methods for nonlinear eigenvalue problems: A systems theoretic approach*, SIAM Review, 65 (2023), pp. 439–470, doi:10.1137/20M1389303.

[4] K. Bruyninckx, D. Huybrechs, and K. Meerbergen, *Compact rational Krylov for parametrized systems with application to BEM frequency sweeping*, 2026, https://arxiv.org/abs/2607.07440, arXiv:2607.07440 [math.NA].

[5] H. Chen and L. Du, *Backward error bounds for polynomial eigenvalue problem solved by a Rayleigh–Ritz type contour integral-based eigensolver*, Applied Mathematics Letters, 102 (2020), p. 106122, doi:10.1016/j.aml.2019.106122.

[6] S. Correnty, E. Jarlebring, and K. M. Soodhalter, *Preconditioned infinite GMRES for parameterized linear systems*, SIAM Journal on Scientific Computing, 46 (2024), pp. S120–S141, doi:10.1137/22M1502380.

[7] H. A. Daas and D. Palitta, *Minimal residual rational Krylov subspace method for sequences of shifted linear systems*, 2025, arXiv:2507.00267 [math.NA]. Revised January 2026.

[8] B. Gavin, A. Międlar, and E. Polizzi, *FEAST eigensolver for nonlinear eigenvalue problems*, Journal of Computational Science, 27 (2018), pp. 107–117, doi:10.1016/j.jocs.2018.05.006.

[9] B. Gavin and E. Polizzi, *Krylov eigenvalue strategy using the FEAST algorithm with inexact system solves*, Numerical Linear Algebra with Applications, 25 (2018), p. e2188, doi:10.1002/nla.2188.

[10] S. Güttel and F. Tisseur, *The nonlinear eigenvalue problem*, Acta Numerica, 26 (2017), pp. 1–94, doi:10.1017/S0962492917000034.

[11] T. Hasegawa, A. Imakura, and T. Sakurai, *Recovering from accuracy deterioration in the contour integral-based eigensolver*, JSIAM Letters, 8 (2016), pp. 1–4, doi:10.14495/jsiaml.8.1.

[12] S. Huber, Y. Futamura, M. Galgon, A. Imakura, B. Lang, and T. Sakurai, *Flexible subspace iteration with moments for an effective contour integration-based eigensolver*, Numerical Linear Algebra with Applications, 29 (2022), p. e2447, doi:10.1002/nla.2447.

[13] A. Imakura, L. Du, and T. Sakurai, *Error bounds of Rayleigh–Ritz type contour integral-based eigensolver for solving generalized eigenvalue problems*, Numerical Algorithms, 71 (2016), pp. 103–120, doi:10.1007/s11075-015-9987-4.

[14] A. Imakura, K. Morikuni, and A. Takayasu, *Verified partial eigenvalue computations using contour integrals for hermitian generalized eigenproblems*, Journal of Computational and Applied Mathematics, 369 (2020), p. 112543, doi:10.1016/j.cam.2019.112543.

[15] A. Imakura, K. Morikuni, and A. Takayasu, *Verified eigenvalue and eigenvector computations using complex moments and the Rayleigh–Ritz procedure for generalized hermitian eigenvalue problems*, Journal of Computational and Applied Mathematics, 424 (2023), p. 114994, doi:10.1016/j.cam.2022.114994.

[16] E. Jarlebring and S. Correnty, *Infinite GMRES for parameterized linear systems*, SIAM Journal on Matrix Analysis and Applications, 43 (2022), pp. 1382–1405, doi:10.1137/21M1410324.

[17] Z. Jia and Q. Zheng, *An analysis of the Rayleigh–Ritz and refined Rayleigh–Ritz methods for regular nonlinear eigenvalue problems*, SIAM Journal on Matrix Analysis and Applications, 46 (2025), pp. 676–701, doi:10.1137/23M161392X.

[18] L. Krämer, E. D. Napoli, M. Galgon, B. Lang, and P. Bientinesi, *Dissecting the FEAST algorithm for generalized eigenproblems*, Journal of Computational and Applied Mathematics, 244 (2013), pp. 1–9, doi:10.1016/j.cam.2012.11.014.

[19] D. Kressner, Y. Liu, J. E. Roman, M. Shao, and N. Shao, *Linear convergence of iterative contour integral-based eigensolvers for nonlinear eigenvalue problems*, 2026, https://arxiv.org/abs/2606.13357, arXiv:2606.13357 [math.NA].

[20] Y. Liu, J. E. Roman, and M. Shao, *Improving performance of contour integral-based nonlinear eigensolvers with infinite GMRES*, SIAM Journal on Scientific Computing, 47 (2025), pp. B595–B617, doi:10.1137/24M1650375.

[21] Y. Liu, J. E. Roman, and M. Shao, *Solving nonlinear eigenvalue problems via contour integration and region partitioning*, Numerical Linear Algebra with Applications, 33 (2026), p. e70072, doi:10.1002/nla.70072.

[22] G. W. Stewart and J. guang Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.

[23] P. T. P. Tang and E. Polizzi, *FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 354–390, doi:10.1137/13090866X.

[24] L. N. Trefethen and J. A. C. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Review, 56 (2014), pp. 385–458, doi:10.1137/130932132.