THE EXACT WORST-CASE ASYMPTOTIC RATE OF NESTEROV ACCELERATION FOR STRONGLY CONVEX OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study Nesterov's accelerated gradient method with its standard constant parameters for smooth strongly convex optimization. We characterize the exact worst-case asymptotic factor for the squared distance to the minimizer. To this end, we construct candidate worst-case trajectories by combining planar spirals with a common contraction factor that satisfy a self-similarity condition. Each component rotates by a fixed angle at every iteration. The exact factor is the optimal value of a linear optimization problem over probability measures on a fixed disk, with constraints derived from smooth strongly convex interpolation. Every optimal measure has a common contraction factor, which is also the largest factor for which the angular interpolation conditions are feasible. We also show that this worst-case asymptotic rate can be realized by a *single* function in a finite-dimensional space whose iterates attain this factor. A matching upper bound holds uniformly over all functions in the class and all finite dimensions. Consequently, allowing the function, dimension, and initial point to vary with the iteration horizon does not increase the worst-case asymptotic factor. We also show that the exact factor is $1 - 4/(3\sqrt{\kappa}) + o(\kappa^{-1/2})$ as $\kappa \to \infty$, where $\kappa = L/\mu$ is the ratio of the gradient Lipschitz constant to the strong convexity parameter.

1 INTRODUCTION

We study Nesterov's accelerated gradient method (NAG) with constant parameters for smooth strongly convex optimization. Our goal is to determine its exact worst-case asymptotic convergence rate, measured by the squared distance to the minimizer.

Let $\kappa = L/\mu$ denote the condition number. The classical convergence guarantee gives an asymptotic squared-distance factor of at most $1 - 1/\sqrt{\kappa}$ (Nesterov, 2018, Theorem 2.2.1, specialized to scheme (2.2.22); Lessard et al., 2016, Proposition 12). The exact worst-case factor among quadratic objectives is $(1 - 1/\sqrt{\kappa})^2$ (Lessard et al., 2016, Proposition 1 and Appendix A). These bounds leave open how slowly the method can converge on a nonquadratic objective.

A function that nearly attains the worst-case error after t iterations need not remain worst-case as the method continues. We therefore distinguish between keeping the function and initial point fixed and allowing the function, dimension, and initial point to depend on the iteration count. We prove that both choices give the same worst-case asymptotic factor, attained by one fixed function and initial point.

1.1 CONTRIBUTION

Our main contributions are:

An optimization characterization of the exact rate. We characterize the exact worst-case asymptotic factor by a linear optimization problem over probability measures on a disk whose radius depends only on the condition number. Every optimal measure is supported on one circle, so its components have a common contraction factor. The value is also the largest factor for which 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.

angular interpolation conditions are feasible; these conditions ensure consistency of the sampled points, gradients, and function values with one function in the class.

Matching lower and upper bounds. We construct one function and initial point attaining the optimal asymptotic factor. A matching upper bound holds uniformly over all functions in the class, all finite dimensions, and all nonoptimal initial points. Thus, allowing the function, dimension, and initial point to vary with the iteration horizon does not increase the worst-case asymptotic factor.

Exact and asymptotic descriptions of the rate. We exhibit a condition number for which the exact asymptotic factor cannot be expressed by radicals. For large condition numbers, we obtain the explicit expansion $1 - 4/(3\sqrt{\kappa}) + o(\kappa^{-1/2})$ as $\kappa \to \infty$, proving that the leading constant $4/3$ in Gannot's upper bound is sharp.

Explicit worst-case examples. At $\kappa = 10$, we construct a worst-case trajectory consisting of one planar spiral. At $\kappa = 100$, our construction combines two planar spirals with different rotation angles.

1.2 Related work

The classical convergence guarantee for NAG can be improved without changing the algorithm. Lessard et al. (2016, Section 4.5) use integral quadratic constraints (IQCs) to compute tighter upper bounds through semidefinite programming, and Safavi et al. (2018) derive explicit bounds by solving the associated inequalities analytically. Gannot (2022, Proposition 2.13) obtains an analytic upper bound using frequency-domain inequalities. For squared distance, his bound already has the leading term $1 - 4/(3\sqrt{\kappa})$ as $\kappa \to \infty$; we prove that the constant $4/3$ is sharp. We extend this IQC and frequency-domain approach by constructing a finite separator for every contraction factor above the interpolation optimum. The resulting upper bound matches the rate attained by one fixed objective.

Lyapunov functions provide a related approach to convergence analysis. Sanz-Serna & Zygalakis (2021) construct such functions for Nesterov methods and connect them to Lyapunov functions for differential equations. More generally, searches over prescribed families of quadratic Lyapunov inequalities can be formulated as semidefinite programs (Taylor et al., 2018; Upadhyaya et al., 2025), an approach implemented in AutoLyap (Upadhyaya et al., 2026). The exactness guarantees for these searches concern the chosen family of Lyapunov inequalities.

Establishing sharpness requires a matching lower bound. Performance estimation formulates worst-case analysis after a fixed number of iterations as an optimization problem (Drori & Teboulle, 2014). The interpolation conditions of Taylor et al. (2017) make this formulation exact for smooth strongly convex functions and provide a way to construct matching examples. For the smooth convex version of Nesterov's method, Kim & Fessler (2018, Theorem 5.2) give an $O(t^{-3})$ upper bound on the smallest squared gradient norm evaluated through iteration t, under an initial distance bound. Du (2026, Theorem 7.2) constructs matching examples for $t \geq 7$ in dimension $t - 4$. We impose these interpolation conditions on every pair of samples along an infinite trajectory. One strongly convex function realizes this trajectory and attains the worst-case asymptotic squared-distance factor from a fixed initial point.

Interpolation can also construct trajectories that demonstrate failure to converge. Goujaud et al. (2023) use it to find periodic trajectories and identify parameter choices for which first-order methods, including Nesterov acceleration, fail to converge. Regis (2026) studies related behavior by characterizing fixed points, two-point cycles, and their stability for heavy-ball and Nesterov momentum. For the Nesterov parameters considered here, convergence is already guaranteed. The trajectories we construct contract while rotating, and their decay determines the worst-case asymptotic factor.

The behavior of these examples depends on the function class. For quadratic objectives, Hagedorn & Jarre (2024) derive iteration-complexity bounds for Nesterov's and Polyak's methods by analyzing their linear recurrences. For twice continuously differentiable objectives, once convergence is established, linearization near the minimizer connects asymptotic convergence to a quadratic model (Van Scoy & Lessard, 2025, Lemma 3). Feng & Yuan (2026) establish local linear convergence of Nesterov's method from sufficiently close initial points under a local Polyak–Łojasiewicz condition and twice continuous differentiability. They obtain an optimal local rate by tuning the parameters. Our function class assumes a Lipschitz continuous gradient but does not require twice continuous

differentiability. The matching nonquadratic examples show that restricting this class to quadratic objectives can miss the worst-case asymptotic factor.

The choice of algorithm parameters matters as well. For the smooth convex version of Nesterov's method, which uses a varying momentum coefficient, Jang & Ryu (2025) prove that the iterates converge to a minimizer even when the minimizer is not unique. He & Fang (2026) study parameter sequences that combine accelerated behavior in the early iterations with a strongly convex asymptotic guarantee. Improved guarantees can also come from modifying the update itself. Park et al. (2023) analyze the optimized gradient method using Lyapunov functions and linear coupling, and introduce a strongly convex variant. This variant changes the momentum coefficient and adds a gradient correction, yielding a better function-value guarantee than the classical Nesterov bound. For strongly convex objectives, Lan et al. (2026, Section 5) develop regularization and restart methods with improved iteration-complexity guarantees for finding a small gradient. We keep the prescribed Nesterov step size and momentum coefficient fixed and determine the exact worst-case asymptotic rate of that iteration. Corollary 4.4 gives the corresponding exact factor for the gradient norm at the current iterate.

Paper organization. Section 2 defines the function class, algorithm, and two asymptotic rates. Section 3 gives the angular interpolation conditions, defines the rate through a disk LP, and proves the common-circle reduction. Section 4 states the main theorem and outlines the matching lower and upper bounds. Section 5 presents a worst-case example at $\kappa = 10$. The appendices provide the full proofs, additional examples, numerical results, and an algebraic obstruction to formulas by radicals.

2 Problem setting and rate definitions

We work in $\mathbf{R}^d$ with the Euclidean inner product and norm. For $0 < \mu \leq L$, let $\mathcal{F}_{\mu,L}$ denote the class of differentiable, μ-strongly convex functions with L-Lipschitz gradients. Dividing the objective by μ gives a function in $\mathcal{F}_{1,\kappa}$, where $\kappa = L/\mu$ is the condition number. This scaling preserves the NAG iterates when the step size changes from $1/L$ to $1/\kappa$. Every function in this class has a unique minimizer, denoted by $x_\star$.

For $\kappa \geq 1$, we consider the NAG iteration

$$y_t = x_t + \beta_\kappa(x_t - x_{t-1}), \qquad x_{t+1} = y_t - \frac{1}{\kappa}\nabla f(y_t), \qquad \beta_\kappa = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}, \tag{1}$$

initialized with $x_{-1} = x_0 \neq x_\star$. Here y_t is the point where the gradient is evaluated, and x_t is the iterate whose distance to the minimizer we measure. The step size $1/\kappa$ and momentum coefficient β_κ remain constant throughout the iteration.

For a fixed function and initial point, define the normalized squared error

$$R_t(f, x_0) = \frac{\|x_t - x_\star\|^2}{\|x_0 - x_\star\|^2}.$$

The quantity $\limsup_{t\to\infty} R_t(f, x_0)^{1/t}$ measures its asymptotic geometric decay. For example, if $R_t = q^t$, this quantity equals q. Throughout, convergence factors refer to squared distance. We compare the *fixed-objective asymptotic factor* and the *horizon-dependent asymptotic factor*:

$$\rho_{\mathrm{fix}}(\kappa) = \sup_{\substack{d\geq 1,\ f\in\mathcal{F}_{1,\kappa}(\mathbf{R}^d)\\ x_0\neq x_\star}} \limsup_{t\to\infty} R_t(f, x_0)^{1/t}, \tag{2}$$

$$\rho_{\mathrm{hor}}(\kappa) = \limsup_{t\to\infty} \left(\sup_{\substack{d\geq 1,\ f\in\mathcal{F}_{1,\kappa}(\mathbf{R}^d)\\ x_0\neq x_\star}} R_t(f, x_0) \right)^{1/t}. \tag{3}$$

In $\rho_{\mathrm{fix}}(\kappa)$, we first compute the asymptotic factor for each fixed function and initial point, then take the supremum over all such choices and dimensions. In $\rho_{\mathrm{hor}}(\kappa)$, we first maximize the error at each iteration t, allowing the function, dimension, and initial point to depend on t, then take the asymptotic factor. Hence $\rho_{\mathrm{fix}}(\kappa) \leq \rho_{\mathrm{hor}}(\kappa)$. Our main theorem proves equality and, for each κ, attainment by a *single fixed function* and initial point, both independent of t.

When $\kappa = 1$, strong convexity and smoothness imply $\nabla f(x) = x - x_\star$. The method reaches the minimizer in one iteration and remains there, so $\rho_{\mathrm{fix}}(1) = \rho_{\mathrm{hor}}(1) = 0$. We therefore focus on $\kappa > 1$.

The quadratic $f(x) = \frac{1}{2}x^2$ belongs to $\mathcal{F}_{1,\kappa}$. On this objective, the NAG iteration becomes

$$x_{t+1} = \left(1 - \frac{1}{\kappa}\right)\big((1+\beta_\kappa)x_t - \beta_\kappa x_{t-1}\big).$$

Its characteristic polynomial is $\big(\lambda - (1 - 1/\sqrt{\kappa})\big)^2$, giving the asymptotic squared-distance factor

$$q_0(s) = \left(1 - \frac{1}{s}\right)^2, \qquad s = \sqrt{\kappa}. \tag{4}$$

Together with Nesterov's classical bound, this gives

$$q_0(s) \le \rho_{\mathrm{fix}}(\kappa) \le \rho_{\mathrm{hor}}(\kappa) \le 1 - \frac{1}{s}, \qquad s = \sqrt{\kappa}.$$

We determine both asymptotic factors exactly over the full class $\mathcal{F}_{1,\kappa}$.

3 Interpolation conditions for contracting trajectories

We construct candidate worst-case trajectories whose components contract and rotate by fixed amounts at each iteration. We first state the conditions under which prescribed query points and gradients come from one function in $\mathcal{F}_{1,\kappa}$. We then express these conditions in terms of the contraction factor, rotation angles, and component weights. We use these conditions to define the feasible factors, then give a disk LP whose value is their largest element. Section 4 identifies this value with the exact worst-case rate.

3.1 Exact interpolation and cyclic monotonicity

Fix $\kappa > 1$. A finite collection (y_i, g_i, f_i) is *interpolable* by $\mathcal{F}_{1,\kappa}$ if there exists one $f \in \mathcal{F}_{1,\kappa}$ satisfying $\nabla f(y_i) = g_i$ and $f(y_i) = f_i$ for every index i. The exact interpolation theorem (Taylor et al., 2017, Theorem 4) states that this holds if and only if every ordered pair satisfies

$$f_i \ge f_j + \langle g_j, y_i - y_j\rangle + \frac{1}{2\kappa}\|g_i - g_j\|^2 + \frac{\kappa}{2(\kappa-1)}\left\|y_i - y_j - \frac{1}{\kappa}(g_i - g_j)\right\|^2. \tag{5}$$

An interpolation inequality is *active* when it holds with equality. To construct a trajectory, we prescribe the query points and gradients and leave the function values free. The following change of variables eliminates the function values from the interpolation conditions. Write $s = \sqrt{\kappa}$ and translate the minimizer to the origin. Set

$$h(y) = f(y) - \frac{1}{2}\|y\|^2, \qquad \sigma = \nabla h(y), \qquad r = y - \frac{\sigma}{\kappa - 1}. \tag{6}$$

Subtracting the quadratic term makes h convex and $(\kappa-1)$-smooth. Its convex conjugate h^* is $1/(\kappa-1)$-strongly convex, so

$$\varphi(\sigma) = h^*(\sigma) - \frac{\|\sigma\|^2}{2(\kappa-1)} \tag{7}$$

is a proper closed convex function, possibly taking the value $+\infty$. The identity $\sigma = \nabla h(y)$ gives $y \in \partial h^*(\sigma)$ and hence $r \in \partial\varphi(\sigma)$. Here ∂ denotes the convex subdifferential. Conversely, a set of pairs (σ_i, r_i) lies in the subdifferential of one proper closed convex function if and only if it is *cyclically monotone*: for every finite cycle $i_0, i_1, \dots, i_m = i_0$,

$$\sum_{j=0}^{m-1} \langle r_{i_j}, \sigma_{i_j} - \sigma_{i_{j+1}}\rangle \ge 0.$$

Rockafellar's cyclic monotonicity theorem guarantees the existence of such a function φ (Rockafellar, 1970). Reversing the change of variables gives an interpolating function in $\mathcal{F}_{1,\kappa}$:

$$f(y) = \left(\varphi + \frac{\|\cdot\|^2}{2(\kappa-1)}\right)^*(y) + \frac{1}{2}\|y\|^2.$$

Thus cyclic monotonicity characterizes whether the prescribed points and gradients admit a function in $\mathcal{F}_{1,\kappa}$. The resulting function determines the previously unspecified values f_i.

3.2 CONTRACTING AND ROTATING TRAJECTORIES

Consider a sequence in $\mathbf{R}^2$ obtained by applying the same contraction and rotation at every iteration. Identify $\mathbf{R}^2$ with the complex plane and set $z = \sqrt{q}\, e^{i\theta}$, where $0 < q < 1$. Multiplication by z contracts lengths by $\sqrt{q}$ and rotates directions by θ, so squared norms contract by q. Thus larger values of q correspond to slower convergence.

Applying (6) at y_t gives $\sigma_t = \nabla f(y_t) - y_t$ and $r_t = y_t - \sigma_t/(\kappa - 1)$. The gradient update gives

$$x_{t+1} = y_t - \frac{1}{\kappa}(y_t + \sigma_t) = \frac{\kappa - 1}{\kappa}\left(y_t - \frac{\sigma_t}{\kappa - 1}\right) = \frac{\kappa - 1}{\kappa} r_t.$$

Thus r_t is a constant multiple of x_{t+1}. We consider candidate trajectories for which these iterates contract and rotate by fixed amounts:

$$r_{t+1} = z r_t, \qquad r_t = a z^t,$$

where $a \neq 0$ specifies the initial magnitude and direction. We first take $a = 1$, then combine components with different angles and amplitudes.

These sequences describe the trajectory after initialization. Lemma C.4 constructs an objective attaining the optimal factor with $x_{-1} = x_0$.

With $r_t = z^t$, we seek $\sigma_t = d z^t$. Using $y_t = r_t + \sigma_t/(\kappa - 1)$ and $x_t = (\kappa - 1) r_{t-1}/\kappa$ in the momentum update gives

$$\left(1 + \frac{d}{\kappa - 1}\right) z^2 = \frac{\kappa - 1}{\kappa}\big((1 + \beta_\kappa) z - \beta_\kappa\big).$$

Solving for d with $\kappa = s^2$ gives

$$r_t = z^t, \qquad \sigma_t = d_s(z) z^t, \qquad d_s(z) = -\frac{(s^2 - 1)(sz - (s - 1))^2}{s^2 z^2}. \tag{8}$$

For $q > 0$, this expression is defined at every angle. At $q = q_0(s)$ and $\theta = 0$, we have $d_s(z) = 0$, so the transformed gradient vanishes. This includes a geometric trajectory of the quadratic objective $f(x) = \|x\|^2/2$. The interpolation conditions below determine which candidate trajectories come from one function in $\mathcal{F}_{1,\kappa}$.

3.3 INTERPOLATION CONDITIONS

We apply the cyclic monotonicity conditions to the contracting and rotating trajectories from Section 3.2. Fix $s > 1$ and $q \in (0, 1)$. Write $z = \sqrt{q}\, e^{i\theta}$, where $0 \leq \theta \leq \pi$. For one such trajectory, define the terms appearing in these inequalities by

$$c_{ij}(\theta) = \langle r_i, \sigma_i - \sigma_j \rangle = \mathrm{Re}\left[\overline{d_s(z)} z^i \left(\bar{z}^i - \bar{z}^j\right)\right], \qquad i, j \geq 0. \tag{9}$$

Here Re denotes the real part, and an overbar on a complex number denotes complex conjugation. The inner products are unchanged when θ is replaced by $-\theta$, so it suffices to consider $\theta \in [0, \pi]$.

We combine m contracting and rotating trajectories with the same factor q. Scaling all components by the same positive constant preserves the interpolation inequalities and convergence factor, so we normalize the weights to sum to one. Choose angles $\theta_1, \ldots, \theta_m$ and positive weights $w_1, \ldots, w_m$ with $\sum_{k=1}^m w_k = 1$. Set $z_k = \sqrt{q}\, e^{i\theta_k}$ and

$$r_t = (\sqrt{w_k}\, z_k^t)_{k=1}^m, \qquad \sigma_t = (\sqrt{w_k}\, d_s(z_k) z_k^t)_{k=1}^m.$$

The inner products add across coordinates:

$$\langle r_i, \sigma_i - \sigma_j \rangle = \sum_{k=1}^m w_k\, c_{ij}(\theta_k).$$

Each complex coordinate represents two real coordinates. An angle of 0 or π requires only one.

This finite collection of modes can be encoded by the finitely supported probability measure $\nu = \sum_{k=1}^{m} w_k \delta_{\theta_k}$, where δ_θ is the unit point mass at θ. Its weighted inner products are then

$$\bar{c}_{ij}(\nu) = \int_0^\pi c_{ij}(\theta)\, d\nu(\theta).$$

For the variational formulation, we now enlarge the domain from finitely supported measures to all probability measures on $[0, \pi]$, denoted by $\mathcal{P}([0, \pi])$. For a general ν in this set, the same integral defines $\bar{c}_{ij}(\nu)$; the interpolation inequalities below still determine feasibility. This enlargement permits compactness arguments without assuming a bound on the number of modes; an infinite-support measure does not itself specify a finite-dimensional modal construction. At the largest feasible factor, if it exceeds $q_0(s)$, Lemma C.3 guarantees that a feasible angular measure can nevertheless be chosen with finite support.

We seek function values φ_i satisfying

$$\varphi_i - \varphi_j \le \bar{c}_{ij}(\nu), \qquad i, j \ge 0.$$

Shifting both indices gives $\bar{c}_{i+t,j+t}(\nu) = q^t \bar{c}_{ij}(\nu)$. Lemma 3.1 shows that, whenever these inequalities are feasible, we can choose $\varphi_i = q^i \varphi_0$ for some scalar φ_0. The inequalities for $(0, t)$ and $(t, 0)$ give the following equivalent conditions.

Lemma 3.1 (Equivalent interpolation conditions)**.** *For a probability measure ν on $[0, \pi]$, the cyclic monotonicity inequalities hold if and only if there is a real scalar φ_0 satisfying*

$$-\bar{c}_{t0}(\nu) \le (1 - q^t)\varphi_0 \le \bar{c}_{0t}(\nu), \qquad \textit{for every integer } t \ge 1. \tag{10}$$

In that case, we may choose $\varphi_i = q^i \varphi_0$.

The proof is given in Appendix B. The averaged terms $\bar{c}_{ij}(\nu)$ depend on s and q through $d_s(z)$ and $z = \sqrt{q}\, e^{i\theta}$.

For fixed $s > 1$, let $\mathcal{Q}(s)$ denote the set of *feasible factors* $q \in (0, 1)$ for which (10) admits an angular probability measure $\nu \in \mathcal{P}([0, \pi])$ and a real scalar φ_0. The choice $q = q_0(s)$, $\nu = \delta_0$, and $\varphi_0 = 0$ is feasible. Lemma B.1 shows that $\mathcal{Q}(s)$ has a largest element and every feasible factor is at most $1 - 1/s$. For fixed q, the constraints are linear in ν and φ_0. They impose interpolation for all iteration indices.

A linear rate formulation and common-circle reduction. In (10), all components share a contraction factor q, and the angular measure ν specifies their rotation angles. We now allow both the contraction factors and angles to vary, represented by a probability measure μ on the disk $D_s = \{z \in \mathbb{C} : |z|^2 \le 1 - 1/s\}$. The objective is the average squared contraction factor, $\int |z|^2\, d\mu(z)$.

Let Γ contain all finite cycles of iteration indices $\gamma = (n_0, \ldots, n_m = n_0)$, with $m \ge 1$ and nonnegative integer indices. Indices may repeat, and no index is required to be zero. For each cycle, define the scaled sum of the interpolation terms in (9) by

$$\mathcal{C}_\gamma^s(z) = \frac{q^2}{s^2 - 1} \sum_{j=0}^{m-1} c_{n_j, n_{j+1}}(\theta), \qquad z = \sqrt{q}\, e^{i\theta}, \quad q > 0. \tag{11}$$

Here $c_{ij}(\theta)$ is evaluated at $q = |z|^2$. Define $\mathcal{C}_\gamma^s$ on the lower half of the disk by $\mathcal{C}_\gamma^s(\bar{z}) = \mathcal{C}_\gamma^s(z)$, and set $\mathcal{C}_\gamma^s(0) = 0$. On each circle, the positive scaling preserves the cyclic monotonicity inequalities. The coefficients are continuous on the whole disk; Appendix B.2 gives their polynomial expressions.

For $s > 1$, define $q_\star(s)$ by the linear optimization problem

$$\begin{aligned} q_\star(s) = \quad & \text{maximize} \quad \int |z|^2\, d\mu(z) \\ & \text{subject to} \quad \int \mathcal{C}_\gamma^s(z)\, d\mu(z) \ge 0, \quad \gamma \in \Gamma. \end{aligned} \tag{12}$$

where the optimization variable μ is a probability measure on D_s.

Proposition 3.2 (Common-circle reduction)**.** *For $s > 1$, problem* (12) *attains its maximum, and every optimal measure is supported on the circle $|z|^2 = q_\star(s)$. For each $q \in (0, 1)$, membership in $\mathcal{Q}(s)$ is equivalent to the existence of a feasible disk measure supported on $|z|^2 = q$. Consequently,*

$$q_\star(s) = \max \mathcal{Q}(s), \qquad q_0(s) \le q_\star(s) \le 1 - 1/s. \tag{13}$$

An *optimal angular pair* is a pair (ν, φ_0) satisfying (10) at $q = q_\star(s)$. Thus every optimal disk measure uses a common contraction factor, although feasible disk measures may mix radii. For such mixed-radius measures, the objective is an average, not an asserted asymptotic trajectory factor. The objective and constraints are linear in μ. For a feasible measure with positive objective, multiplying its weights by $|z|^2$ and normalizing preserves feasibility. This strictly increases the objective unless $|z|^2$ is constant on the support. This argument uses all cycles, including those obtained by shifting every index by the same amount. Appendix B.2 gives the proof and relaxations that provide upper bounds. The disk LP remains infinite-dimensional; for trajectory constructions and numerical evaluation we use the angular conditions (10) and the identity (13). The next section identifies $q_\star(s)$ with the worst-case asymptotic factor over the full function class.

4 Exact rate theorem

Theorem 4.1 (Exact asymptotic rate)**.** *For every $\kappa > 1$, the two worst-case asymptotic factors defined in Section 2 satisfy*

$$\rho_{\mathrm{fix}}(\kappa) = \rho_{\mathrm{hor}}(\kappa) = q_\star(\sqrt{\kappa}). \tag{14}$$

This factor is attained by a single function in a finite-dimensional space and an initial point, both independent of the iteration count. The exact rate satisfies

$$\left(1 - \frac{1}{\sqrt{\kappa}}\right)^2 \le q_\star(\sqrt{\kappa}) \le 1 - \frac{1}{\sqrt{\kappa}}. \tag{15}$$

In particular, the rate tends to zero as $\kappa \downarrow 1$ and to one as $\kappa \to \infty$.

Proof sketch. Set $s = \sqrt{\kappa} > 1$. For the lower bound, suppose $q_\star(s) > q_0(s)$. An optimal angular pair can then be chosen with a finitely supported measure, so the construction in Section 3 uses finitely many planar trajectories. We construct a function in $\mathcal{F}_{1,\kappa}$ for which Nesterov's method, initialized with $x_{-1} = x_0$, attains this asymptotic factor. If $q_\star(s) = q_0(s)$, the quadratic $f(x) = x^2/2$ attains the factor. For the upper bound, fix $q \in (q_\star(s), 1)$. The interpolation conditions are infeasible at this factor. A separation argument gives a nonnegative combination of finitely many cyclic monotonicity inequalities. We turn this combination into an IQC multiplier for the exponentially rescaled recurrence, following the control-theoretic approach of Lessard et al. (2016); Safavi et al. (2018); Gannot (2022). The classical strict-IQC theorem (Zhang et al., 2019, Theorem 3), together with a direct performance estimate, then yields

$$\|x_t - x_\star\|^2 \le C q^t \|x_0 - x_\star\|^2,$$

where C depends only on s and q. Taking tth roots and letting $t \to \infty$ gives $\rho_{\mathrm{hor}}(s^2) \le q$. Letting $q \downarrow q_\star(s)$ proves the upper bound. The full proof is given in Appendix C. □

Relation to Gannot's upper bound. Proposition 2.13 of Gannot (2022) gives an upper bound $q_{\mathrm{G}}(\kappa)$ on the asymptotic squared-distance factor. At $\kappa = 10$ and $\kappa = 100$, we determine the exact factor using exact arithmetic. Numerical evaluation of Gannot's upper bound gives a larger value in both cases. The differences reported in Table 2 are numerical approximations.

Remark 4.2 (No general formula by radicals). At $\kappa = (1421/1250)^2$, the exact factor $q_\star(1421/1250)$ has a minimal polynomial of degree 20 over $\mathbb{Q}$, with Galois group S_{20}. Consequently, no formula built from rational functions of s with rational coefficients and finitely many radicals can be defined and valid for every $s > 1$. Appendix G gives the polynomial and its exact certificate. This does not exclude other notions of closed form.

Corollary 4.3 (Asymptotic rate for large condition numbers)**.** *The exact factor satisfies*

$$\lim_{\kappa\to\infty} \sqrt{\kappa}\left(1 - q_\star(\sqrt{\kappa})\right) = \frac{4}{3}, \qquad q_\star(\sqrt{\kappa}) = 1 - \frac{4}{3\sqrt{\kappa}} + o(\kappa^{-1/2}). \tag{16}$$

The corresponding asymptotic factor for the distance $\|x_t - x_\star\|$ is $1 - 2/(3\sqrt{\kappa}) + o(\kappa^{-1/2})$.

Proof. The necessary condition $\bar{c}_{01}(\nu) \ge 0$ gives an upper bound on the exact factor. For the lower bound, fix $2/3 < \alpha < 3/4$. For all sufficiently large κ, we construct a probability measure satisfying every interpolation inequality with $\varphi_0 = 0$ at

$$q = \exp(-2\alpha/\sqrt{\kappa}).$$

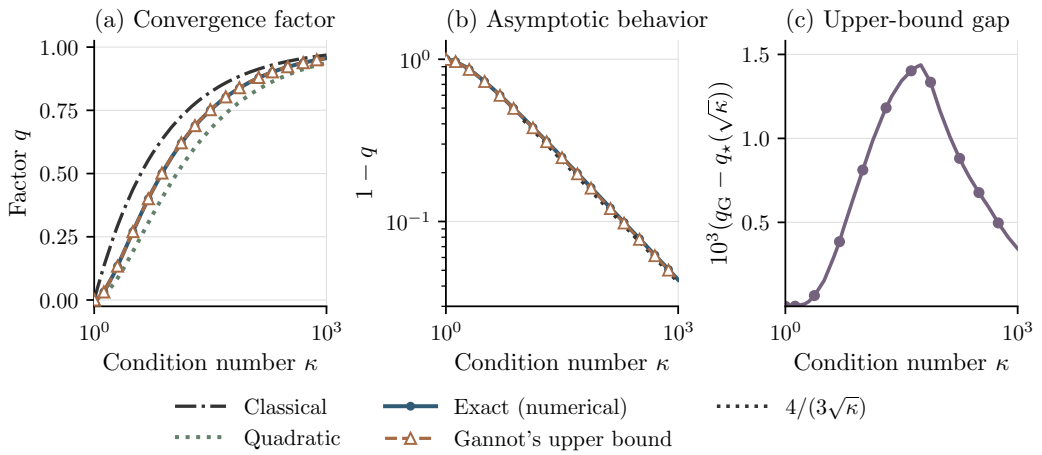


Figure 1: Convergence factors for $1 \le \kappa \le 1000$. (a) Numerical evaluations of $q_\star(\sqrt{\kappa})$ using the angular conditions (10) and (13), compared with the quadratic rate and the classical and Gannot upper bounds. At $\kappa = 1$ we plot the limiting value zero; exact arithmetic certifies the values at $\kappa = 10$ and 100. The remaining values are numerical approximations. Filled circles show numerical evaluations of our rate formula, and hollow triangles show Gannot's upper bound. (b) Comparison of $1 - q$ with the asymptotic expression $4/(3\sqrt{\kappa})$ from Corollary 4.3. (c) The difference between Gannot's upper bound and the evaluated exact rate. The largest displayed gap is 1.44×10^{-3} (less than 0.18% of Gannot's upper bound).

Letting $\alpha \downarrow 2/3$ gives the matching coefficient $4/3$. The construction and full proof are given in Appendix D. □

Figure 1 compares numerical evaluations of the rate formula with Gannot's upper bound for $1 \le \kappa \le 1000$. We compute Gannot's upper bound from the smaller positive root of the cubic in (Gannot, 2022, Proposition 2.13). The two curves nearly coincide at this scale; panel (c) shows their difference. The cubic defining Gannot's upper bound also arises from a subset of our interpolation inequalities involving consecutive iteration indices and the minimizer. Keeping only these inequalities enlarges the set of feasible factors $\mathcal{Q}(s)$ and therefore gives an upper bound on the exact factor by (13). Selected numerical values and details of their verification are given in Appendix H.

Corollary 4.4 (Function gap and gradient norm)**.** *Replacing $R_t(f, x_0)$ in the two asymptotic factor definitions* (2)–(3) *by either*

$$\frac{f(x_t) - f(x_\star)}{f(x_0) - f(x_\star)} \qquad \textit{or} \qquad \frac{\|\nabla f(x_t)\|^2}{\|\nabla f(x_0)\|^2}$$

gives the same exact factor as in Theorem 4.1. The asymptotic factors for the distance $\|x_t - x_\star\|$ and the gradient norm $\|\nabla f(x_t)\|$ equal the square root of this factor.

Proof. For every $f \in \mathcal{F}_{1,\kappa}$ and x, strong convexity and smoothness give

$$\frac{1}{2}\|x - x_\star\|^2 \le f(x) - f(x_\star) \le \frac{\kappa}{2}\|x - x_\star\|^2, \qquad \|x - x_\star\|^2 \le \|\nabla f(x)\|^2 \le \kappa^2\|x - x_\star\|^2.$$

Dividing by the corresponding initial values bounds the first ratio between $\kappa^{-1}R_t$ and κR_t, and the second between $\kappa^{-2}R_t$ and $\kappa^2 R_t$. These constants depend only on κ, so the bounds remain valid when taking the supremum over dimensions, functions, and initial points. Their tth roots tend to one, which proves equality of the asymptotic factors in both definitions. □

Classical convergence guarantees also give upper bounds on the gradient norm through strong convexity and smoothness (Nesterov, 2018; Lessard et al., 2016). Corollary 4.4 identifies its exact asymptotic factor, attained by the same fixed function and initialization as the distance factor. Equality of these asymptotic factors does not imply equality of finite-iteration bounds.

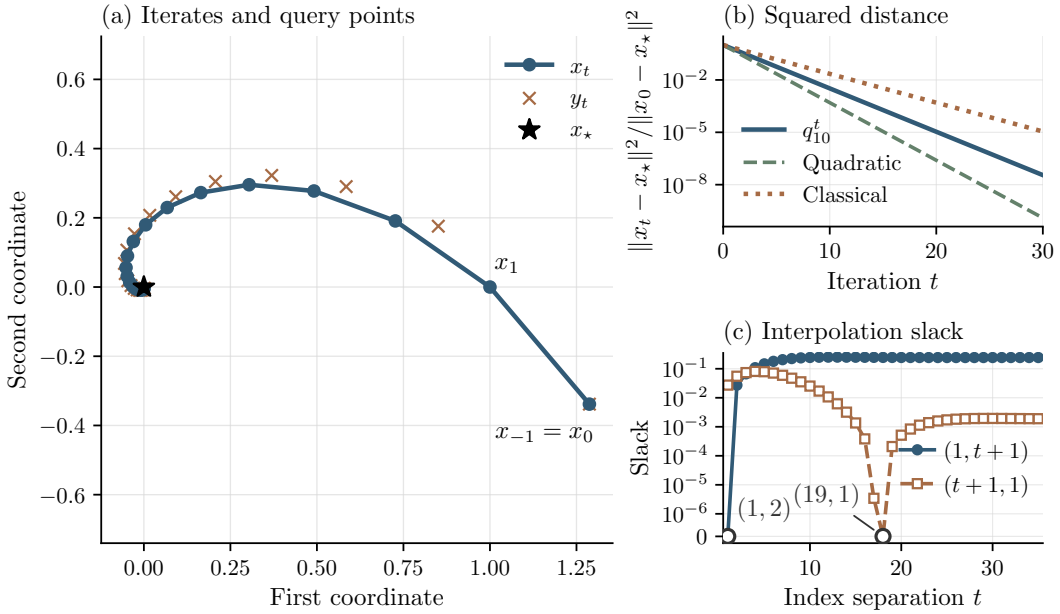


Figure 2: The exact $\kappa = 10$ example. (a) Iterates and query points from the standard initialization. (b) The normalized squared distance equals q_{10}^t, compared with the powers of the quadratic and classical factors. (c) Slack in the interpolation inequalities for the ordered pairs $(1, t+1)$ and $(t+1, 1)$. Equality holds at $(1, 2)$ and $(19, 1)$. The vertical axis is logarithmic above 10^{-6} and linear near zero. Every comparison involving the initialization query is strict.

5 A WORST-CASE EXAMPLE AT $\kappa = 10$

At $\kappa = 10$, an optimal angular measure is concentrated at a single angle in $(0, \pi)$. The corresponding trajectory therefore lies in $\mathbf{R}^2$. Identifying $\mathbf{R}^2$ with the complex plane, the optimal factor and corresponding complex multiplier are

$$q_{10} = q_\star(\sqrt{10}) \approx 0.564423,$$
$$z_\star \approx 0.726601 + 0.190984\, i, \qquad |z_\star|^2 = q_{10}.$$

The exact algebraic definitions are given in Appendix E.

There exists a function $f \in \mathcal{F}_{1,10}(\mathbf{R}^2)$, with minimizer at the origin, for which Nesterov's method produces

$$x_{-1} = x_0 = z_\star^{-1}, \qquad x_t = z_\star^{t-1}, \quad t \geq 1. \tag{17}$$

Each iteration rotates the error by the angle $\theta_\star = \arg z_\star \approx 0.257032$ radians and multiplies its squared norm by q_{10}. Therefore,

$$\frac{\|x_t - x_\star\|^2}{\|x_0 - x_\star\|^2} = q_{10}^t, \qquad t \geq 0. \tag{18}$$

Appendix E gives formulas for the query points, gradients, and function values. These data satisfy the interpolation inequalities (5) for every ordered pair of indices.

For this example, the inequalities are active for the ordered pairs

$$(i, j) = (t, t+1) \qquad \text{and} \qquad (i, j) = (t+18, t), \qquad t \geq 1.$$

All inequalities comparing the initial query point y_0 with a later query point are strict. The matching upper bound uses interpolation inequalities between samples at most eighteen iterations apart. It proves that the worst-case asymptotic factor is at most q_{10}. Independent checks using rigorous interval arithmetic and exact integer arithmetic verify the algebraic solution, all interpolation inequalities, and positive semidefiniteness of the matrix used for the upper bound.

AI USE STATEMENT

Generative AI tools assisted with literature comparison, formulation and revision of mathematical arguments, proof writing, implementation of verification and plotting scripts, formalization, and manuscript editing. The associated checks include symbolic calculations, numerical and adversarial tests, and machine-checked proofs for the results covered by the formalization. The authors have reviewed all AI-assisted work, verified the claims, and take responsibility for the final manuscript.

REFERENCES

Yoel Drori and Marc Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1–2):451–482, 2014. doi: 10.1007/s10107-013-0653-0.

Yi Du. When a relaxed PEP is exact: The sharp queried-gradient rate of Nesterov's fast gradient method, 2026.

Zixu Feng and Hao Yuan. Optimal local linear convergence of Nesterov's accelerated gradient method for C^2 functions under the Polyak–Łojasiewicz inequality, 2026.

Oran Gannot. A frequency-domain analysis of inexact gradient methods. *Mathematical Programming*, 194(1–2):975–1016, 2022. doi: 10.1007/s10107-021-01665-8. arXiv:1912.13494.

Baptiste Goujaud, Aymeric Dieuleveut, and Adrien B. Taylor. Counter-examples in first-order optimization: A constructive approach. *IEEE Control Systems Letters*, 7:2485–2490, 2023. doi: 10.1109/LCSYS.2023.3286277. arXiv:2303.10503.

Melinda Hagedorn and Florian Jarre. Iteration complexity of fixed-step methods by Nesterov and Polyak for convex quadratic functions. *Journal of Optimization Theory and Applications*, 202: 456–474, 2024. doi: 10.1007/s10957-023-02261-w. arXiv:2211.10234.

Xin He and Ya-Ping Fang. A unified continuous-discrete framework for Nesterov acceleration: Transitions between convex and strongly convex regimes, 2026.

Uijeong Jang and Ernest K. Ryu. Point convergence of Nesterov's accelerated gradient method: An AI-assisted proof, 2025.

Donghwan Kim and Jeffrey A. Fessler. Generalizing the optimized gradient method for smooth convex minimization. *SIAM Journal on Optimization*, 28(2):1920–1950, 2018. doi: 10.1137/17M112124X.

Guanghui Lan, Yuyuan Ouyang, and Zhe Zhang. Optimal and parameter-free gradient minimization methods for convex and nonconvex optimization. *Mathematical Programming*, 2026. doi: 10.1007/s10107-026-02352-2.

Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95, 2016. doi: 10.1137/15M1009597.

Alexandre Megretski and Anders Rantzer. System analysis via integral quadratic constraints. *IEEE Transactions on Automatic Control*, 42(6):819–830, 1997. doi: 10.1109/9.587335.

James S. Milne. Fields and Galois theory, 2022. URL `https://www.jmilne.org/math/CourseNotes/FT.pdf`. Version 5.10, September 2022.

Yurii Nesterov. *Lectures on Convex Optimization*, volume 137 of *Springer Optimization and Its Applications*. Springer, 2 edition, 2018. ISBN 978-3-319-91578-4. doi: 10.1007/978-3-319-91578-4.

Chanwoo Park, Jisun Park, and Ernest K. Ryu. Factor-$\sqrt{2}$ acceleration of accelerated gradient methods. *Applied Mathematics & Optimization*, 88:77, 2023. doi: 10.1007/s00245-023-10047-9. arXiv:2102.07366.

Eric Regis. Variational structure at the edge of stability, 2026.

R. Tyrrell Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematical Series*. Princeton University Press, 1970.

Sam Safavi, Bikash Joshi, Guilherme França, and José Bento. An explicit convergence rate for Nesterov's method from SDP. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1560–1564, 2018. doi: 10.1109/ISIT.2018.8437794. arXiv:1801.04492.

J. M. Sanz-Serna and Konstantinos C. Zygalakis. The connections between Lyapunov functions for some optimization algorithms and differential equations. *SIAM Journal on Numerical Analysis*, 59 (3):1542–1565, 2021. doi: 10.1137/20M1364138. arXiv:2009.00673.

Elias M. Stein and Rami Shakarchi. *Complex Analysis*, volume 2 of *Princeton Lectures in Analysis*. Princeton University Press, 2003. ISBN 978-0-691-11385-2.

Ian Stewart. *Galois Theory*. Chapman and Hall/CRC, Boca Raton, 5 edition, 2022. ISBN 9781003213949. doi: 10.1201/9781003213949.

Adrien B. Taylor, Julien M. Hendrickx, and François Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161(1–2): 307–345, 2017. doi: 10.1007/s10107-016-1009-3. arXiv:1502.05666.

Adrien B. Taylor, Bryan Van Scoy, and Laurent Lessard. Lyapunov functions for first-order methods: Tight automated convergence guarantees. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4897–4906, 2018.

Manu Upadhyaya, Sebastian Banert, Adrien B. Taylor, and Pontus Giselsson. Automated tight Lyapunov analysis for first-order methods. *Mathematical Programming*, 209(1–2):133–170, 2025. doi: 10.1007/s10107-024-02061-8. arXiv:2302.06713.

Manu Upadhyaya, Shuvomoy Das Gupta, Adrien B. Taylor, Sebastian Banert, and Pontus Giselsson. The AutoLyap software suite for computer-assisted Lyapunov analyses of first-order methods, 2026.

Bryan Van Scoy and Laurent Lessard. The fastest known first-order method for minimizing twice continuously differentiable smooth strongly convex functions. *IEEE Control Systems Letters*, 9: 655–660, 2025. doi: 10.1109/LCSYS.2025.3576681. arXiv:2506.01168.

Jingfan Zhang, Peter Seiler, and Joaquin Carrasco. Noncausal FIR Zames–Falb multiplier search for exponential convergence rate, 2019. URL `https://arxiv.org/abs/1902.09473v1`. arXiv:1902.09473v1; Definitions 1 and 5 and Theorem 3.

A Proof roadmap and dependencies

The proof of Theorem 4.1 has four steps. First, we eliminate the function values from the interpolation conditions (5) using cyclic monotonicity. Second, we construct one objective and initial point whose Nesterov trajectory satisfies $x_{-1} = x_0$ and attains the asymptotic factor $q_\star(s)$. Third, for any $q \in (q_\star(s), 1)$, we find a nonnegative combination of finitely many cycle functions from (11) that is strictly negative at every angle. We call this combination a *strict cycle separator*. Fourth, we use this separator to prove $\|x_t - x_\star\|^2 \leq b(s,q)q^t\|x_0 - x_\star\|^2$. The constant $b(s,q)$ is independent of the objective, dimension, initialization, and iteration count.

Appendix C.3 introduces integral quadratic constraints (IQCs) and uses them to derive the uniform upper bound from a strict cycle separator. It states the stability theorem and verifies its hypotheses for the rescaled Nesterov recurrence.

When $q_\star(s) > q_0(s)$, Lemma C.3 proves that an optimal angular measure can be chosen with finite support. The proof uses Montel's theorem (Stein & Shakarchi, 2003, Chapter 8, Theorem 3.3) to obtain a convergent subsequence of weighted sums of cycle functions. The limit is nonpositive on $[0, \pi]$, while feasibility requires its integral against the optimal measure to be nonnegative. Hence the measure is supported on the zeros of the limit. The proof shows that there are only finitely many such zeros. The argument assumes no bound on the number of support points and allows comparisons between iterations arbitrarily far apart.

B PROOF OF THE INTERPOLATION CONDITIONS

Proof of Lemma 3.1. Fix $s > 1$, $q \in (0, 1)$, and a probability measure ν on $[0, \pi]$. To prove the forward implication, we construct interpolation values of the form $\varphi_i = q^i \varphi_0$. We first solve the pairwise inequalities, then average shifted and rescaled solutions to obtain this form. Consider the directed graph with vertices $0, 1, \ldots$ and weight $\bar{c}_{ij}(\nu)$ on each edge $i \to j$. The weight of a path or cycle is the sum of its edge weights. Assume cyclic monotonicity, so every finite directed cycle has nonnegative weight.

For $j \geq 0$, let u_j be the infimum of the weights of all finite directed paths from 0 to j. The direct edge $0 \to j$ gives $u_j \leq \bar{c}_{0j}(\nu)$. Closing any such path with the edge $j \to 0$ gives $u_j \geq -\bar{c}_{j0}(\nu)$. Thus every u_j is finite and $u_0 = 0$. Appending the edge $i \to j$ to paths from 0 to i and taking infima gives

$$u_j - u_i \leq \bar{c}_{ij}(\nu), \qquad i, j \geq 0. \tag{19}$$

Conversely, summing (19) around any finite directed cycle cancels the u_i terms and gives a nonnegative total edge weight.

Let $\mathcal{U}$ be the set of real sequences $u = (u_i)_{i \geq 0}$ satisfying (19) and $u_0 = 0$. The preceding construction shows that $\mathcal{U}$ is nonempty. It is convex because its constraints are linear. Every $u \in \mathcal{U}$ satisfies

$$-\bar{c}_{i0}(\nu) \leq u_i \leq \bar{c}_{0i}(\nu), \qquad i \geq 0.$$

Each coordinate therefore lies in a compact interval. A diagonal subsequence argument shows that every sequence of elements of $\mathcal{U}$ has a subsequence converging in every coordinate. The limit satisfies the same inequalities and normalization, so it belongs to $\mathcal{U}$. Thus $\mathcal{U}$ is compact under convergence in every coordinate.

The edge weights satisfy $\bar{c}_{i+1,j+1}(\nu) = q\bar{c}_{ij}(\nu)$. Define a linear map T on real sequences by

$$(Tu)_i = q^{-1}(u_{i+1} - u_1), \qquad i \geq 0.$$

Subtracting u_1 ensures $(Tu)_0 = 0$. Dividing by q preserves the interpolation inequalities: for every $u \in \mathcal{U}$,

$$(Tu)_j - (Tu)_i = q^{-1}(u_{j+1} - u_{i+1}) \leq q^{-1}\bar{c}_{i+1,j+1}(\nu) = \bar{c}_{ij}(\nu).$$

Thus T maps $\mathcal{U}$ into itself. Each coordinate of Tu is a linear combination of two coordinates of u, so T is continuous under convergence in every coordinate.

We now construct $a \in \mathcal{U}$ with $Ta = a$. Choose $u \in \mathcal{U}$ and form the averages

$$a^{(N)} = \frac{1}{N} \sum_{n=0}^{N-1} T^n u, \qquad N \geq 1.$$

Every $T^n u$ lies in $\mathcal{U}$, so the averages lie in $\mathcal{U}$ by convexity. Linearity of T gives

$$Ta^{(N)} - a^{(N)} = \frac{T^N u - u}{N}.$$

Each coordinate of $T^N u$ stays in a fixed bounded interval, so the right-hand side tends to zero in every coordinate. By compactness, a subsequence of the averages converges in every coordinate to some $a \in \mathcal{U}$. Continuity of T then gives $Ta = a$.

The equation $Ta = a$ gives $a_{i+1} = qa_i + a_1$. Since $a_0 = 0$, solving this recurrence gives

$$a_i = (1 - q^i)\varphi_0, \qquad \varphi_0 = \frac{a_1}{1 - q}.$$

Define $\varphi_i = \varphi_0 - a_i = q^i \varphi_0$. Since $a \in \mathcal{U}$, these values satisfy $\varphi_i - \varphi_j = a_j - a_i \leq \bar{c}_{ij}(\nu)$. For $t \geq 1$, taking $(i, j) = (0, t)$ and $(i, j) = (t, 0)$ gives the two bounds in (10).

Conversely, suppose φ_0 satisfies (10) and set $\varphi_i = q^i \varphi_0$ for $i \geq 0$. For $j > i$, the upper bound with $t = j - i$ gives

$$\varphi_i - \varphi_j = q^i(1 - q^{j-i})\varphi_0 \leq q^i \bar{c}_{0,j-i}(\nu) = \bar{c}_{ij}(\nu).$$

For $i > j$, the lower bound with $t = i - j$ gives

$$\varphi_i - \varphi_j = -q^j(1 - q^{i-j})\varphi_0 \leq q^j \bar{c}_{i-j,0}(\nu) = \bar{c}_{ij}(\nu).$$

For $i = j$, the inequality holds because both sides are zero. Thus the interpolation inequalities hold for every pair of indices. Summing them around any finite directed cycle proves cyclic monotonicity. □

B.1 FEASIBILITY AND ATTAINMENT

Lemma B.1 (Feasibility, the classical bound, and attainment)**.** *For $s > 1$, the set $\mathcal{Q}(s)$ of feasible factors is nonempty and has a largest element. Every $q \in \mathcal{Q}(s)$ satisfies $q \le 1 - 1/s$.*

Proof. We first prove the upper bound on q using Nesterov's classical convergence bound. Fix (q, ν, φ_0) satisfying (10). To apply the interpolation construction, we represent the averaged terms $\bar{c}_{ij}(\nu)$ as inner products. Regard the functions r_t and σ_t in (8) as vectors in $L^2(\nu; \mathbb{C})$, the space of square-integrable complex-valued functions. Use the real inner product

$$\langle u, v \rangle = \int_0^\pi \mathrm{Re}\big(\overline{u(\theta)} v(\theta)\big)\, d\nu(\theta).$$

This averages the Euclidean inner product on $\mathbf{R}^2$ over θ, with weights given by ν. It gives $\langle r_i, \sigma_i - \sigma_j \rangle = \bar{c}_{ij}(\nu)$. We include the sample $(\sigma, r, \varphi) = (0, 0, 0)$ to ensure that the interpolating function has minimizer zero. By (6), the conditions $\sigma = r = 0$ give $y = 0$ and $\nabla f(0) = 0$. To check compatibility with the other samples, set $\varphi_i = q^i \varphi_0$. Letting $t \to \infty$ in (10) and multiplying by q^i gives

$$0 \le \varphi_i \le \langle r_i, \sigma_i \rangle, \qquad i \ge 0.$$

These are precisely the interpolation inequalities between each prescribed sample and $(0, 0, 0)$. To define the initial points below, extend $r_t = z^t$ in (8) to $t = -2, -1$. For any integer $T \ge 1$, the vectors r_{-2}, r_{-1} and r_t, σ_t for $0 \le t < T$ span a finite-dimensional real space. The interpolation construction in Section 3.1 gives a function $f \in \mathcal{F}_{1,s^2}$ on this space with minimizer zero. Set $c = (s^2 - 1)/s^2$. Starting from $x_{-1} = c r_{-2}$ and $x_0 = c r_{-1}$, Nesterov's method applied to f gives $x_{t+1} = c r_t$ for $0 \le t < T$. The initial points x_{-1} and x_0 need not coincide. To bound the squared distance to the minimizer, use the Lyapunov function

$$V_t = f(x_t) - f(0) + \frac{1}{2} \| s x_t - (s-1) x_{t-1} \|^2.$$

Strong convexity gives $\|x_t\|^2/2 \le f(x_t) - f(0) \le V_t$. For any initial points, Nesterov's method satisfies

$$V_{t+1} \le (1 - 1/s) V_t, \qquad 0 \le t < T.$$

To prove this, set $v_t = s x_t - (s-1) x_{t-1}$ and $h = 1 - 1/s$. Nesterov's updates give

$$y_t = \frac{s x_t + v_t}{s + 1}, \qquad v_{t+1} = h v_t + \frac{y_t - \nabla f(y_t)}{s}.$$

Apply strong convexity at y_t to bound $f(x_t)$ and $f(0)$ from below. Weight these bounds by h and $1/s$, respectively. Combining them with the smoothness bound for the gradient step and the expression for v_{t+1} gives

$$V_{t+1} - h V_t \le -\frac{h}{2} \left(\|x_t - y_t\|^2 + \frac{1}{s} \|v_t - y_t\|^2 \right) \le 0.$$

Since $x_T = c r_{T-1}$, we have $\|x_T\|^2 = c^2 q^{T-1}$. The lower bound on V_T and the decrease inequality give

$$\frac{c^2}{2} q^{T-1} \le V_T \le (1 - 1/s)^T V_0.$$

The interpolating function may depend on T, so we need a common upper bound on V_0. Smoothness gives $f(x_0) - f(0) \le s^2 \|x_0\|^2/2$. Together with $\|x_0\|^2 = c^2/q$ and $\|x_{-1}\|^2 = c^2/q^2$, this bounds V_0 independently of T and of the interpolating function. Taking Tth roots and letting $T \to \infty$ proves $q \le 1 - 1/s$.

To show that $\mathcal{Q}(s)$ is nonempty, choose $q = q_0(s)$, $\nu = \delta_0$, and $\varphi_0 = 0$. At $\theta = 0$, we have $z = 1 - 1/s$ and $d_s(z) = 0$ by (8). Thus $\bar{c}_{ij}(\nu) = 0$ for all $i, j \ge 0$, so (10) holds. To find the largest feasible factor, we can therefore restrict q to $[q_0(s), 1 - 1/s]$. To prove attainment, we show that the feasible triples (q, ν, φ_0) with q in this interval form a compact set. The function $d_s(\sqrt{q}\, e^{i\theta})$ is continuous and bounded on $[q_0(s), 1 - 1/s] \times [0, \pi]$. The terms $\bar{c}_{01}(\nu)$ and $\bar{c}_{10}(\nu)$ are therefore uniformly bounded over all such q and all probability measures ν. Since $1 - q \ge 1/s$, the bounds (10) with $t = 1$ place every feasible φ_0 in a common compact interval. Probability measures on $[0, \pi]$ form a compact set under weak convergence. Here weak convergence means that the integral of

each continuous function converges to its integral under the limiting measure. For each fixed $t \geq 1$, the integrands in (10) are jointly continuous in q and θ on the compact domain above. Thus the inequalities remain valid when q and φ_0 converge and ν converges weakly. The restricted feasible set is therefore nonempty and compact. The continuous objective q attains a maximum on this set, giving the largest element of $\mathcal{Q}(s)$. □

B.2 Proof of the fixed-domain linear formulation

We prove Proposition 3.2 using the bound $q \leq 1 - 1/s$ from Lemma B.1 and the equivalent interpolation conditions in Lemma 3.1. This proof does not use Theorem 4.1. Put $\alpha_s = 1 - 1/s$ and $J(\mu) = \int |z|^2 \, d\mu(z)$. For $z = \sqrt{q}\, e^{i\theta}$ with $q > 0$, define the individual scaled terms by

$$\widehat{c}^s_{ij}(z) = \frac{q^2}{s^2 - 1} c_{ij}(\theta). \tag{20}$$

Then (11) gives $\mathcal{C}^s_\gamma(z) = \sum_{j=0}^{m-1} \widehat{c}^s_{n_j, n_{j+1}}(z)$. Since $d_s(z) = -(s^2-1)(z-\alpha_s)^2/z^2$, substitution into (20) gives the polynomial expression

$$\widehat{c}^s_{ij}(z) = -\operatorname{Re}\left[z^{i+2}(\bar{z} - \alpha_s)^2(\bar{z}^i - \bar{z}^j)\right], \qquad i, j \geq 0. \tag{21}$$

The scaling in (20) is positive and common to all edges on each nonzero circle. The expression (21) is continuous at $z = 0$ and vanishes there, justifying the extension used in the main text. Conjugating z preserves every cost. For a nonnegative integer t, let $\gamma + t$ denote the cycle obtained by adding t to every index of γ. For any z and $i, j, t \geq 0$, the polynomial costs satisfy

$$\widehat{c}^s_{i+t,j+t}(z) = |z|^{2t} \widehat{c}^s_{ij}(z), \qquad \mathcal{C}^s_{\gamma+t}(z) = |z|^{2t} \mathcal{C}^s_\gamma(z). \tag{22}$$

Lemma B.2 (Radial reweighting)**.** *Let μ be feasible for* (12) *with $m = J(\mu) > 0$. Define $d\eta(z) = |z|^2 \, d\mu(z)/m$, which gives greater relative weight to larger values of $|z|^2$. Then η is feasible for the same problem and*

$$J(\eta) - J(\mu) = \frac{\int |z|^4 \, d\mu - m^2}{m} = \frac{\operatorname{Var}_\mu(|z|^2)}{m}. \tag{23}$$

Proof. Since $|z|^2 \geq 0$ and $\int |z|^2 \, d\mu = m$, η is a probability measure on D_s. For every cycle γ, the shift identity (22) and feasibility of μ give

$$\int \mathcal{C}^s_\gamma \, d\eta = \frac{1}{m} \int |z|^2 \mathcal{C}^s_\gamma \, d\mu = \frac{1}{m} \int \mathcal{C}^s_{\gamma+1} \, d\mu \geq 0.$$

Also $J(\eta) = \int |z|^4 \, d\mu/m$, which proves the identity. □

Proof of Proposition 3.2. Probability measures on the compact disk form a weakly compact set. Every cycle cost is continuous, so the feasible set of (12) is closed. It is nonempty: δ_{α_s} has zero costs and objective $\alpha_s^2 = q_0(s)$. Hence the feasible set is nonempty and compact under weak convergence. Since $z \mapsto |z|^2$ is continuous and bounded on D_s, the objective J is continuous under weak convergence. The extreme value theorem, applied in this topology, therefore gives a maximum $m \geq q_0(s) > 0$. If an optimal measure μ had $\operatorname{Var}_\mu(|z|^2) > 0$, Lemma B.2 would give a feasible measure with a strictly larger objective value. Thus every optimal measure μ satisfies $\operatorname{Var}_\mu(|z|^2) = 0$ and is supported on the circle $|z|^2 = m$, with $0 < m \leq \alpha_s < 1$.

We now prove the fixed-factor equivalence. For any $q \in \mathcal{Q}(s)$, Lemma B.1 gives $q \leq \alpha_s$. Choose an angular measure ν satisfying (10). Replace each angle θ by the point $\sqrt{q}\, e^{i\theta}$ to obtain a probability measure μ on the corresponding circle in D_s. By Lemma 3.1, ν satisfies all cycle inequalities. The positive scaling in (20) then makes μ feasible for (12), with objective value q. Conversely, let μ be feasible for (12) and supported on $|z|^2 = q$ for some $q \in (0, 1)$. Map each point $z = \sqrt{q}\, e^{\pm i\theta}$ to $\theta \in [0, \pi]$ to obtain an angular probability measure ν. Conjugation preserves every cost, and division by the common positive factor $q^2/(s^2 - 1)$ gives all cycle inequalities for ν. Lemma 3.1 then gives a scalar φ_0 satisfying (10), so $q \in \mathcal{Q}(s)$. This argument remains valid at $q = q_0(s)$, even where $d_s(z) = 0$, because it does not divide by $d_s(z)$.

Applying this equivalence to an optimizing disk measure gives $m \in \mathcal{Q}(s)$. Conversely, for every $q \in \mathcal{Q}(s)$, the construction above gives a feasible measure for (12) with objective value q, so $q \leq m$. Thus $m = \max \mathcal{Q}(s)$, proving (13). The lower bound follows from $q_0(s) \in \mathcal{Q}(s)$, and the upper bound from Lemma B.1. □

At $s = 1$ the disk reduces to $\{0\}$ and the polynomial problem has value zero; the division by m is not needed in that case. The proof above does not assert that every feasible measure has one radius, only that every optimizing measure does.

Finite-index relaxations and duality. Let $Q_N(s)$ be the value of (12) when only cycles with vertices in $\{0, \dots, N\}$ are imposed. These conditions can equivalently be written using real potentials $u_0 = 0, u_1, \dots, u_N$ as

$$u_j - u_i \le \int \widehat{c}_{ij}^{\,s}\, d\mu, \qquad 0 \le i, j \le N.$$

Potentials imply the cycle inequalities by telescoping; conversely, shortest-path costs from vertex zero give such potentials when no cycle is negative.

Proposition B.3 (Finite-index bounds and duality)**.** *Each $Q_N(s)$ is attained, and*

$$\alpha_s = Q_0(s) \ge Q_1(s) \ge \cdots \ge q_\star(s), \qquad \lim_{N\to\infty} Q_N(s) = q_\star(s).$$

Moreover,

$$q_\star(s) = \inf \left\{ b : \begin{array}{l} \Gamma_0 \subset \Gamma \textit{ finite}, \quad \lambda_\gamma \ge 0, \\ b - |z|^2 - \displaystyle\sum_{\gamma\in\Gamma_0} \lambda_\gamma \mathcal{C}_\gamma^s(z) \ge 0, \quad z \in D_s \end{array} \right\}. \tag{24}$$

The dual constraint requires a real polynomial in $\operatorname{Re} z$ and $\operatorname{Im} z$ to be nonnegative on D_s. For every $b > q_\star(s)$ such a polynomial can be strictly positive on D_s. No finite dual attainment at $b = q_\star(s)$ is asserted.

Proof. Each finite-index feasible set is closed in the compact probability space and contains δ_{α_s}, so its maximum exists. The sets are nested. A weak limit of optimizers along $N \to \infty$ satisfies every fixed cycle constraint, since that constraint holds eventually. Continuity of J proves that the decreasing limit of $Q_N(s)$ equals $q_\star(s)$.

Integrating any polynomial inequality in (24) against a feasible measure gives $J(\mu) \le b$. Conversely, fix $b > q_\star(s)$. Compactness supplies finitely many cycle constraints whose intersection with $J(\mu) \ge b$ is already empty. Their compact convex moment image is disjoint from $\{t \in \mathbf{R} : t \ge b\} \times \mathbf{R}_+^{|\Gamma_0|}$. Strict separation gives nonnegative coefficients $(\lambda_0, (\eta_\gamma)_{\gamma\in\Gamma_0})$ with

$$\lambda_0 |z|^2 + \sum_{\gamma\in\Gamma_0} \eta_\gamma \mathcal{C}_\gamma^s(z) < \lambda_0 b, \quad z \in D_s.$$

The coefficient λ_0 is positive: otherwise evaluating at $z = \alpha_s$, where all cycle costs vanish, contradicts strictness. Division by λ_0 proves the claim. □

For fixed cycles and degree bounds, one can search for a sufficient dual certificate of the form

$$b - |z|^2 - \sum_{\gamma\in\Gamma_0} \lambda_\gamma \mathcal{C}_\gamma^s = S_0 + (\alpha_s - |z|^2) S_1,$$

where S_0, S_1 are sums of squares of real polynomials in $\operatorname{Re} z$ and $\operatorname{Im} z$. Coefficient matching and positive semidefinite Gram matrices give a semidefinite program with fixed data. A checked polynomial certificate gives an upper bound, but no finite exact degree or runtime improvement is claimed. Likewise, replacing the disk by a finite support grid changes the feasible set: a grid computation with truncated cycles is not automatically an upper or lower bound for the original problem.

C Full proof of the exact-rate theorem

This section proves Theorem 4.1. We first show that, when the interpolation conditions are infeasible, a nonnegative combination of finitely many cycle functions is strictly negative at every angle. This separation argument is used in both the lower and upper bounds.

For the lower bound, we show that, if $q_\star(s) > q_0(s)$, an optimal feasible measure can be supported on finitely many angles. We use this measure to construct one finite-dimensional objective and initial

point satisfying $x_{-1} = x_0$ and attaining the asymptotic factor $q_\star(s)$. If $q_\star(s) = q_0(s)$, a quadratic objective attains the factor.

For the upper bound, fix any $q \in (q_\star(s), 1)$. The negative combination of cycle functions gives a convergence bound with a constant depending only on s and q. The bound therefore holds uniformly over objectives, dimensions, initial points, and iteration counts. Letting $q \downarrow q_\star(s)$ identifies both worst-case asymptotic factors with $q_\star(s)$.

The measure arguments use weak compactness and finite-dimensional convex separation. Rockafellar's cyclic-monotonicity theorem supplies the interpolating objective, and Montel's theorem establishes finite support. The rate estimates use Parseval's identity, Nesterov's classical estimate-sequence bound, and the discrete strict-IQC theorem stated in Theorem C.5. We state and check the required hypotheses where these results are used.

C.1 Finite alternatives, attainment, and finite support

Lemma C.1 (Finite certificate of infeasibility)**.** *Fix $s > 1$ and $q \in (q_0(s), 1)$. For a finite cycle $\gamma = (i_0, i_1, \dots, i_m = i_0)$, define the sum of the interpolation terms* (9) *by*

$$c_\gamma(\theta) = \sum_{j=0}^{m-1} c_{i_j i_{j+1}}(\theta).$$

Exactly one of the following holds:

1. *There is a probability measure ν on $[0, \pi]$ such that*

$$\int_0^\pi c_\gamma(\theta)\, d\nu(\theta) \geq 0, \qquad \text{for every finite cycle } \gamma.$$

2. *There are finitely many cycles $\gamma_1, \dots, \gamma_n$ and weights $\lambda_j \geq 0$ such that*

$$\phi(\theta) = \sum_{j=1}^{n} \lambda_j c_{\gamma_j}(\theta) < 0, \qquad 0 \leq \theta \leq \pi.$$

Proof. Let $\mathcal{K} = \mathcal{P}([0, \pi])$ and write

$$\bar{c}_\gamma(\nu) = \int_0^\pi c_\gamma(\theta)\, d\nu(\theta)$$

for the average cycle cost. The set $\mathcal{K}$ is convex and compact for weak convergence, meaning convergence of integrals against continuous functions. Since each cycle function c_γ is continuous, its average $\bar{c}_\gamma(\nu)$ is continuous in ν for this convergence.

Suppose alternative 1 fails. Then every probability measure in $\mathcal{K}$ violates at least one cycle inequality. For each cycle γ, the measures satisfying $\bar{c}_\gamma(\nu) < 0$ form an open subset of $\mathcal{K}$. These open sets cover $\mathcal{K}$. By compactness, finitely many of them still cover $\mathcal{K}$. Thus there are cycles $\gamma_1, \dots, \gamma_n$ whose inequalities have no common feasible probability measure.

Collect the average cycle costs in the set

$$\mathcal{M} = \left\{ \left(\bar{c}_{\gamma_1}(\nu), \dots, \bar{c}_{\gamma_n}(\nu)\right) : \nu \in \mathcal{K} \right\} \subset \mathbf{R}^n.$$

This set is compact because the averages depend continuously on ν. It is convex because averaging is linear in the measure. It contains no vector with all coordinates nonnegative, since such a vector would correspond to a measure satisfying all the selected inequalities. Thus $\mathcal{M}$ and the closed convex set $\mathbf{R}^n_+$ are disjoint. Strict separation gives a vector $\lambda \in \mathbf{R}^n$ such that

$$\sup_{u \in \mathcal{M}} \lambda^T u < \inf_{v \in \mathbf{R}^n_+} \lambda^T v.$$

If any λ_j were negative, increasing the corresponding coordinate v_j without bound would make the infimum equal to $-\infty$. Hence every $\lambda_j \geq 0$, and the infimum equals zero.

For each angle θ, take the probability measure δ_θ concentrated at that angle. Its average cycle costs are $c_{\gamma_j}(\theta)$, so

$$\phi(\theta) = \sum_{j=1}^{n} \lambda_j c_{\gamma_j}(\theta) \le \sup_{u \in \mathcal{M}} \lambda^T u < 0, \qquad 0 \le \theta \le \pi.$$

The upper bound is independent of θ, so the negativity is uniform. This proves alternative 2 when alternative 1 fails.

Finally, if both alternatives held, a feasible measure ν would satisfy

$$\int_0^\pi \phi(\theta)\, d\nu(\theta) = \sum_{j=1}^{n} \lambda_j \bar{c}_{\gamma_j}(\nu) \ge 0.$$

But the integral must be strictly negative because ϕ is uniformly negative. This contradiction proves that exactly one alternative holds. □

Lemma C.2 (Feasible factors stay below one)**.** *Fix $s > 1$. For every $q \in (q_0(s), 1)$ sufficiently close to one, there is an integer $m \ge 1$ such that the cycle $\gamma = (0, 1, \dots, m, 0)$ satisfies*

$$c_\gamma(\theta) < 0, \qquad 0 \le \theta \le \pi.$$

Consequently, $\sup \mathcal{Q}(s) < 1$.

Proof. Put $r = \sqrt{q}$, $z = re^{i\theta}$, and $x = 2r\cos\theta \in [-2r, 2r]$. Equations (8) and (9) give

$$c_{01}(\theta) = -\frac{s^2 - 1}{2s^2 q^2} \psi_s(q, x),$$

where $\psi_s(q, x)$ is the convex quadratic in x

$$\psi_s(q, x) = (s-1)^2 x^2 - q(qs^2 + 3s^2 - 4s + 1)x + 6q^2 s^2 - 4q^2 s - 2qs^2 + 4qs - 2q. \tag{25}$$

The unconstrained minimizer x_{v} and the value of ψ_s at the right endpoint $x = 2r$ are

$$x_{\mathrm{v}} = \frac{q(qs^2 + 3s^2 - 4s + 1)}{2(s-1)^2}, \qquad \psi_s(q, 2r) = -2q(r-1)(sr - s + 1)^2 > 0. \tag{26}$$

For fixed $s > 1$, we have

$$\lim_{q \uparrow 1} (x_{\mathrm{v}} - 2\sqrt{q}) = \frac{4s - 3}{2(s-1)^2} > 0.$$

Hence $x_{\mathrm{v}} > 2r$ for every $q < 1$ sufficiently close to one. For these q, the minimum of $\psi_s(q, x)$ over $[-2r, 2r]$ is its value at $x = 2r$, which is positive by (26). It follows that $c_{01}(\theta) < 0$ for every $\theta \in [0, \pi]$.

Fix one such $q \in (q_0(s), 1)$. For the cycle $\gamma_m = (0, 1, \dots, m, 0)$, the identity $c_{t,t+1}(\theta) = q^t c_{01}(\theta)$ gives

$$c_{\gamma_m}(\theta) = \frac{1 - q^m}{1 - q} c_{01}(\theta) + c_{m0}(\theta).$$

Equation (9) also gives

$$|c_{m0}(\theta)| \le |d_s(z)|(q^m + q^{m/2}).$$

Since d_s is bounded on the fixed circle $|z| = \sqrt{q}$, the closing term tends to zero uniformly in θ. Thus the cycle costs converge uniformly on $[0, \pi]$ to $c_{01}(\theta)/(1 - q)$. This limit is continuous and strictly negative on the compact interval, so for all sufficiently large m the cycle cost is negative at every angle. Lemma C.1 therefore rules out feasibility for every $q < 1$ sufficiently close to one. Consequently, $\sup \mathcal{Q}(s) < 1$. □

Lemma C.3 (Finite support at the optimal factor)**.** *Fix $s > 1$ and suppose $q_\star(s) > q_0(s)$. There are a probability measure $\nu_\star$ supported on finitely many angles in $[0, \pi]$ and a real scalar $\varphi_0^\star$ that satisfy* (10) *at $q = q_\star(s)$.*

Proof. Proposition 3.2 gives $q_\star(s) = \max \mathcal{Q}(s)$. Choose a pair $(\nu_\star, \varphi_0^\star)$ satisfying (10) at $q = q_\star(s)$.

We will construct a continuous function that is nonpositive on $[0, \pi]$ and has finitely many zeros. Feasibility will force its integral against $\nu_\star$ to be nonnegative. The measure must therefore be supported on those zeros.

For $q \in (q_\star(s), 1)$, the interpolation conditions are infeasible. Lemma C.1 then supplies nonnegative weighted sums of cycle functions that are strictly negative at every angle. We will scale these sums to equal -1 at $\theta = 0$ and obtain the desired function as a limit along a sequence $q \downarrow q_\star(s)$.

To compare nearby factors, choose $q_0(s) < q_- < q_\star(s) < q_+ < 1$ and set $\mathcal{I} = [q_-, q_+]$.

Since q varies in this proof, write $c_\gamma(q, \theta)$ for the cycle cost at factor q.

Every cycle can be split at repeated indices until each resulting cycle visits each index only once before returning to its starting point. These are called simple cycles, and their costs sum to the original cycle cost. Cycles containing only one index have zero cost and can be discarded.

For a remaining cycle γ, let k be its smallest index. Subtracting k from every index gives a cycle γ_0 whose smallest index is zero, with

$$c_\gamma(q, \theta) = q^k c_{\gamma_0}(q, \theta).$$

Since $q^k > 0$, this factor can be absorbed into the cycle's nonnegative weight. It therefore suffices to consider simple cycles with at least two indices and minimum index zero.

Write such a cycle as $\gamma = (n_0, \dots, n_{m-1}, n_m = n_0)$ and put $r = \sqrt{q}$. We first show that its cost is strictly negative at $\theta = 0$. Expanding the squares and using $n_m = n_0$ gives

$$c_\gamma(q, 0) = \frac{d_s(r)}{2} \sum_{j=0}^{m-1} \left(r^{n_{j+1}} - r^{n_j}\right)^2.$$

Every other index is at least one, so the two steps entering and leaving index zero each contribute at least $(1 - r)^2$ to the sum. Since $d_s(r) < 0$ for $q \in \mathcal{I}$, it follows that

$$d_\gamma(q) := -c_\gamma(q, 0) \geq |d_s(r)|(1 - r)^2 \geq d_0 > 0,$$

where

$$d_0 = \min_{q \in \mathcal{I}} |d_s(\sqrt{q})|(1 - \sqrt{q})^2.$$

The constant d_0 is positive because $\mathcal{I}$ is a compact interval contained in $(q_0(s), 1)$. Dividing each cycle cost by $d_\gamma(q)$ makes its value at zero equal to -1. The lower bound on the divisor holds for every such cycle and every $q \in \mathcal{I}$.

We next extend the cycle costs to complex values of θ. This lets us use compactness for analytic functions and show that the limiting function has only finitely many zeros. Here, analytic means complex differentiable on an open set. Choose $\delta > 0$ small enough that

$$1 - 1/s < \sqrt{q_-} e^{-\delta}, \qquad r_+ := \sqrt{q_+} e^{\delta} < 1.$$

For $q \in \mathcal{I}$ and $|\operatorname{Im} \theta| < \delta$, set

$$z_+ = r e^{i\theta}, \qquad z_- = r e^{-i\theta}.$$

The constant r_+ bounds both $|z_+|$ and $|z_-|$, independently of q and θ in these domains. For real θ, the numbers z_+ and z_- are complex conjugates. Thus the function

$$\widetilde{c}_\gamma(q, \theta) = \frac{1}{2} \sum_{(i,j) \in \gamma} \left[d_s(z_-) z_+^i (z_-^i - z_-^j) + d_s(z_+) z_-^i (z_+^i - z_+^j) \right]$$

equals $c_\gamma(q, \theta)$ for real θ. The sum runs over the consecutive index pairs in the cycle. Since d_s is a rational function with its only pole at zero, this formula is analytic throughout the strip $|\operatorname{Im} \theta| < \delta$.

We need bounds that hold for every cycle and every $q \in \mathcal{I}$. Each product satisfies $|z_+^i z_-^j| \leq r_+^{i+j}$. The inequality $2r_+^{i+j} \leq r_+^{2i} + r_+^{2j}$ gives

$$\sum_{(i,j) \in \gamma} r_+^{i+j} \leq \sum_{n \in \{n_0, \dots, n_{m-1}\}} r_+^{2n} \leq \sum_{n \geq 0} r_+^{2n} < \infty,$$

because each index occurs once as the first index of a pair and once as the second. The factors $d_s(z_\pm)$ and their derivatives with respect to q are uniformly bounded, since their arguments stay bounded and separated from zero. Differentiating a product with respect to q gives

$$\partial_q(z_+^i z_-^j) = \frac{i+j}{2q} z_+^i z_-^j.$$

The resulting sums are also bounded independently of the cycle, since $q \geq q_- > 0$ and

$$\sum_{(i,j)\in\gamma} (i+j) r_+^{i+j} \leq 2 \sum_{n\geq 0} n r_+^n < \infty.$$

These bounds control both $\widetilde{c}_\gamma$ and its first derivative with respect to q throughout the strip. At $\theta = 0$, they also bound $\partial_q d_\gamma(q)$. The lower bound $d_\gamma(q) \geq d_0$ and the quotient rule therefore give uniform bounds on

$$\psi_{\gamma,q}(\theta) := \frac{\widetilde{c}_\gamma(q,\theta)}{d_\gamma(q)}, \qquad \psi_{\gamma,q}(0) = -1,$$

and its first derivative with respect to q. Each function $\psi_{\gamma,q}$ is analytic on the same strip.

Choose $q_n \downarrow q_\star(s)$ with $q_\star(s) < q_n \leq q_+$. For each n, Lemma C.1 and the cycle reduction above give finitely many nonnegative weights $\lambda_{\gamma,n}$ such that

$$\sum_\gamma \lambda_{\gamma,n} c_\gamma(q_n,\theta) < 0, \qquad 0 \leq \theta \leq \pi.$$

Define

$$a_{\gamma,n} = \frac{\lambda_{\gamma,n} d_\gamma(q_n)}{\sum_\eta \lambda_{\eta,n} d_\eta(q_n)}.$$

The denominator is positive, the weights $a_{\gamma,n}$ are nonnegative, and their sum is one. The scaled combination

$$\phi_n(\theta) = \sum_\gamma a_{\gamma,n} \psi_{\gamma,q_n}(\theta)$$

therefore satisfies $\phi_n(\theta) < 0$ on $[0,\pi]$ and $\phi_n(0) = -1$. To use feasibility at $q_\star(s)$, evaluate the same combination at that factor:

$$\psi_n(\theta) = \sum_\gamma a_{\gamma,n} \psi_{\gamma,q_\star(s)}(\theta).$$

The uniform derivative bound gives a constant b, independent of n, such that

$$\sup_{0\leq\theta\leq\pi} |\psi_n(\theta) - \phi_n(\theta)| \leq b|q_n - q_\star(s)| \longrightarrow 0.$$

The functions ψ_n are analytic and uniformly bounded on the fixed strip. Montel's theorem gives a subsequence that converges uniformly on every compact subset of the strip to an analytic function $\phi_\star$. In particular, convergence is uniform on $[0,\pi]$. The preceding bound and $\phi_n < 0$ on this interval imply

$$\phi_\star(\theta) \leq 0, \quad 0 \leq \theta \leq \pi, \qquad \phi_\star(0) = -1.$$

Feasibility of $\nu_\star$ at $q_\star(s)$ means that every cycle has a nonnegative average cost. Dividing by the positive number $d_\gamma(q_\star(s))$ gives

$$\int_0^\pi \psi_{\gamma,q_\star(s)}(\theta)\, d\nu_\star(\theta) \geq 0.$$

Taking nonnegative weighted sums and then the uniform limit yields

$$\int_0^\pi \psi_n(\theta)\, d\nu_\star(\theta) \geq 0, \qquad \int_0^\pi \phi_\star(\theta)\, d\nu_\star(\theta) \geq 0.$$

Since $\phi_\star$ is continuous and nonpositive, it must vanish on the support of $\nu_\star$. Otherwise, a neighborhood of a support point would have positive measure and a strictly negative integral.

The analytic function $\phi_\star$ is not identically zero because $\phi_\star(0) = -1$. Its zeros are therefore isolated. The interval $[0,\pi]$ is compact and lies inside the strip, so it contains only finitely many zeros. Thus $\nu_\star$ has finite support. The corresponding construction uses one real coordinate for each support angle equal to 0 or π, and two for each support angle in $(0,\pi)$. □

C.2 FINITE-DIMENSIONAL REALIZATION WITH THE STANDARD INITIALIZATION

An optimal angular measure with finite support determines a finite-dimensional self-similar trajectory. We must still enforce the standard initialization $x_{-1} = x_0$. We do this by scaling the prescribed trajectory and adding one initial sample and one additional coordinate generated by Nesterov's method on a scalar quadratic. The additional coordinate supplies slack in the interpolation inequalities involving the initial sample. It decays faster than the prescribed trajectory, so the asymptotic factor is unchanged.

Lemma C.4 (Enforcing the standard initialization)**.** *Fix $s > 1$ with $q_\star(s) > q_0(s)$, and set $\kappa = s^2$. Consider any self-similar trajectory associated with a finitely supported optimal angular pair. After scaling this trajectory by a positive constant and adding one coordinate, it can be realized from iteration one onward by Nesterov's method on a single finite-dimensional objective in $\mathcal{F}_{1,\kappa}$. The method starts with $x_{-1} = x_0$, and its squared-distance asymptotic factor is $q_\star(s)$.*

Proof. Write $\beta = \beta_\kappa \in (0, 1)$ and set

$$a = \frac{\beta}{4}, \qquad \lambda = \kappa(1 - a).$$

Since $4\beta/(1+\beta)^2 = 1 - 1/\kappa$, we have $0 < a < 1 - 1/\kappa$, and hence $1 < \lambda < \kappa$. Apply Nesterov's method to the scalar quadratic $u \mapsto \lambda u^2/2$, starting from $u_{-1} = u_0 = 1$. Denote its query points by v_t, so that

$$v_t = (1 + \beta)u_t - \beta u_{t-1}, \qquad u_{t+1} = av_t.$$

We first show that $|u_t| \le \beta^t$ for all $t \ge 0$. The bound holds at $t = 0$ and $t = 1$, since $u_0 = 1$ and $u_1 = \beta/4$. If it holds at $t - 1$ and t, then

$$|u_{t+1}| \le \frac{\beta}{4}\big((1 + \beta)\beta^t + \beta\beta^{t-1}\big) = \frac{2 + \beta}{4}\beta^{t+1} \le \beta^{t+1}.$$

Since $v_t = u_{t+1}/a$, the query points satisfy $|v_t| \le 4\beta^t$. Thus both sequences decay geometrically, with squared asymptotic factor at most

$$\beta^2 = \left(\frac{s-1}{s+1}\right)^2 < \left(\frac{s-1}{s}\right)^2 = q_0(s) < q_\star(s).$$

We next show that every later query stays at least $1/4$ away from the initial query $v_0 = 1$. We have $v_1 = -\beta(3 - \beta)/4$, so $|v_1| < 1/2$. The query points satisfy the same recurrence as the iterates. For $t \ge 1$, if $|v_{t-1}| \le 1$ and $|v_t| \le 1$, then

$$|v_{t+1}| = a|(1 + \beta)v_t - \beta v_{t-1}| \le a(1 + 2\beta) < \frac{3}{4}.$$

Induction therefore gives $|v_t| < 3/4$ for every $t \ge 1$, and hence

$$|v_0 - v_t| \ge 1 - |v_t| > \frac{1}{4}.$$

The same distance bound holds for the optimum $v_\star = 0$.

Write the interpolation condition (5) for each ordered pair as $e_{ij} \le 0$, where

$$e_{ij} = f_j - f_i + \langle g_j, y_i - y_j\rangle + \frac{\|g_i - g_j\|^2}{2\kappa} + \frac{\kappa}{2(\kappa - 1)}\left\|y_i - y_j - \frac{g_i - g_j}{\kappa}\right\|^2.$$

For the scalar quadratic, substituting $y_i = v_i$, $g_i = \lambda v_i$, and $f_i = \lambda v_i^2/2$ gives

$$e_{ij} = -c(v_i - v_j)^2, \qquad c = \frac{(\lambda - 1)(\kappa - \lambda)}{2(\kappa - 1)} > 0.$$

The preceding distance bound therefore gives

$$e_{0t} = e_{t0} \le -\frac{c}{16}, \qquad t \in \{\star, 1, 2, \ldots\}.$$

Thus both interpolation inequalities between the initial sample and any later sample or the optimum have slack at least $c/16$.

Use the superscript tail for the prescribed trajectory, beginning at $t = 1$. The extrapolation equation requires

$$y_1^{\mathrm{tail}} = (1+\beta)x_1^{\mathrm{tail}} - \beta x_0^{\mathrm{tail}},$$

so set

$$x_0^{\mathrm{tail}} = \frac{(1+\beta)x_1^{\mathrm{tail}} - y_1^{\mathrm{tail}}}{\beta}.$$

Set $x_{-1}^{\mathrm{tail}} = y_0^{\mathrm{tail}} = x_0^{\mathrm{tail}}$ and prescribe

$$g_0^{\mathrm{tail}} = \kappa(x_0^{\mathrm{tail}} - x_1^{\mathrm{tail}}).$$

Then the first gradient step reaches x_1^{tail}, and the next extrapolation gives y_1^{tail}. The NAG updates do not depend on f_0^{tail}, so assign it any finite value. Scaling the prescribed trajectory and adding the scalar coordinate will enforce the interpolation inequalities involving this new sample.

Include the optimal sample

$$(y_\star^{\mathrm{tail}}, g_\star^{\mathrm{tail}}, f_\star^{\mathrm{tail}}) = (0,0,0),$$

whose compatibility with the prescribed trajectory was established in the proof of Lemma B.1. Self-similarity gives $\|y_t^{\mathrm{tail}}\| + \|g_t^{\mathrm{tail}}\| = O(q_\star(s)^{t/2})$. The interpolation inequalities with the optimum imply

$$0 \le f_t^{\mathrm{tail}} \le \langle g_t^{\mathrm{tail}}, y_t^{\mathrm{tail}} \rangle \le \|g_t^{\mathrm{tail}}\| \|y_t^{\mathrm{tail}}\| = O(q_\star(s)^t).$$

Thus the tail samples converge to the optimal sample. With the initial sample fixed, the residual formula gives

$$e_{0t}^{\mathrm{tail}} \to e_{0\star}^{\mathrm{tail}}, \qquad e_{t0}^{\mathrm{tail}} \to e_{\star 0}^{\mathrm{tail}}.$$

Both sequences converge and are therefore bounded, so

$$b = \sup_{t \in \{\star,1,2,\ldots\}} \max\{0, e_{0t}^{\mathrm{tail}}, e_{t0}^{\mathrm{tail}}\} < \infty.$$

For $\varepsilon > 0$, define the combined iterates and samples by

$$x_t = (u_t, \varepsilon x_t^{\mathrm{tail}}), \qquad y_t = (v_t, \varepsilon y_t^{\mathrm{tail}}),$$

$$g_t = (\lambda v_t, \varepsilon g_t^{\mathrm{tail}}), \qquad f_t = \frac{\lambda}{2} v_t^2 + \varepsilon^2 f_t^{\mathrm{tail}}.$$

Include the optimal sample $(y_\star, g_\star, f_\star) = (0,0,0)$. Inner products and squared norms add across coordinates, so

$$e_{ij} = -c(v_i - v_j)^2 + \varepsilon^2 e_{ij}^{\mathrm{tail}}.$$

Pairs that do not involve the initial sample have $e_{ij}^{\mathrm{tail}} \le 0$, and therefore $e_{ij} \le 0$. Choose

$$\varepsilon^2 = \frac{c}{16(b+1)}.$$

For every $t \in \{\star, 1, 2, \ldots\}$, this gives

$$\max\{e_{0t}, e_{t0}\} \le -\frac{c}{16} + \varepsilon^2 b = -\frac{c}{16(b+1)} < 0.$$

Together with $e_{00} = 0$, this proves all the interpolation inequalities.

We now construct one objective interpolating all the samples. Set $\sigma_i = g_i - y_i$ and define

$$p_i(\sigma) = \langle y_i, \sigma \rangle - f_i + \frac{1}{2}\|y_i\|^2 + \frac{\|\sigma - \sigma_i\|^2}{2(\kappa - 1)}.$$

Let $p(\sigma) = \sup_i p_i(\sigma)$, with the supremum taken over all samples, including the optimum. The samples are bounded, so p is finite everywhere. Subtracting $\|\sigma\|^2/(2(\kappa-1))$ leaves a supremum of affine functions. Thus p is $1/(\kappa-1)$-strongly convex.

The interpolation inequalities give

$$p_i(\sigma_j) - p_j(\sigma_j) = e_{ij} \le 0.$$

Consequently, each quadratic attains the supremum at its own sample:

$$p(\sigma_i) = p_i(\sigma_i), \qquad y_i = \nabla p_i(\sigma_i) \in \partial p(\sigma_i).$$

Define

$$f(y) = p^*(y) + \frac{1}{2}\|y\|^2,$$

where p^* is the convex conjugate of p. Strong convexity of p makes p^* differentiable with $(\kappa-1)$-Lipschitz gradient, so $f \in \mathcal{F}_{1,\kappa}$. Conjugacy gives $\nabla p^*(y_i) = \sigma_i$, hence $\nabla f(y_i) = g_i$. Fenchel equality also gives

$$f(y_i) = \langle y_i, \sigma_i \rangle - p_i(\sigma_i) + \frac{1}{2}\|y_i\|^2 = f_i.$$

The constructed sequences satisfy the NAG updates (1) because $\nabla f(y_t) = g_t$. Thus (x_t) is the trajectory of Nesterov's method on f, with $x_{-1} = x_0$. Interpolation at the optimal sample gives $\nabla f(0) = 0$, so $x_\star = 0$. The initial point is nonoptimal because its first coordinate is $u_0 = 1$.

For $t \ge 1$, self-similarity of the prescribed trajectory gives

$$\|x_t\|^2 = u_t^2 + \varepsilon^2 \|x_t^{\mathrm{tail}}\|^2 = u_t^2 + \varepsilon^2 q_\star(s)^{t-1} \|x_1^{\mathrm{tail}}\|^2.$$

The vector x_1^{tail} is nonzero, and $|u_t| \le \beta^t$ with $\beta^2 < q_\star(s)$. Taking tth roots therefore yields

$$\lim_{t\to\infty} \left(\frac{\|x_t - x_\star\|^2}{\|x_0 - x_\star\|^2} \right)^{1/t} = q_\star(s).$$

□

C.3 Uniform upper bound from a strict separator

Fix $s > 1$ and $q \in (q_\star(s), 1)$. We will bound $\|x_t - x_\star\|^2$ by a constant multiple of $q^t \|x_0 - x_\star\|^2$ for every objective in $\mathcal{F}_{1,s^2}$ and every standard initialization. The constant will depend only on s and q, so the bound is uniform over objectives, dimensions, initial points, and iteration counts.

Lemma C.1 supplies a nonnegative combination of cycle costs that is strictly negative at every angle. To use it, we first translate the minimizer to zero and rescale each sequence at time t by $q^{-t/2}$. The cycle inequalities then give a quadratic inequality for the rescaled sequences. Writing this inequality as an integral over frequency gives an *integral quadratic constraint* (IQC). Together with the strict separator, a standard IQC stability theorem shows that the rescaled query points and transformed gradients have finite sums of squared norms. We then derive a common bound on these sums and recover the desired bound on the iterates.

This proof uses the IQC framework (Lessard et al., 2016; Safavi et al., 2018; Gannot, 2022). We recall the discrete stability theorem of Zhang et al. (2019, Theorem 3), based on Megretski & Rantzer (1997, Theorem 1).

Let ℓ^d be the set of all real d-vector sequences indexed by $\mathbb{Z}_{\ge 0}$. Its square-summable subspace ℓ_2^d consists of the sequences $u = (u_t)$ satisfying

$$\|u\|_{\ell_2}^2 = \sum_{t=0}^{\infty} \|u_t\|^2 < \infty.$$

A sequence operator maps an input sequence u to an output sequence v. It is *causal* if v_t depends only on $u_0, \dots, u_t$. It is *bounded* if there is a finite $b \ge 0$ such that

$$\|v\|_{\ell_2} \le b\|u\|_{\ell_2}, \qquad u \in \ell_2^d.$$

The transformed gradient in (6) is (s^2-1)-Lipschitz and satisfies $\nabla h(0) = 0$ after translating the minimizer to zero. Thus $v_t = \nabla h(u_t)$ defines a bounded causal operator with $b = s^2 - 1$.

A linear recurrence with fixed coefficients computes each output from finitely many earlier outputs and current or earlier inputs. With zero values before time zero, its output sequence has the form

$$v_t = (Gu)_t = \sum_{j=0}^{t} G_j u_{t-j},$$

where the real matrix G_j describes the effect of an input j iterations earlier. The *transfer matrix* collects these coefficients:

$$G(z) = \sum_{j=0}^{\infty} G_j z^{-j},$$

with the series converging for sufficiently large $|z|$. Each factor z^{-1} represents a delay of one iteration.

For such a recurrence, the entries of $G(z)$ are rational functions with real coefficients. The transfer matrix is *proper*, meaning that it has a finite limit as $|z| \to \infty$. Its poles are the singularities of these rational entries. It is *stable* if every pole has modulus less than one. A stable proper transfer matrix defines a bounded causal sequence operator.

For $u \in \ell_2^d$, write

$$\widehat{u}(\theta) = \sum_{t=0}^{\infty} u_t e^{-it\theta}, \qquad -\pi \le \theta \le \pi,$$

for its Fourier transform. The partial sums converge in L^2, meaning that their integrated squared error tends to zero. For a stable transfer matrix, $v = Gu$ gives $\widehat{v}(\theta) = G(e^{i\theta})\widehat{u}(\theta)$ for almost every θ.

For any real symmetric $d \times d$ matrix M, Parseval's identity gives

$$\sum_{t=0}^{\infty} u_t^T M u_t = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{u}(\theta)^* M \widehat{u}(\theta)\, d\theta.$$

Thus a nonnegativity constraint on the sum is equivalent to a nonnegativity constraint on the integral. In our construction, the quadratic expressions also involve delayed sequence values. The corresponding factors $e^{-ij\theta}$ for delays of j iterations produce a frequency-dependent matrix inside the integral.

Let Π be a bounded measurable Hermitian $2d \times 2d$ matrix function on the unit circle. Here $*$ denotes conjugate transpose, and Hermitian means $\Pi(z)^* = \Pi(z)$ for $|z| = 1$. Two sequences $v, w \in \ell_2^d$ satisfy the *integral quadratic constraint* defined by Π if

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \begin{bmatrix} \widehat{v}(\theta) \\ \widehat{w}(\theta) \end{bmatrix}^* \Pi(e^{i\theta}) \begin{bmatrix} \widehat{v}(\theta) \\ \widehat{w}(\theta) \end{bmatrix} d\theta \ge 0.$$

The matrix function Π is called the *multiplier*. A bounded causal operator Δ satisfies this IQC if the inequality holds for every $v \in \ell_2^d$ with $w = \Delta(v)$. The multiplier may be indefinite, since the inequality is required only for such input-output pairs.

Let G and Δ be sequence operators on ℓ^d, with G linear. For $\tau \in [0, 1]$ and prescribed input sequences $e_v, e_w \in \ell^d$, consider the equations

$$v = e_v + Gw, \qquad w = e_w + \tau\Delta(v), \tag{27}$$

for the unknown sequences $v, w \in \ell^d$. They form a *feedback interconnection*: the output of each operator contributes to the input of the other.

The parameter τ scales the contribution of Δ. At $\tau = 0$, the equations give $w = e_w$ and $v = e_v + Ge_w$ directly. At $\tau = 1$, they give the interconnection whose stability we want to prove.

For fixed τ, the interconnection is *well posed* if every $(e_w, e_v) \in \ell^d \times \ell^d$ determines a unique solution (v, w) and the map from (e_w, e_v) to (v, w) is causal. It is ℓ_2 *stable* if it is well posed and $(v, w) \in \ell_2^d \times \ell_2^d$ for every $(e_w, e_v) \in \ell_2^d \times \ell_2^d$. A matrix inequality required at every point on the unit circle is called a *frequency-domain inequality* (FDI).

Theorem C.5 (Stability from an integral quadratic constraint)**.** *Let G be a stable proper real-rational $d \times d$ transfer matrix, and let $\Delta : \ell^d \to \ell^d$ be a bounded causal operator. Suppose the following conditions hold.*

1. *The feedback interconnection* (27) *is well posed for every* $\tau \in [0,1]$.
2. *There is one bounded measurable Hermitian multiplier* Π *such that* $\tau\Delta$ *satisfies its IQC for every* $\tau \in [0,1]$.
3. *There is an* $\epsilon > 0$ *such that*

$$\begin{bmatrix} G(\zeta) \\ I \end{bmatrix}^* \Pi(\zeta) \begin{bmatrix} G(\zeta) \\ I \end{bmatrix} \preceq -\epsilon I, \qquad |\zeta| = 1.$$

Then the interconnection at $\tau = 1$ *is* ℓ_2 *stable.*

We next apply the theorem to the rescaled Nesterov recurrence. The proof first establishes square summability and then derives a bound with a constant depending only on s and q.

Proposition C.6 (Uniform rate bound from a strict cycle separator)**.** *Fix* $s > 1$ *and* $q > q_0(s)$. *Suppose a nonnegative linear combination of finitely many cycle costs is strictly negative at every angle* $\theta \in [0,\pi]$. *Then there is a finite* $b(s,q)$ *such that every NAG trajectory* (1) *with* $\kappa = s^2$, *on any finite-dimensional objective* $f \in \mathcal{F}_{1,s^2}$ *and with any initialization* $x_{-1} = x_0$, *satisfies*

$$\|x_t - x_\star\|^2 \le b(s,q) q^t \|x_0 - x_\star\|^2, \qquad t \ge 0.$$

Proof. The upper-bound calculation expresses r in terms of σ. Recall from (8) that $d_s(z)$ is the complex coefficient relating the transformed gradients and iterates by $\sigma_t = d_s(z) r_t$. For the contracting and rotating trajectories, we use $\sigma_t = z^t$ instead of the choice $r_t = z^t$ in (8). We first relate these two choices while preserving the interpolation inequalities. For $z = \sqrt{q}\, e^{i\theta}$ with $q > q_0(s)$, $d_s(z)$ is nonzero for every angle, so define

$$p_s(z) = \frac{1}{d_s(z)} = -\frac{s^2 z^2}{(s^2-1)(sz-(s-1))^2}. \tag{28}$$

Multiplying both vectors in (8) by $p_s(z)$ gives $\sigma_t = z^t$ and $r_t = p_s(z) z^t$. For each fixed angle, this applies the same rotation and scaling to both vectors. Since $p_s(z) = |p_s(z)|^2 \overline{d_s(z)}$, multiplying (9) by $|p_s(z)|^2$ gives

$$\mathrm{Re}\big[p_s(z) z^i (\bar{z}^i - \bar{z}^j)\big] = |p_s(z)|^2 c_{ij}(\theta). \tag{29}$$

Let η be a probability measure on $[0,\pi]$ specifying the weights of the angles when $\sigma_t = z^t$. In this normalization, write the prescribed values of the convex interpolation function in (7) as $q^t v$, where v is its value at time zero. To return to $r_t = z^t$, set

$$d\nu = \frac{|p_s(z)|^2\, d\eta}{\int |p_s(z)|^2\, d\eta}, \qquad \varphi_0 = \frac{v}{\int |p_s(z)|^2\, d\eta}. \tag{30}$$

The factor $|p_s(z)|^2$ in the measure accounts for the change in inner products at each angle. The denominator makes ν a probability measure, and

$$\bar{c}_{ij}(\nu) = \frac{\int |p_s(z)|^2 c_{ij}(\theta)\, d\eta(\theta)}{\int |p_s(z)|^2\, d\eta}.$$

Thus the averaged inner products and the prescribed values $q^t v$ are divided by the same positive number. This preserves every interpolation inequality and gives the values $q^t \varphi_0$.

The inverse change is

$$d\eta = \frac{|d_s(z)|^2\, d\nu}{\int |d_s(z)|^2\, d\nu}, \qquad v = \frac{\varphi_0}{\int |d_s(z)|^2\, d\nu}.$$

Since $p_s(z) d_s(z) = 1$, substitution recovers both η and v. For fixed $s > 1$ and $q > q_0(s)$, the weighting functions $|p_s(z)|^2$ and $|d_s(z)|^2$ are continuous and strictly positive on $[0,\pi]$. They are therefore bounded above and away from zero. Both changes are valid for arbitrary probability measures and preserve their supports.

At each fixed angle, the same positive factor $|p_s(z)|^2$ multiplies every sum of interpolation terms around a cycle. It therefore also multiplies every fixed linear combination of these sums. The signs

and zeros of these functions are unchanged. This explains why the exact certificates below also certify the angular conditions (10) in the original trajectory normalization.

Translate the minimizer to zero and subtract its function value, so $x_\star = 0$ and $f(0) = 0$. Set $c = (s^2-1)^{-1}$ and $h(x) = f(x) - \|x\|^2/2$. Define the gradient operator and transformed iterates by

$$(\Delta y)_t = \nabla h(y_t) = \sigma_t, \qquad r_t = y_t - c\sigma_t.$$

The function h is convex with an (s^2-1)-Lipschitz gradient and $\nabla h(0) = 0$. Thus Δ is bounded with constant s^2-1 and is causal, with σ_t depending only on y_t.

Treating σ as an input, the remaining NAG equations form a linear recurrence with scalar transfer

$$G(z) = p_s(z) + c.$$

The same transfer acts in each coordinate. It describes the response with zero initial values. The prescribed initialization contributes a separate response.

Since $p_s(z) \to -c$ as $|z| \to \infty$, we have $G(z) \to 0$. Thus G is *strictly causal*: its response at time t uses only inputs σ_j with $j < t$. Its only pole is $1 - 1/s$, with multiplicity two.

Rescaling both the input and output at time t by $q^{-t/2}$ gives the transfer

$$G_q(\zeta) = G(\sqrt{q}\,\zeta).$$

Its pole is $(1-1/s)/\sqrt{q} < 1$ because $q > q_0(s)$, so G_q is stable.

We next construct the IQC from inequalities valid for any sequence of points $(y_t)_{t\geq 0}$, with σ_t and r_t defined above.

Fix one of the separator cycles $\gamma = (n_0, \dots, n_m = n_0)$ and let $n_\gamma = \max_j n_j$. For these inequalities, extend the samples by $(y_t, \sigma_t, r_t) = (0,0,0)$ for $t < 0$. Each added triple is the sample at the minimizer, since $\nabla h(0) = 0$.

The relation $r_t \in \partial\varphi(\sigma_t)$, with φ defined in (7), therefore also holds at negative indices. For every integer shift $t \geq -n_\gamma$, cyclic monotonicity gives

$$a_{\gamma,t} = \sum_{j=0}^{m-1} \left\langle r_{t+n_j}, \sigma_{t+n_j} - \sigma_{t+n_{j+1}} \right\rangle \geq 0. \tag{31}$$

The largest index in this sum is $t + n_\gamma$. For $t < -n_\gamma$ all entries are zero, so the sum vanishes.

Theorem C.5 requires one IQC valid for every $\tau \in [0,1]$. For such a τ, define

$$f_\tau(x) = \frac{1}{2}\|x\|^2 + \tau h(x).$$

Since h is convex and (s^2-1)-smooth, f_τ is 1-strongly convex and has gradient Lipschitz constant at most $1 + \tau(s^2-1) \leq s^2$. Thus $f_\tau \in \mathcal{F}_{1,s^2}$, with minimizer zero.

Using the same coefficient c, its transformed samples are

$$\sigma_t^\tau = \tau\sigma_t, \qquad r_t^\tau = y_t - c\sigma_t^\tau.$$

Let $a_{\gamma,t}^\tau$ denote the cycle sum (31) evaluated at these samples. Applying (31) to f_τ gives, for every integer shift $t \geq -n_\gamma$,

$$a_{\gamma,t}^\tau \geq 0, \qquad 0 \leq \tau \leq 1. \tag{32}$$

The added zero samples remain valid because every f_τ has minimizer zero.

For each sequence $u = y, \sigma, r$, define $\bar{u}_t = q^{-t/2}u_t$. Since each cycle sum is quadratic in r and σ, define $\bar{a}_{\gamma,t} = q^{-t}a_{\gamma,t}$. Substituting $u_{t+n_j} = q^{(t+n_j)/2}\bar{u}_{t+n_j}$ into (31) gives

$$\bar{a}_{\gamma,t} = \sum_{j=0}^{m-1} \left\langle q^{n_j/2}\bar{r}_{t+n_j}, q^{n_j/2}\bar{\sigma}_{t+n_j} - q^{n_{j+1}/2}\bar{\sigma}_{t+n_{j+1}} \right\rangle \geq 0. \tag{33}$$

The coefficients in this quadratic expression are independent of t. Rescaling r^τ and σ^τ in the same way gives the same formula with $(\bar{r}^\tau, \bar{\sigma}^\tau)$. The corresponding sum $\bar{a}_{\gamma,t}^\tau = q^{-t}a_{\gamma,t}^\tau$ is nonnegative by (32).

At time $k = t + n_\gamma \geq 0$, each sample index is $t + n_j = k - (n_\gamma - n_j)$. The delays $n_\gamma - n_j$ lie between zero and n_γ, so the sum uses samples from a finite window ending at k. For a vector sequence u, define

$$(\omega_\gamma u)_k = \begin{bmatrix} u_k \\ u_{k-1} \\ \vdots \\ u_{k-n_\gamma} \end{bmatrix}, \qquad k \geq 0,$$

with $u_j = 0$ for $j < 0$. This defines a causal finite-delay filter, also called a *finite-impulse-response (FIR) filter*.

Since $\bar{r} = \bar{y} - c\bar{\sigma}$, (33) is a quadratic expression in the delayed values of $\bar{y}$ and $\bar{\sigma}$. Thus there is a real symmetric matrix M_γ such that

$$\bar{a}_{\gamma,t} = \left(\omega_\gamma \begin{bmatrix} \bar{y} \\ \bar{\sigma} \end{bmatrix}\right)_k^* M_\gamma \left(\omega_\gamma \begin{bmatrix} \bar{y} \\ \bar{\sigma} \end{bmatrix}\right)_k.$$

Its coefficients are independent of k and τ.

The transfer matrix of this filter is

$$\omega_\gamma(z) = \begin{bmatrix} I \\ z^{-1}I \\ \vdots \\ z^{-n_\gamma}I \end{bmatrix}.$$

Each delay of j iterations contributes the factor z^{-j}. Thus $\omega_\gamma u$ denotes the filter applied to a sequence, whereas $\omega_\gamma(e^{i\theta})$ denotes its transfer matrix evaluated at $z = e^{i\theta}$. The associated *FIR multiplier* is the bounded Hermitian matrix function

$$\Pi_\gamma(e^{i\theta}) = \omega_\gamma(e^{i\theta})^* M_\gamma \omega_\gamma(e^{i\theta}).$$

The filter has zero initial state because its stored negative-time entries are zero. Using $k = t + n_\gamma$ and summing the scaled cycle inequalities gives, for every integer $T \geq 0$ and every $\tau \in [0, 1]$,

$$\sum_{k=0}^{T} \left(\omega_\gamma \begin{bmatrix} \bar{y} \\ \bar{\sigma}^\tau \end{bmatrix}\right)_k^* M_\gamma \left(\omega_\gamma \begin{bmatrix} \bar{y} \\ \bar{\sigma}^\tau \end{bmatrix}\right)_k = \sum_{t=-n_\gamma}^{T-n_\gamma} \bar{a}^\tau_{\gamma,t} \geq 0.$$

Every term on the right is nonnegative by (32) and the positive rescaling q^{-t}. This is a *hard IQC*: the inequality holds at every finite horizon, without assuming square summability.

Using the nonnegative weights from the strict separator, define

$$\Pi(e^{i\theta}) = \sum_\gamma \lambda_\gamma \Pi_\gamma(e^{i\theta}).$$

The sum is finite, so Π is a bounded Hermitian FIR multiplier. Summing the preceding inequalities with these weights gives its hard IQC for every $\tau \in [0, 1]$. The same multiplier therefore works for the whole family.

For any sequence $\bar{y}$, recover the original points as $y_t = q^{t/2}\bar{y}_t$. Evaluating ∇h at these points and rescaling gives the operator

$$(\Delta_q \bar{y})_t = q^{-t/2} \nabla h(q^{t/2}\bar{y}_t).$$

For $\tau \in [0, 1]$, set $\Delta_{q,\tau} = \tau\Delta_q$, so $\bar{\sigma}^\tau = \Delta_{q,\tau}\bar{y}$. Each output at time t depends only on $\bar{y}_t$, so these operators are causal.

Since $\nabla h(0) = 0$ and ∇h is $(s^2 - 1)$-Lipschitz, the rescaling factors cancel in the estimate

$$\|(\Delta_{q,\tau}\bar{y})_t\| \leq \tau q^{-t/2}(s^2 - 1)q^{t/2}\|\bar{y}_t\| = \tau(s^2 - 1)\|\bar{y}_t\|.$$

Squaring and summing over t proves that these operators are bounded:

$$\|\Delta_{q,\tau}\bar{y}\|_{\ell_2} \leq \tau(s^2 - 1)\|\bar{y}\|_{\ell_2}, \qquad \bar{y} \in \ell_2^d. \tag{34}$$

We next evaluate the multiplier on the frequency response of the NAG recurrence. Fix $\theta \in [0, \pi]$ and set $z = \sqrt{q}\, e^{i\theta}$. Since $G_q(e^{i\theta}) - c = p_s(z)$, the transfer from $\bar{\sigma}$ to $\bar{r}$ at this frequency is $p_s(z)$.

The offset n_j in (33) corresponds to delay $n_\gamma - n_j$ and weight $q^{n_j/2}$, giving the factor

$$q^{n_j/2} e^{-i(n_\gamma - n_j)\theta} = e^{-in_\gamma\theta} z^{n_j}.$$

The common factor $e^{-in_\gamma\theta}$ has modulus one and cancels from the quadratic form. Equation (29) therefore gives

$$\begin{bmatrix} G_q(e^{i\theta}) \\ I \end{bmatrix}^* \Pi_\gamma(e^{i\theta}) \begin{bmatrix} G_q(e^{i\theta}) \\ I \end{bmatrix} = |p_s(z)|^2 c_\gamma(s, q, \theta) I.$$

Recall that $\phi(\theta) = \sum_\gamma \lambda_\gamma c_\gamma(s, q, \theta) < 0$ for every $\theta \in [0, \pi]$. Since p_s is continuous and nonzero on the circle $|z| = \sqrt{q}$, the product $|p_s(\sqrt{q}\, e^{i\theta})|^2 \phi(\theta)$ is continuous and strictly negative.

Its maximum on the compact interval $[0, \pi]$ is therefore negative, so define

$$\epsilon = -\max_{\theta \in [0,\pi]} \big(|p_s(\sqrt{q}\, e^{i\theta})|^2 \phi(\theta) \big) > 0.$$

Using $\Pi = \sum_\gamma \lambda_\gamma \Pi_\gamma$, the weighted sum of the preceding identities gives

$$\begin{bmatrix} G_q(e^{i\theta}) \\ I \end{bmatrix}^* \Pi(e^{i\theta}) \begin{bmatrix} G_q(e^{i\theta}) \\ I \end{bmatrix} = |p_s(\sqrt{q}\, e^{i\theta})|^2 \phi(\theta) I \preceq -\epsilon I. \tag{35}$$

We check the conditions of Theorem C.5 for G_q and Δ_q. For any $(e_w, e_v) \in \ell^d \times \ell^d$ and $\tau \in [0, 1]$, strict causality of G_q determines v_t from $e_{v,t}$ and earlier w values. Then $w_t = e_{w,t} + \tau(\Delta_q v)_t$. This recursion defines a unique causal solution, so the interconnection is well posed.

For every $\bar{y} \in \ell_2^d$, the bound in (34) gives $\bar{\sigma}^\tau \in \ell_2^d$. The FIR outputs are square summable, so their quadratic series converge absolutely. Taking limits in the finite-horizon inequalities and applying Parseval gives the IQC for the same Π and every $\tau \in [0, 1]$.

Real coefficients extend (35) to the whole unit circle by complex conjugation. The theorem therefore gives ℓ_2 stability. In particular, for any input $d_{\mathrm{fr}} \in \ell_2^d$, the equations

$$\bar{y} = G_q \bar{\sigma} + d_{\mathrm{fr}}, \qquad \bar{\sigma} = \Delta_q \bar{y}$$

have $(\bar{y}, \bar{\sigma}) \in \ell_2^d \times \ell_2^d$.

The theorem establishes square summability for each objective. We now derive a norm bound with a constant independent of the objective and dimension.

The finite-delay construction gives $\Pi = \omega^* M \omega$. Let T be the linear operator with transfer matrix

$$T(e^{i\theta}) = \begin{bmatrix} G_q(e^{i\theta}) & I \\ I & 0 \end{bmatrix}.$$

The feedback equations then give

$$\begin{bmatrix} \bar{y} \\ \bar{\sigma} \end{bmatrix} = T \begin{bmatrix} \bar{\sigma} \\ d_{\mathrm{fr}} \end{bmatrix}.$$

In the following matrix inequalities, T and Π denote $T(e^{i\theta})$ and $\Pi(e^{i\theta})$. These inequalities hold for every θ.

Stability of G_q and the FIR construction of Π imply that T and Π are bounded on the unit circle. By (35), we can choose $\alpha > 0$ small enough that

$$T^* \Pi T + \alpha T^* T = \begin{bmatrix} A & B \\ B^* & D \end{bmatrix}, \qquad A \preceq -\frac{\epsilon}{2} I.$$

Since T and Π are bounded on the unit circle, so are the blocks B, D. The bound on A gives $\|A^{-1}\| \le 2/\epsilon$. Thus we can choose $b_0 > 0$ such that $D - B^* A^{-1} B \prec b_0 I$ for every θ.

The Schur complement gives

$$T^* \Pi T + \alpha T^* T \preceq b_0 \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}. \tag{36}$$

All blocks of T and Π are scalar multiples of I with coefficients depending only on s, q, and the finite separator. Hence α, b_0 can be chosen independently of the objective and dimension.

The IQC theorem gives $\bar{y}, \bar{\sigma} \in \ell_2^d$ for every $d_{\mathrm{fr}} \in \ell_2^d$.

Apply (36) to the Fourier transforms of $\bar{\sigma}$ and d_{fr}, and integrate over the unit circle. The integral involving Π is nonnegative by the IQC for $\bar{y}$ and $\bar{\sigma}$.

Dropping this term and applying Parseval gives

$$\alpha\big(\|\bar{y}\|_{\ell_2}^2 + \|\bar{\sigma}\|_{\ell_2}^2\big) \le b_0 \|d_{\mathrm{fr}}\|_{\ell_2}^2. \tag{37}$$

We now account for the initialization $x_{-1} = x_0$. The *free response* is the sequence of query points produced by the linear recurrence from this initialization with $\sigma_t = 0$.

Solving the recurrence and rescaling gives

$$(d_{\mathrm{fr}})_t = \left(1 + \frac{t}{s+1}\right)\left(\frac{1 - 1/s}{\sqrt{q}}\right)^t x_0, \qquad t \ge 0.$$

By linearity, the actual trajectory satisfies $\bar{y} = G_q \bar{\sigma} + d_{\mathrm{fr}}$.

Since $q > q_0(s)$, the ratio $(1 - 1/s)/\sqrt{q}$ is less than one. Thus d_{fr} is square summable, and its ℓ_2 norm is bounded by a finite multiple of $\|x_0\|$ depending only on s, q.

The NAG update gives $s^2 x_{t+1} = (s^2 - 1) y_t - \sigma_t$. By Cauchy–Schwarz and the definition of the ℓ_2 norm, for every $t \ge 0$,

$$q^{-(t+1)} \|x_{t+1}\|^2 \le \frac{(s^2-1)^2 + 1}{s^4 q}\big(\|\bar{y}\|_{\ell_2}^2 + \|\bar{\sigma}\|_{\ell_2}^2\big).$$

By (37) and the bound on $\|d_{\mathrm{fr}}\|_{\ell_2}$, there is a finite $b(s, q)$ such that

$$\|x_t\|^2 \le b(s,q) q^t \|x_0\|^2, \qquad t \ge 1.$$

Enlarging $b(s, q)$ to at least one also covers $t = 0$. The constants depend only on s, q and the chosen separator. Thus $b(s, q)$ is uniform over the objective, dimension, and horizon.

Undoing the translation of the minimizer gives the claimed bound. □

C.4 Assembly of the lower and upper bounds

Remark C.7 (The case $\kappa = 1$). For $f \in \mathcal{F}_{1,1}$, strong convexity and smoothness at the minimizer give

$$\frac{1}{2}\|x - x_\star\|^2 \le f(x) - f(x_\star) \le \frac{1}{2}\|x - x_\star\|^2.$$

Hence $f(x) = f(x_\star) + \frac{1}{2}\|x - x_\star\|^2$ and $\nabla f(x) = x - x_\star$ for every x. When $\kappa = 1$, the momentum coefficient is $\beta_1 = 0$ and the step size is one. Thus $x_{t+1} = x_t - \nabla f(x_t) = x_\star$ for every $t \ge 0$. Both asymptotic factors in (2)–(3) are therefore zero.

Proof of Theorem 4.1. Fix $\kappa > 1$ and set $s = \sqrt{\kappa}$. Proposition 3.2 gives

$$q_0(s) \le q_\star(s) \le 1 - 1/s < 1.$$

We first prove $q_\star(s) \le \rho_{\mathrm{fix}}(s^2)$ by constructing one fixed objective and initial point that attain $q_\star(s)$.

If $q_\star(s) = q_0(s)$, the scalar quadratic $f(x) = x^2/2$ attains this factor from any nonzero standard initialization by (4).

Suppose $q_\star(s) > q_0(s)$. Lemma C.3 supplies a pair $(\nu_\star, \varphi_0^\star)$ satisfying (10) at $q = q_\star(s)$, with $\nu_\star$ supported on finitely many angles. The associated contracting and rotating trajectory uses at most two real coordinates per support angle, so it is finite-dimensional. By Lemma 3.1, the corresponding interpolation values can be chosen as $\varphi_t = q_\star(s)^t \varphi_0^\star$. Lemma C.4 then gives a single objective in $\mathcal{F}_{1,s^2}$ and an initialization $x_{-1} = x_0 \ne x_\star$ whose NAG trajectory has asymptotic squared-distance factor $q_\star(s)$.

Thus in both cases,

$$q_\star(s) \le \rho_{\mathrm{fix}}(s^2) \le \rho_{\mathrm{hor}}(s^2).$$

For the upper bound, fix any $q \in (q_\star(s), 1)$. Since $q_\star(s) = \max \mathcal{Q}(s)$ by (13), the interpolation conditions (10) are infeasible at this q. By Lemma 3.1, no probability measure on $[0, \pi]$ satisfies all finite-cycle inequalities. Lemma C.1 therefore gives a nonnegative combination of finitely many cycle costs that is strictly negative at every angle.

Since $q > q_0(s)$, Proposition C.6 gives a finite $b(s, q) \ge 1$ that depends only on s, q. Taking the supremum over objectives, dimensions, and initial points yields

$$\sup_{d \ge 1} \sup_{f \in \mathcal{F}_{1,s^2}(\mathbf{R}^d)} \sup_{x_0 \ne x_\star} R_t(f, x_0) \le b(s, q) q^t, \qquad t \ge 0.$$

Taking tth roots and the limit superior, with q fixed, gives

$$\rho_{\mathrm{hor}}(s^2) \le \lim_{t \to \infty} b(s, q)^{1/t} q = q.$$

This holds for every $q \in (q_\star(s), 1)$, so letting $q \downarrow q_\star(s)$ gives $\rho_{\mathrm{hor}}(s^2) \le q_\star(s)$. The constant $b(s, q)$ need not remain bounded as $q \downarrow q_\star(s)$, since the limit in t was taken first. Combining this upper bound with the lower bound proves (14) for $\kappa > 1$.

The bounds in (13) come from the quadratic example and Nesterov's classical estimate (Nesterov, 2018; Sanz-Serna & Zygalakis, 2021). Substituting $q_0(s) = (1 - 1/s)^2$ and $s = \sqrt{\kappa}$ gives (15). Both bounds tend to zero as $\kappa \downarrow 1$ and to one as $\kappa \to \infty$, so $q_\star(\sqrt{\kappa})$ has the same limits. □

D Proof of the sharp large-condition-number expansion

Proof of Corollary 4.3. Write $s = \sqrt{\kappa}$. The quadratic lower bound in (15) already implies $q_\star(s) \to 1$, so it suffices to prove

$$\lim_{s \to \infty} -s \log q_\star(s) = \frac{4}{3}. \tag{38}$$

We first show that, for each fixed $0 < \lambda < 4/3$ and all sufficiently large s, no $q \in [e^{-\lambda/s}, 1)$ is feasible for (10). We then construct a feasible measure at $q = e^{-2\alpha/s}$ for each fixed $2/3 < \alpha < 3/4$ and all sufficiently large s. These give upper and lower bounds on $q_\star(s)$, respectively, and the inequalities reverse for $-s \log q_\star(s)$.

Upper bound. Fix $q \in (q_0(s), 1)$ and a pair (ν, φ_0) satisfying (10). Because $d_s(z)$ is bounded on $|z| = \sqrt{q} < 1$, the terms in (9) defining $c_{t0}(\theta)$ tend to zero uniformly in θ as $t \to \infty$. Hence $\bar{c}_{t0}(\nu) \to 0$, so the lower inequality in (10) gives $\varphi_0 \ge 0$ in the limit. The upper inequality with $t = 1$ then gives $\bar{c}_{01}(\nu) \ge 0$. Put $r = \sqrt{q}$ and $x = 2r \cos\theta \in [-2r, 2r]$. Equations (8) and (9) give

$$c_{01}(\theta) = -\frac{s^2 - 1}{2s^2 q^2} \psi_s(q, x),$$

where $\psi_s(q, x)$ and its unconstrained minimizer x_{v} are given in (25) and (26). If $x_{\mathrm{v}} \ge 2r$, the minimum of $\psi_s(q, x)$ over $[-2r, 2r]$ is its value at the right endpoint, which is positive by (26). Hence $c_{01}(\theta) < 0$ for every angle, contradicting $\bar{c}_{01}(\nu) \ge 0$. Fix $0 < \lambda < 4/3$. At $q = e^{-\lambda/s}$,

$$s(2\sqrt{q} - x_{\mathrm{v}}) \longrightarrow \frac{3\lambda}{2} - 2 < 0.$$

Since $3\lambda/2 - 2 < 0$, the displayed limit implies $x_{\mathrm{v}} > 2r$ at $q = e^{-\lambda/s}$ for all sufficiently large s. To cover every larger $q < 1$, write the condition $x_{\mathrm{v}} \ge 2r$ as

$$r(r^2 s^2 + 3s^2 - 4s + 1) - 4(s - 1)^2 \ge 0.$$

Its left-hand side is strictly increasing in r for $s > 1$, so the condition also holds for every $q \in [e^{-\lambda/s}, 1)$. Since $e^{-\lambda/s} > q_0(s)$ for all sufficiently large s, the preceding argument rules out feasibility throughout this interval. Therefore $q_\star(s) \le e^{-\lambda/s}$ and $-s \log q_\star(s) \ge \lambda$. Taking the limit inferior as $s \to \infty$ and then letting $\lambda \uparrow 4/3$ gives

$$\liminf_{s \to \infty} -s \log q_\star(s) \ge \frac{4}{3}.$$

The same asymptotic upper bound follows from Gannot's result (Gannot, 2022, Proposition 2.13).

Lower bound. Fix $2/3 < \alpha < 3/4$ and set $r = e^{-\alpha/s}$ and $q = r^2$. We seek a probability measure ν on $[0, \pi]$ such that $\bar{c}_{t0}(\nu) \geq 0$ and $\bar{c}_{0t}(\nu) \geq 0$ for every integer $t \geq 1$. Then (10) holds with $\varphi_0 = 0$. For $0 < h < r$, let ν be the probability measure on $[0, \pi]$ with density

$$\frac{(1-h^2)^5}{\pi(1+4h^2+h^4)|1-he^{i\theta}|^6}. \tag{39}$$

This density gives explicit formulas for the two integrals needed to check feasibility. Write $z = re^{i\theta}$ and $m_s = \int \operatorname{Re} \overline{d_s(z)}\, d\nu$. Equation (9) gives

$$\bar{c}_{t0}(\nu) = q^t m_s - \int \operatorname{Re}\left[\overline{d_s(z)} z^t\right] d\nu,$$
$$\bar{c}_{0t}(\nu) = m_s - \int \operatorname{Re}\left[\overline{d_s(z)} \bar{z}^t\right] d\nu.$$

The subscripts 1 and 2 below correspond to the integrals containing z^t and $\bar{z}^t$ in these two costs, respectively. To evaluate them, use the absolutely convergent series

$$(1-he^{i\theta})^{-3} = \sum_{k=0}^{\infty} \binom{k+2}{2} h^k e^{ik\theta}.$$

Substitute this series and its complex conjugate into the density, and use (8). After taking real parts, all nonconstant terms integrate to zero because $\int_0^\pi \cos(j\theta)\, d\theta = 0$ for every nonzero integer j. With

$$\delta = s^2r^2(h^4+4h^2+1) - 6sr(s-1)h(h^2+1) + 6(s-1)^2h^2,$$
$$e_1 = sr - (s-1)h, \qquad e_2 = srh - (s-1),$$

this gives

$$m_s = -\frac{(s^2-1)\delta}{s^2r^2(1+4h^2+h^4)}.$$

For $\delta \neq 0$ and every integer $t \geq 0$, we obtain

$$\int \operatorname{Re}\left[\overline{d_s(z)} z^t\right] d\nu = m_s(rh)^t \left(1 + a_1 t + b_1 \binom{t}{2}\right),$$
$$\int \operatorname{Re}\left[\overline{d_s(z)} \bar{z}^t\right] d\nu = m_s(rh)^t \left(1 + a_2 t + b_2 \binom{t}{2}\right), \tag{40}$$

where the coefficients are

$$a_1 = -\frac{(h^2-1)e_1[(s-1)h^3 + srh^2 - 4(s-1)h + 2sr]}{\delta}, \qquad b_1 = \frac{(h^2-1)^2 e_1^2}{\delta},$$
$$a_2 = -\frac{(h^2-1)e_2[srh^2 - 3(s-1)h + 2sr]}{h\delta}, \qquad b_2 = \frac{(h^2-1)^2 e_2^2}{h^2\delta}.$$

To satisfy (10) with $\varphi_0 = 0$, it suffices to have

$$m_s > 0, \qquad a_1 \leq r/h - 1, \qquad a_2 \leq 1/(rh) - 1. \tag{41}$$

Indeed, $m_s > 0$ implies $\delta < 0$, and hence $b_1 \leq 0$ and $b_2 \leq 0$. For $u \geq 0$, $a \leq u$, and $b \leq 0$, Bernoulli's inequality gives

$$1 + at + b\binom{t}{2} \leq 1 + ut \leq (1+u)^t, \qquad t = 0, 1, \ldots.$$

Apply this with $u = r/h - 1$ and $u = 1/(rh) - 1$, respectively. The first integral in (40) is then at most $m_s q^t$, and the second is at most m_s. Thus both averaged costs are nonnegative for every integer $t \geq 1$.

We now choose h to satisfy (41). Write $\epsilon = 1/s$ and $h = e^{-\eta\epsilon}$, and define the two slacks

$$g_1 = r/h - 1 - a_1, \qquad g_2 = 1/(rh) - 1 - a_2.$$

We will choose η so that these slacks are equal and positive. Set

$$\gamma = \sqrt{\frac{3\alpha(1-\alpha)^2}{3\alpha-2}}.$$

The function $g(\epsilon,\eta) = (g_1 - g_2)/\epsilon$ extends smoothly to a neighborhood of $(0,\gamma)$, with

$$g(0,\eta) = -2\,\frac{3\alpha(1-\alpha)^2 - (3\alpha-2)\eta^2}{3\alpha^2 - 6\alpha - \eta^2 + 3}.$$

Our choice of γ gives $g(0,\gamma) = 0$, and

$$\partial_\eta g(0,\gamma) = \frac{4(3\alpha-2)\gamma}{3\alpha^2 - 6\alpha - \gamma^2 + 3} \neq 0.$$

The denominator equals $6(\alpha-1)^3/(3\alpha-2) < 0$. The implicit function theorem therefore gives a smooth choice $\gamma_s = \gamma + O(s^{-1})$ for which $g_1 = g_2$ when $\eta = \gamma_s$. For this choice, Taylor expansion gives

$$g_1 = g_2 = \frac{\alpha}{2(3\alpha-2)} s^{-2} + O(s^{-3}). \tag{42}$$

The same substitution gives

$$m_s \longrightarrow -\frac{3\alpha^2 - 6\alpha - \gamma^2 + 3}{3} = \frac{2(1-\alpha)^3}{3\alpha-2} > 0.$$

Thus m_s and both slacks are positive for all sufficiently large s. Also, $\gamma^2 - \alpha^2 = \alpha(3-4\alpha)/(3\alpha-2) > 0$, so $\gamma > \alpha$. Consequently, $h = e^{-\gamma_s/s}$ satisfies $0 < h < r < 1$ for all sufficiently large s. Since $\alpha < 1$, we have $r = e^{-\alpha/s} > 1 - \alpha/s > 1 - 1/s$, so $q > q_0(s)$. The measure ν therefore satisfies (10) with $\varphi_0 = 0$ at $q = e^{-2\alpha/s}$ for all sufficiently large s. Hence $q_\star(s) \geq e^{-2\alpha/s}$, which gives

$$\limsup_{s\to\infty} -s \log q_\star(s) \leq 2\alpha.$$

Letting $\alpha \downarrow 2/3$ and combining with the upper bound proves (38). Since $(-\log u)/(1-u) \to 1$ as $u \uparrow 1$, this is equivalent to (16). The square-root expansion follows from $\sqrt{1-v} = 1 - v/2 + o(v)$. □

E EXACT CHARACTERIZATION OF $q_\star$ FOR $\kappa = 10$

We explain how the interpolation inequalities determine the algebraic data for the planar example at $\kappa = 10$. For completeness we also provide full polynomial coefficients further below. Throughout, we identify $\mathbb{C}$ with $\mathbf{R}^2$, so $\langle a, b\rangle = \mathrm{Re}(\bar{a}b)$ and $\|a\| = |a|$. The momentum parameter is

$$\beta = \beta_{10} = \frac{\sqrt{10}-1}{\sqrt{10}+1} = \frac{11 - 2\sqrt{10}}{9}.$$

Candidate query points, gradients, and function values. Let $z = \sqrt{q}\, e^{i\theta}$, where $0 < q < 1$, and put $\chi = z + \bar{z} = 2\sqrt{q}\cos\theta$. Thus q is the squared contraction factor, while the real coordinate χ encodes the rotation angle. Define the query and gradient amplitudes

$$e \doteq e(z) = 1 + \beta - \beta z^{-1}, \qquad g \doteq g(z) = 10(e - z).$$

These expressions come directly from the momentum and gradient steps on a geometric trajectory; in particular, $e - g/10 = z$. We first consider only its contracting part, with candidate samples

$$\widehat{y}_k \doteq z^k e, \qquad \widehat{g}_k \doteq z^k g, \qquad \widehat{f}_k \doteq q^k F, \qquad k \geq 0,$$

where F is an unknown scalar function value. The hats distinguish these samples from the full trajectory: below they become the queries with indices $k+1$, after a separate initial query. We must still verify that the samples belong to one function in $\mathcal{F}_{1,10}$.

The inequalities producing the polynomial. For $\kappa = 10$, the interpolation condition (5) reads

$$f_i - f_j \geq \langle g_j, y_i - y_j \rangle + \frac{1}{20}\|g_i - g_j\|^2 + \frac{5}{9}\left\| y_i - y_j - \frac{g_i - g_j}{10} \right\|^2 .$$

To substitute the candidate samples, define

$$B \doteq B(q, z) = \frac{|g|^2}{20} + \frac{5q}{9},$$

and, for each integer $m \geq 1$, we have to compute the interpolation inequalities for $(i, j) = (0, m)$ and $(i, j) = (m, 0)$. To this end observe that for the case $(i, j) = (0, m)$ we have $g_j = z^m g$ and $y_i - y_j = (1 - z^m)e$ and for $(i, j) = (m, 0)$ we have $g_j = g$ and $y_i - y_j = (z^m - 1)e$. Further for the first configuration we have $g_i - g_j = (1 - z^m)g$ and $y_i - y_j = (1 - z^m)e$, so that using $e - g/10 = z$ we obtain $y_i - y_j - \frac{g_i - g_j}{10} = (1 - z^m)(e - \frac{g}{10}) = (1 - z^m)z$, so that

$$\frac{1}{20}\|g_i - g_j\|^2 + \frac{5}{9}\left\| y_i - y_j - \frac{g_i - g_j}{10} \right\|^2 = |1 - z^m|^2 B,$$

follows; we obtain the same quadratic term when expanding for the second case. With this we arrive at the right hand sides for the two interpolation inequalities

$$\begin{aligned} C_m^+(q, z) &\doteq \langle z^m g, (1 - z^m)e \rangle + |1 - z^m|^2 B, \\ C_m^-(q, z) &\doteq \langle g, (z^m - 1)e \rangle + |1 - z^m|^2 B, \end{aligned} \tag{43}$$

and we obtain the bounds

$$C_m^+(q, z) \leq (1 - q^m)F \leq -C_m^-(q, z)$$

Shifting both indices by k turns $(0, m)$ and $(m, 0)$ into $(k, k + m)$ and $(k + m, k)$, respectively, and multiplies either inequality by the positive factor q^k. The factor $1 - q^m$ is then included in the difference of function values:

$$\widehat{f}_k - \widehat{f}_{k+m} = q^k F - q^{k+m} F = q^k (1 - q^m) F.$$

We now select two different members of this family: the forward inequality with $m = 1$ and the reverse inequality with $m = 18$. These are the bounds that are active (hold with equality) at the optimal $\kappa = 10$ example; the choice of eighteen is specific to this solution. Taking $m = 1$ in the forward inequality and shifting by $k = 0, \ldots, 17$ gives the consecutive comparisons $(0, 1), (1, 2), \ldots, (17, 18)$. Taking $m = 18$ in the reverse inequality gives $(18, 0)$, which closes this chain. The particular inequalities we sum are

$$\begin{aligned} \widehat{f}_k - \widehat{f}_{k+1} &\geq q^k C_1^+(q, z), && k = 0, \ldots, 17, \\ \widehat{f}_{18} - \widehat{f}_0 &\geq C_{18}^-(q, z). \end{aligned} \tag{44}$$

After summing we obtain for the left-hand sides the two terms:

$$\sum_{k=0}^{17} (\widehat{f}_k - \widehat{f}_{k+1}) = (1 - q)F \sum_{k=0}^{17} q^k = (1 - q^{18})F,$$
$$\widehat{f}_{18} - \widehat{f}_0 = -(1 - q^{18})F.$$

Thus F is cancelled out from the summed inequalities. The first eighteen comparisons go from index 0 to index 18, and the last returns to index 0, i.e., a cycle of interpolation comparisons. Consequently, with $H_{18}(q) \doteq \sum_{k=0}^{17} q^k$, feasibility requires

$$W_{18}(q, z) \doteq H_{18}(q) C_1^+(q, z) + C_{18}^-(q, z) \leq 0.$$

Equivalently, the two active comparisons bound the same function value:

$$\frac{C_1^+(q, z)}{1 - q} \leq F \leq -\frac{C_{18}^-(q, z)}{1 - q^{18}}.$$

When both are equalities, $W_{18}(q, z) = 0$ and these bounds coincide.

To express W_{18} using only the real quantities q and χ, introduce the sums of conjugate powers

$$T_j \doteq z^j + \bar{z}^j, \qquad j \geq 0.$$

Since $z + \bar{z} = \chi$ and $z\bar{z} = q$, multiplying out gives $\chi T_j = T_{j+1} + qT_{j-1}$ for $j \geq 1$. Thus these sums satisfy

$$T_0 = 2, \qquad T_1 = \chi, \qquad T_{j+1} = \chi T_j - qT_{j-1}, \quad j \geq 1.$$

This recurrence expresses each T_j as a polynomial in (q, χ); for example, $T_2 = \chi^2 - 2q$ and $T_3 = \chi^3 - 3q\chi$. We use these polynomials to replace the complex powers in (43). In particular, the common quadratic factor becomes

$$|1 - z^m|^2 = 1 + q^m - (z^m + \bar{z}^m) = 1 + q^m - T_m(q, \chi).$$

For the inner-product terms, substitute $\bar{z} = q/z$ and expand into real parts of powers of z, then use

$$\mathrm{Re}(z^j) = \frac{T_j(q,\chi)}{2}, \qquad \mathrm{Re}(z^{-j}) = \frac{T_j(q,\chi)}{2q^j}, \qquad j \geq 1.$$

Applying these replacements to $C_1^+(q, z)$ and $C_{18}^-(q, z)$ eliminates z and $\bar{z}$ from their combination $W_{18}(q, z)$. Simplifying the resulting rational expression in (q, χ) and factoring gives the exact identity

$$W_{18}(q, z) = \frac{1 + q - \chi}{9q} P(q, \chi), \tag{45}$$

via $|z|^2 = q$ and $\chi = z + \bar{z}$. Here P is a polynomial with real coefficients of total degree 19 in the two real variables q, χ; in fact $P \in \mathbb{Q}(\sqrt{10})[q, \chi] \subseteq \mathbb{R}[q, \chi]$ and its complete coefficient table appears at the end of this section. The factor multiplying P is positive, since $1+q-\chi = |1-z|^2 \geq (1-\sqrt{q})^2 > 0$. Thus this cycle requires $P(q, \chi) \leq 0$, and its active boundary is at equality, i.e., when $P(q, \chi) = 0$; we call $P(q, \chi)$ the *active-cycle polynomial*.

Selecting an exact candidate. We seek a value of q for which the graph of $\chi \mapsto P(q, \chi)$ touches the horizontal axis at a minimum strictly between the endpoints $-2\sqrt{q}$ and $2\sqrt{q}$. Since $\chi = 2\sqrt{q}\cos\theta$, this means that the trajectory's rotation angle lies strictly between $0°$ and $180°$. At such a minimum, both the polynomial and its derivative vanish. We therefore look for candidate pairs satisfying

$$P(q, \chi) = 0, \qquad \partial_\chi P(q, \chi) = 0. \tag{46}$$

These are necessary conditions, not a guarantee that the graph has the desired shape. The box below selects one candidate, denoted $(q_{10}, \chi_\star)$; we then establish that it gives the required minimum and, separately, the optimal rate. In particular, the identification $q_{10} = q_\star(\sqrt{10})$ remains to be established by a realizable trajectory and a matching upper bound. Specifically, our pair is the unique solution in the rational box

$$\begin{aligned} |q - 0.56442327797289351191064244846217017114 26| &< 10^{-40}, \\ |\chi - 1.45320111296571604837294212481974106632 14| &< 10^{-40}. \end{aligned} \tag{47}$$

The decimal centers here are exact rational numbers defining the box. Directed interval arithmetic certifies that the Jacobian of $(P, \partial_\chi P)$ is nonsingular throughout this box and that an interval Newton image lies strictly inside it, establishing existence and uniqueness. The computation also establishes $q_0(\sqrt{10}) < q_{10} < 1 - 1/\sqrt{10}$ and $\chi_\star^2 < 4q_{10}$. The latter places $\chi_\star$ strictly inside the allowed interval $[-2\sqrt{q_{10}}, 2\sqrt{q_{10}}]$. Since $\chi = 2\sqrt{q}\cos\theta$, this corresponds to an angle strictly between 0 and π, rather than either real-mode endpoint. This establishes the interiority of the candidate.

To show that the stationary zero actually touches zero rather than crosses it, we use the exact polynomial factorization

$$P(q_{10}, \chi) = (\chi - \chi_\star)^2 R(\chi), \tag{48}$$

where R has degree 17. Interval bounds on 64 rational subintervals certify $R(\chi) > 17$ for $|\chi| \leq 1.503$, an interval containing $[-2\sqrt{q_{10}}, 2\sqrt{q_{10}}]$. Hence $P(q_{10}, \chi)$ is nonnegative throughout the allowed frequency interval and vanishes only at $\chi_\star$: the curve genuinely touches zero at this interior point; see Figure 3 for an illustration. This establishes the contact for the exact candidate q_{10}, but does not yet show that $q_{10} = q_\star(\sqrt{10})$. The remaining interpolation conditions and the matching upper bound below are essential.

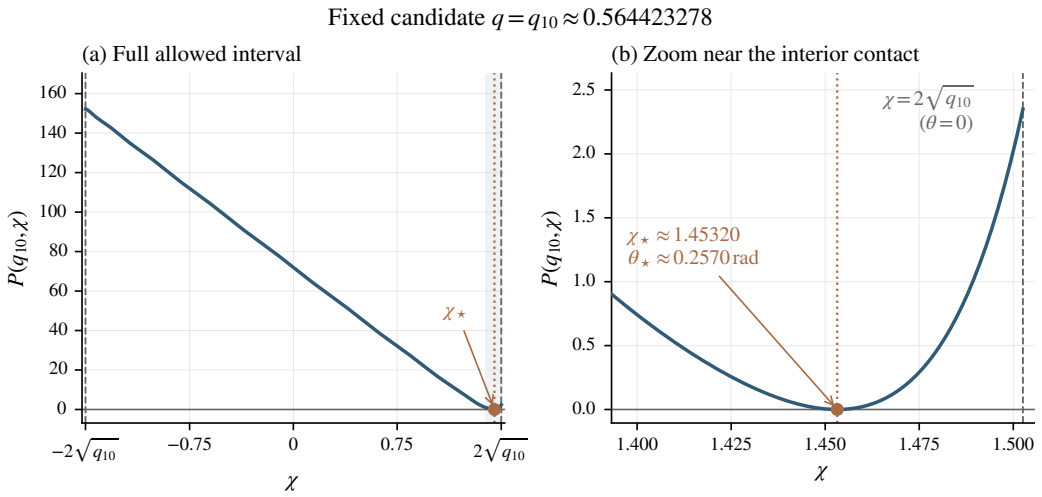


Figure 3: Numerical illustration of $\chi \mapsto P(q_{10}, \chi)$ at the (algebraic) candidate q_{10}. Left: the full allowed interval $[-2\sqrt{q_{10}}, 2\sqrt{q_{10}}]$, with the shaded region enlarged on the right. Right: the contact at $\chi_\star$ is strictly inside the interval, away from the real-mode endpoint $\chi = 2\sqrt{q_{10}}$ ($\theta = 0$). The marked angle is $\theta_\star = \arccos(\chi_\star/(2\sqrt{q_{10}}))$. Here P comes from one active comparison cycle, whereas Figures 4 and 6 plot the weighted dual functions Φ_{100} and $\Phi_\dagger$ combining several interpolation constraints. All three use the real frequency coordinate χ.

The trajectory with the standard initialization realizing q_{10}. We now show that q_{10} is realizable, i.e., we construct a trajectory that contracts with a factor q_{10} and thus $q_\star(\sqrt{10}) \geq q_{10}$. Set

$$z_\star \doteq \frac{\chi_\star}{2} + i\sqrt{q_{10} - \frac{\chi_\star^2}{4}}, \qquad e_\star \doteq 1 + \beta_{10} - \beta_{10} z_\star^{-1}, \qquad g_\star \doteq 10(e_\star - z_\star).$$

The active forward equality explicitly determines the function-value amplitude:

$$F_\star \doteq \frac{C_1^+(q_{10}, z_\star)}{1 - q_{10}} = 0.63565632612066339 98\ldots. \tag{49}$$

We translate coordinates to place the minimizer at the origin and subtract the minimum value from the objective. The full query points, gradients, and function values are

$$\begin{aligned} y_0 &= z_\star^{-1}, \qquad g_0 = 10(z_\star^{-1} - 1), \qquad f_0 = \frac{5}{2}, \\ y_{t+1} &= z_\star^t e_\star, \quad g_{t+1} = z_\star^t g_\star, \qquad f_{t+1} = q_{10}^t F_\star, \qquad t \geq 0. \end{aligned} \tag{50}$$

In particular, the initial value $f_0 = 5/2$ is different from the amplitude $F_\star$ of the contracting samples. Starting with $x_{-1} = x_0 = z_\star^{-1}$, the first gradient step gives $x_1 = y_0 - g_0/10 = 1$. Thereafter $e_\star - g_\star/10 = z_\star$ makes the Nesterov identities give $x_t = z_\star^{t-1}$ for every $t \geq 0$. Consequently the normalized squared distance is exactly q_{10}^t, as in (17), including at $t = 0$.

Checking that one objective realizes the trajectory. We still have to show that the prescribed points, gradients, and function values come from one function in $\mathcal{F}_{1,10}$. For the contracting part, this means checking precisely the bounds already derived:

$$C_m^+(q_{10}, z_\star) \leq (1 - q_{10}^m)F_\star \leq -C_m^-(q_{10}, z_\star), \qquad m \geq 1. \tag{51}$$

Equality in the left inequality at $m = 1$ follows from the definition of $F_\star$ in (49). Equality in the right inequality at $m = 18$ follows from the cycle identity (45) and $P(q_{10}, \chi_\star) = 0$; this is an exact identity. Outward-rounded interval arithmetic checks both inequalities for $m = 1, \ldots, 400$. Apart from the two equalities for $m = 1$ and $m = 18$, each checked inequality has a margin greater than $3 \cdot 10^{-6}$, where the margin is the right-hand side minus the left-hand side. This finite check alone

does not cover all iteration indices. To handle the remaining ones, observe that $z_\star^m \to 0$ and $q_{10}^m \to 0$, so the bounds in (51) tend to

$$B(q_{10}, z_\star) \le F_\star \le \langle g_\star, e_\star \rangle - B(q_{10}, z_\star). \tag{52}$$

Both limiting inequalities also have margins greater than $3 \cdot 10^{-6}$. They are exactly the interpolation bounds between the contracting sample $(e_\star, g_\star, F_\star)$ and the minimizer's data $(0, 0, 0)$. Expanding the formulas for $C_m^\pm$ and using the triangle inequality shows that each finite-index margin differs from its limiting margin by at most

$$(F_\star + 2|g_\star||e_\star| + 3B(q_{10}, z_\star))\, q_{10}^{m/2}.$$

This quantity is less than 10^{-49} for every $m \ge 401$, so both inequalities remain strict at all remaining indices.

The interpolation inequalities for all distinct pairs are covered by three cases:

1. *Two contracting queries.* For $1 \le i < j$, set $m = j - i$ and multiply (51) by q_{10}^{i-1}. Using the data in (50), these are exactly the interpolation inequalities (5) for the ordered pairs (i, j) and (j, i).
2. *A contracting query and the minimizer.* For query $i \ge 1$, multiply (52) by q_{10}^{i-1} to obtain both ordered comparisons with the minimizer's data $(0, 0, 0)$.
3. *The initial query with index* 0. Substitute (50) into (5) for $(0, j)$ and $(j, 0)$ with $j \ge 1$, and for both comparisons between the initial query and the minimizer. The infinitely many comparisons with contracting queries are covered by finite interval checks and geometric bounds for the remaining indices.

Together with interpolation, this establishes *one objective* realizing the entire trajectory, not a different objective for each horizon. Its normalized squared distance is q_{10}^t, which gives the lower bound $q_\star(\sqrt{10}) \ge q_{10}$.

Ruling out larger asymptotic factors. The upper bound must apply to every admissible objective, not just the one constructed above. We obtain it from the same polynomial P and Proposition C.6. That proposition applies when a cycle has a strictly negative cost at every angle. We will establish this for the cycle $\gamma = (0, 1, \ldots, 18, 0)$ from (44), with $q > q_{10}$ sufficiently close to q_{10}.

First we relate W_{18} to the cycle cost used in that proposition. The change of variables (6)–(7) gives, at $\kappa = 10$,

$$\varphi_i + f_i = \langle g_i, y_i \rangle - \frac{1}{2}\|y_i\|^2 - \frac{1}{18}\|g_i - y_i\|^2.$$

Substituting in (5) converts its lower bounds on $f_i - f_j$ into upper bounds on $\varphi_i - \varphi_j$. The additional sample-dependent terms telescope around the cycle, reversing the sign of the summed costs. There is also a scaling factor: for the contracting samples defining $C_m^\pm$, we have $e - (g - e)/9 = 10z/9$, so their r-sequence is $(10z/9)z^k$, rather than z^k in (8). The inner products are therefore multiplied by $|10z/9|^2 = 100q/81$. Thus the cycle cost from Lemma C.1 satisfies

$$c_\gamma(\sqrt{10}, q, \theta) = -\frac{81}{100q} W_{18}(q, z) = -\frac{9(1 + q - \chi)}{100q^2} P(q, \chi), \qquad \chi = 2\sqrt{q}\cos\theta. \tag{53}$$

The last equality uses (45). Since $1 + q - \chi > 0$, it remains to show that $P(q, \chi) > 0$ throughout the allowed interval for q just above q_{10}, not merely at one angle.

Using the coefficients in Table 1, we differentiate $P(q, \chi)$ in (57) with respect to q and evaluate the derivative with outward-rounded intervals on (47), obtaining

$$114 < \partial_q P(q_{10}, \chi_\star) < 115. \tag{54}$$

By continuity, $\partial_q P(q, \chi) > 0$ near $(q_{10}, \chi_\star)$. For $q > q_{10}$ and χ in this neighborhood,

$$P(q, \chi) = P(q_{10}, \chi) + \int_{q_{10}}^{q} \partial_q P(u, \chi)\, du > 0,$$

because (48) makes the first term nonnegative. Away from a fixed small neighborhood of $\chi_\star$, that factorization and $R > 17$ give a strictly positive minimum on the remaining closed subset

of $[-1.503, 1.503]$. Continuity preserves this positivity for q sufficiently close to q_{10}. Moreover, $2\sqrt{q} < 1.503$ for such q, so this fixed interval still contains the entire allowed interval. Consequently there exists $\varepsilon > 0$, with $q_{10} + \varepsilon < 1$, such that

$$P(q, \chi) > 0 \quad \text{whenever} \quad q_{10} < q < q_{10} + \varepsilon, \qquad |\chi| \le 2\sqrt{q}. \tag{55}$$

For each such q, (53) is strictly negative at every angle, so this single cycle satisfies the strict alternative in Lemma C.1. Since $q > q_{10} > q_0(\sqrt{10})$, Proposition C.6 now gives

$$\|x_t - x_\star\|^2 \le b(\sqrt{10}, q) q^t \|x_0 - x_\star\|^2, \qquad t \ge 0, \tag{56}$$

for every finite-dimensional $f \in \mathcal{F}_{1,10}$ with the standard initialization. The finite constant depends only on q and the fixed condition number, not on f, the dimension, x_0, or t. For $x_0 \neq x_\star$ and $t \ge 1$, (56) gives

$$R_t(f, x_0)^{1/t} \le b(\sqrt{10}, q)^{1/t} q.$$

The same bound holds after taking the supremum over objectives, dimensions, and starting points. Letting $t \to \infty$ therefore bounds both rates in (2)–(3) by q. This holds for every $q \in (q_{10}, q_{10} + \varepsilon)$; letting $q \downarrow q_{10}$ gives the upper bound q_{10}. Combining it with the realizing trajectory and Theorem 4.1 proves

$$\rho_{\mathrm{fix}}(10) = \rho_{\mathrm{hor}}(10) = q_\star(\sqrt{10}) = q_{10}.$$

The limit in q is taken after the asymptotic limit in t; no bound on $b(\sqrt{10}, q)$ as $q \downarrow q_{10}$ is needed. The tangency equations alone are insufficient: the positive factorization, the derivative bound (54), and the uniform-rate proposition are the additional ingredients of the upper bound.

Full polynomial coefficients. To specify the polynomial in (45) without any external data, write

$$P(q, \chi) = 5 \sum_{j=0}^{19} \sum_{k=0}^{19-j} (a_{jk} + b_{jk}\sqrt{10}) q^k \chi^j. \tag{57}$$

The table below lists all 210 integer pairs (a_{jk}, b_{jk}), including the zero coefficients. The common factor 5 is part of this normalization. For example, the terms with the two highest powers of χ are

$$5(11 - 2\sqrt{10})\chi^{19} + 5(11 - 2\sqrt{10} - 9q)\chi^{18}.$$

For the full set of values see the table.

Table 1: Complete coefficient table for (57). The coefficient of $q^k \chi^j$ is $5(a_{jk} + b_{jk}\sqrt{10})$. Read each of the three blocks from top to bottom. All other coefficients are zero.

j	k	a_{jk}	b_{jk}	j	k	a_{jk}	b_{jk}	j	k	a_{jk}	b_{jk}
19	0	11	−2	8	4	9195	−1320	3	4	885	−110
18	0	11	−2	8	5	−25707	3564	3	5	−1516	182
18	1	−9	0	8	6	16533	0	3	6	2394	−280
17	0	11	−2	8	7	495	0	3	7	−3560	408
17	1	−218	38	8	8	−165	0	3	8	5055	−570
17	2	1	0	8	9	45	0	3	9	−120	0
16	0	11	−2	8	10	−9	0	3	10	84	0
16	1	−207	36	8	11	1	0	3	11	−56	0
16	2	163	0	7	0	11	−2	3	12	35	0
16	3	1	0	7	1	−108	18	3	13	−20	0
15	0	11	−2	7	2	566	−88	3	14	10	0
15	1	−196	34	7	3	−2120	312	3	15	−4	0
15	2	1826	−304	7	4	6390	−900	3	16	1	0
15	3	−16	0	7	5	−16512	2244	2	0	11	−2
15	4	1	0	7	6	38016	−5016	2	1	−53	8
14	0	11	−2	7	7	−792	0	2	2	136	−18
14	1	−185	32	7	8	330	0	2	3	−260	32
14	2	1630	−270	7	9	−120	0	2	4	425	−50
14	3	−1230	0	7	10	36	0	2	5	−631	72

j	k	a_{jk}	b_{jk}	j	k	a_{jk}	b_{jk}	j	k	a_{jk}	b_{jk}
14	4	−15	0	7	11	−8	0	2	6	878	−98
14	5	1	0	7	12	1	0	2	7	−1166	128
13	0	11	−2	6	0	11	−2	2	8	1495	−162
13	1	−174	30	6	1	−97	16	2	9	−765	0
13	2	1445	−238	6	2	458	−70	2	10	−36	0
13	3	−8400	1330	6	3	−1554	224	2	11	28	0
13	4	105	0	6	4	4270	−588	2	12	−21	0
13	5	−14	0	6	5	−10122	1344	2	13	15	0
13	6	1	0	6	6	21504	−2772	2	14	−10	0
12	0	11	−2	6	7	−12936	0	2	15	6	0
12	1	−163	28	6	8	−462	0	2	16	−3	0
12	2	1271	−208	6	9	210	0	2	17	1	0
12	3	−6955	1092	6	10	−84	0	1	0	0	0
12	4	5005	0	6	11	28	0	1	1	−53	8
12	5	91	0	6	12	−7	0	1	2	72	−8
12	6	−13	0	6	13	1	0	1	3	−135	16
12	7	1	0	5	0	11	−2	1	4	154	−16
11	0	11	−2	5	1	−86	14	1	5	−217	24
11	1	−152	26	5	2	361	−54	1	6	236	−24
11	2	1108	−180	5	3	−1096	154	1	7	−299	32
11	3	−5684	884	5	4	2716	−364	1	8	318	−32
11	4	23075	−3458	5	5	−5852	756	1	9	−381	40
11	5	−364	0	5	6	11382	−1428	1	10	−2	2
11	6	78	0	5	7	−20472	2508	1	11	−19	2
11	7	−12	0	5	8	462	0	1	12	−4	2
11	8	1	0	5	9	−252	0	1	13	−17	2
10	0	11	−2	5	10	126	0	1	14	−6	2
10	1	−141	24	5	11	−56	0	1	15	−15	2
10	2	956	−154	5	12	21	0	1	16	−8	2
10	3	−4576	704	5	13	−6	0	1	17	−13	2
10	4	17391	−2574	5	14	1	0	1	18	−10	2
10	5	−11869	0	4	0	11	−2	0	0	0	0
10	6	−286	0	4	1	−75	12	0	1	−22	4
10	7	66	0	4	2	275	−40	0	2	60	−4
10	8	−11	0	4	3	−735	100	0	3	−22	4
10	9	1	0	4	4	1620	−210	0	4	60	−4
9	0	11	−2	4	5	−3136	392	0	5	−22	4
9	1	−130	22	4	6	5530	−672	0	6	60	−4
9	2	815	−130	4	7	−9090	1080	0	7	−22	4
9	3	−3620	550	4	8	5070	0	0	8	60	−4
9	4	12815	−1870	4	9	210	0	0	9	−22	4
9	5	−38522	5434	4	10	−126	0	0	10	38	0
9	6	715	0	4	11	70	0	0	11	20	0
9	7	−220	0	4	12	−35	0	0	12	18	0
9	8	55	0	4	13	15	0	0	13	20	0
9	9	−10	0	4	14	−5	0	0	14	18	0
9	10	1	0	4	15	1	0	0	15	20	0
8	0	11	−2	3	0	11	−2	0	16	18	0
8	1	−119	20	3	1	−64	10	0	17	20	0
8	2	685	−108	3	2	200	−28	0	18	18	0
8	3	−2805	420	3	3	−460	60	0	19	20	0

F AN EXACT TWO-COMPONENT OPTIMAL SOLUTION AT $\kappa = 100$

The $\kappa = 100$ case provides a nontrivial example of the finite-support property beyond one planar spiral. The calculation is similar to Appendix E and we only repeat key steps.

Proposition F.1 (An optimal solution requiring two components)**.** *At $\kappa = 100$,*

$$q_{100} = q_\star(\sqrt{100}) = 0.859899778052548365740446035906789693\ldots.$$

There is an optimal angular pair whose measure is supported on two frequencies, and every optimal angular measure has exactly these two frequencies in its support. The resulting exact self-similar tail

spans four real dimensions; after enforcing $x_{-1} = x_0$, one fixed function attaining this factor can be chosen in $\mathbf{R}^5$.

Proof. We state the exact certificate in the normalization $\sigma_t = z^t$, as in Appendix C.3. Let

$$\xi_\star \doteq (q_{100}, v_\star, w_\star, \chi_{1\star}, \chi_{2\star}, \lambda_{49}, \lambda_{63}, \lambda_{64})$$

be the unique zero in the rational box of radius 10^{-58} around

$$\begin{aligned} q_{100} &= 0.859899778052548365740446035906789693\ldots, \\ v_\star &= 0.002854749733611906316731334872665651\ldots, \\ w_\star &= 0.432893789374821564486049487863967595\ldots, \\ \chi_{1\star} &= 1.847585923705342323643049369488243114\ldots, \\ \chi_{2\star} &= 1.833337063888573702334307789415459859\ldots. \end{aligned}$$

Here $\chi_j \doteq 2\sqrt{q}\cos\theta_j$ is a *real frequency coordinate*. The corresponding angles are

$$\theta_1 = 0.08709509851058300453\ldots, \qquad \theta_2 = 0.15162655445126233148\ldots,$$

and the certificate measure is $\eta_\star \doteq w_\star\delta_{\theta_1} + (1-w_\star)\delta_{\theta_2}$. The optimal angular pair $(\nu_\star, \varphi_0^\star)$ satisfying (10) at $q = q_{100}$ is obtained from $(\eta_\star, v_\star)$ by (30).

To specify the algebraic system, fix $s = \sqrt{100} = 10$ and define, for integers $m \geq 1$,

$$L_m(q,\chi) \doteq -\frac{|p_{10}(z)|^2 c_{m0}(\theta)}{1-q^m}, \qquad U_m(q,\chi) \doteq \frac{|p_{10}(z)|^2 c_{0m}(\theta)}{1-q^m},$$

where $z = \sqrt{q}\,e^{i\theta}$, $\theta = \arccos(\chi/(2\sqrt{q}))$, and c_{ij} is from (9) at $s = 10$. Write

$$\langle A\rangle_w \doteq wA(q,\chi_1) + (1-w)A(q,\chi_2),$$

and set

$$\Phi_{100}(q,\chi) \doteq \lambda_{49}L_{49}(q,\chi) + \lambda_{63}L_{63}(q,\chi) + \lambda_{64}L_{64}(q,\chi) - U_1(q,\chi).$$

Dependence on the dual weights is suppressed in Φ_{100}; derivatives with respect to χ keep those weights fixed. The eight equations are

$$\langle U_1\rangle_w = v, \qquad \langle L_{49}\rangle_w = \langle L_{63}\rangle_w = \langle L_{64}\rangle_w = v, \tag{58}$$

$$\lambda_{49} + \lambda_{63} + \lambda_{64} = 1, \qquad \Phi_{100}(q,\chi_1) = \partial_\chi\Phi_{100}(q,\chi_1) = \partial_\chi\Phi_{100}(q,\chi_2) = 0. \tag{59}$$

The omitted equality $\Phi_{100}(q,\chi_2) = 0$ follows from (58) and the positive component weights. An interval Krawczyk inclusion proves existence and uniqueness of the root in the rational box. All three dual weights are positive and equal approximately to

$$(\lambda_{49}, \lambda_{63}, \lambda_{64}) \approx (0.7275254636036373,\ 0.0389115479752263,\ 0.2335629884211364).$$

Figure 4 plots $\chi \mapsto \Phi_{100}(q_{100},\chi)$, keeping the three dual multipliers fixed at the same algebraic solution. Its two contacts correspond to the two rotation angles above.

The four-dimensional self-similar tail. Set

$$z_j \doteq \frac{\chi_{j\star}}{2} + i\sqrt{q_{100} - \frac{\chi_{j\star}^2}{4}}, \qquad a_1 \doteq \sqrt{w_\star}, \qquad a_2 \doteq \sqrt{1-w_\star}.$$

Thus $|z_1|^2 = |z_2|^2 = q_{100}$. In $\mathbb{C}^2 \cong \mathbf{R}^4$, define

$$\begin{aligned} &\sigma_t \doteq (a_1 z_1^t, a_2 z_2^t), \\ &r_t \doteq (a_1 p_{10}(z_1) z_1^t, a_2 p_{10}(z_2) z_2^t), && x_{t+1} \doteq \frac{99}{100} r_t, \qquad t \in \mathbb{Z}, \\ &y_t \doteq r_t + \frac{1}{99}\sigma_t, && g_t \doteq y_t + \sigma_t, \qquad t \geq 0. \end{aligned} \tag{60}$$

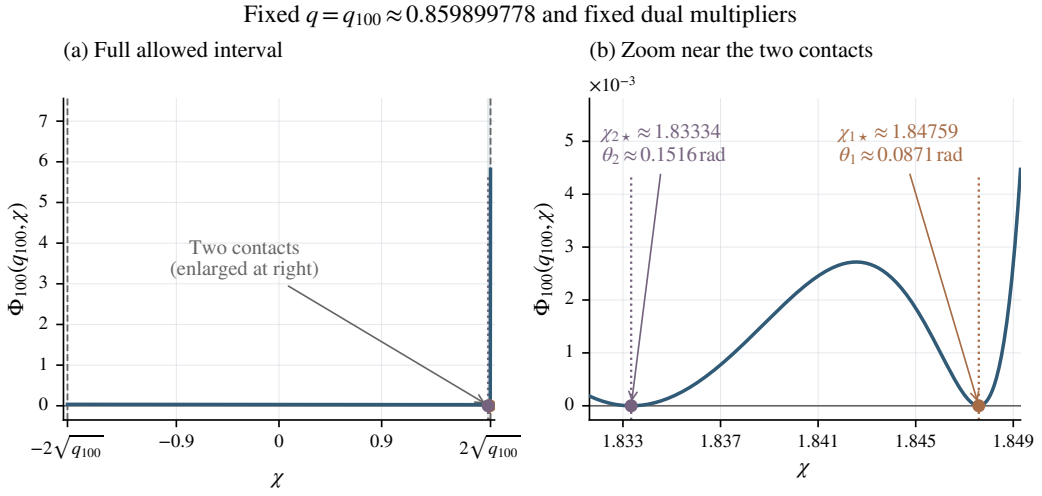


Figure 4: Numerical illustration of $\Phi_{100}(q_{100}, \chi)$ with q and the dual multipliers fixed. Left: the full allowed interval $[-2\sqrt{q_{100}}, 2\sqrt{q_{100}}]$. Right: enlargement of the shaded region, showing both contacts $\chi_{1\star}$ and $\chi_{2\star}$, labeled by their rotation angles $\theta_j = \arccos(\chi_{j\star}/(2\sqrt{q_{100}}))$.

Here p_{10} means (28) with $s = 10$. Its definition makes these identities exactly consistent with the Nesterov momentum and gradient steps. In particular,

$$\frac{\|x_{t+1}\|^2}{\|x_1\|^2} = q_{100}^t, \qquad t \geq 0. \tag{61}$$

The four eigenvalues $z_1, \bar{z}_1, z_2, \bar{z}_2$ are pairwise distinct and all four modal coefficients are nonzero. The associated four-column Vandermonde determinant is therefore nonzero, so the real span of the tail is exactly four-dimensional. For completeness, with the objective minimum normalized to $f(0) = 0$, the corresponding function values can be taken as

$$f_t \doteq \langle \sigma_t, y_t \rangle - q_{100}^t v_\star - \frac{1}{198}\|\sigma_t\|^2 + \frac{1}{2}\|y_t\|^2. \tag{62}$$

Indeed, after fixing the additive constant by $\varphi(0) = 0$, the potential $u_t \doteq (1 - q_{100}^t)v_\star = v_\star - \varphi(\sigma_t)$ satisfies every pairwise difference constraint by Lemma 3.1. Reversing the transform (7) gives exactly (60)– (62), so all of these countably many samples belong to one function in $\mathcal{F}_{1,100}(\mathbf{R}^4)$.

Why two components are necessary. As in the case for $\kappa = 10$ in Appendix E, we exactly check the two interpolation inequalities for each index $m = 1, \ldots, 180$ and then use geometric bounds for both families at larger indices. Equality holds only in the upper inequality at $m = 1$ and the lower inequalities at $m = 49, 63, 64$. The interval calculation also proves the global factorization

$$\Phi_{100}(q_{100}, \chi) = \frac{(\chi - \chi_{1\star})^2(\chi - \chi_{2\star})^2 R_{62}(\chi)}{D(\chi)}, \qquad \frac{R_{62}(\chi)}{D(\chi)} > 0, \quad |\chi| \leq 2\sqrt{q_{100}}.$$

For any feasible angular measure ν at $q = q_{100}$, the corresponding angular measure η from Appendix C.3 satisfies $\int \Phi_{100}(q_{100}, \chi(\theta))\, d\eta(\theta) \leq 0$, while the factorization gives $\Phi_{100}(q_{100}, \chi) \geq 0$. We call its two zeros the *contact frequencies*, because the dual inequality is tight there. Such a measure must therefore be supported on these two frequencies, and the change of weights preserves support. Neither contact works alone: directed interval arithmetic proves

$$L_{56}(q_{100}, \chi_{1\star}) - U_1(q_{100}, \chi_{1\star}) > 0.00511062771457,$$
$$L_{35}(q_{100}, \chi_{2\star}) - U_1(q_{100}, \chi_{2\star}) > 0.0281844327967.$$

Each singleton therefore has an empty feasible interval for φ_0. Every optimal angular measure consequently has support cardinality exactly two, not merely at most two. This counts angles in $[0, \pi]$; a disk measure may split the mass at each angle between conjugate points.

For the matching upper bound, combine the directed costs for $\sigma_t = z^t$ and $r_t = p_{10}(z)z^t$ with positive coefficients derived from the same dual weights. Denote the resulting function by $S(r, \theta)$, where $q = r^2$, and set

$$N(r, \theta) \doteq 99|10re^{i\theta} - 9|^4 S(r, \theta).$$

The multiplicative factor is positive on the relevant rectangle. At $r = \sqrt{q_{100}}$, the factorization above gives $S(r, \theta) \leq 0$. The checker proves $\partial_r N(r, \theta) < -0.585$ throughout $\sqrt{q_{100}} \leq r \leq \sqrt{0.9}$ by covering this rectangle with 1544 directed rational boxes. Hence $S(r, \theta) < 0$ for every $r > \sqrt{q_{100}}$ in this interval. Its coefficients remain nonnegative weights on finite cycle inequalities, so every $q_{100} < q \leq 0.9$ is infeasible. The classical estimate-sequence bound already gives $q_\star(10) \leq 0.9$. Thus

$$\rho_{\mathrm{fix}}(100) = \rho_{\mathrm{hor}}(100) = q_\star(10) = q_{100}.$$

The exact tail (60) occupies four real dimensions and does not by itself enforce $x_{-1} = x_0$. The splice in Lemma C.4 scales this tail and adds one orthogonal scalar quadratic coordinate whose squared rate is strictly below $q_0(10) = 0.81 < q_{100}$. It therefore produces one function in $\mathcal{F}_{1,100}(\mathbf{R}^5)$ with the standard initialization and asymptotic factor q_{100}. Figure 5 shows an orthogonal three-dimensional projection of the genuine four-dimensional tail; it is not a three-dimensional realization. □

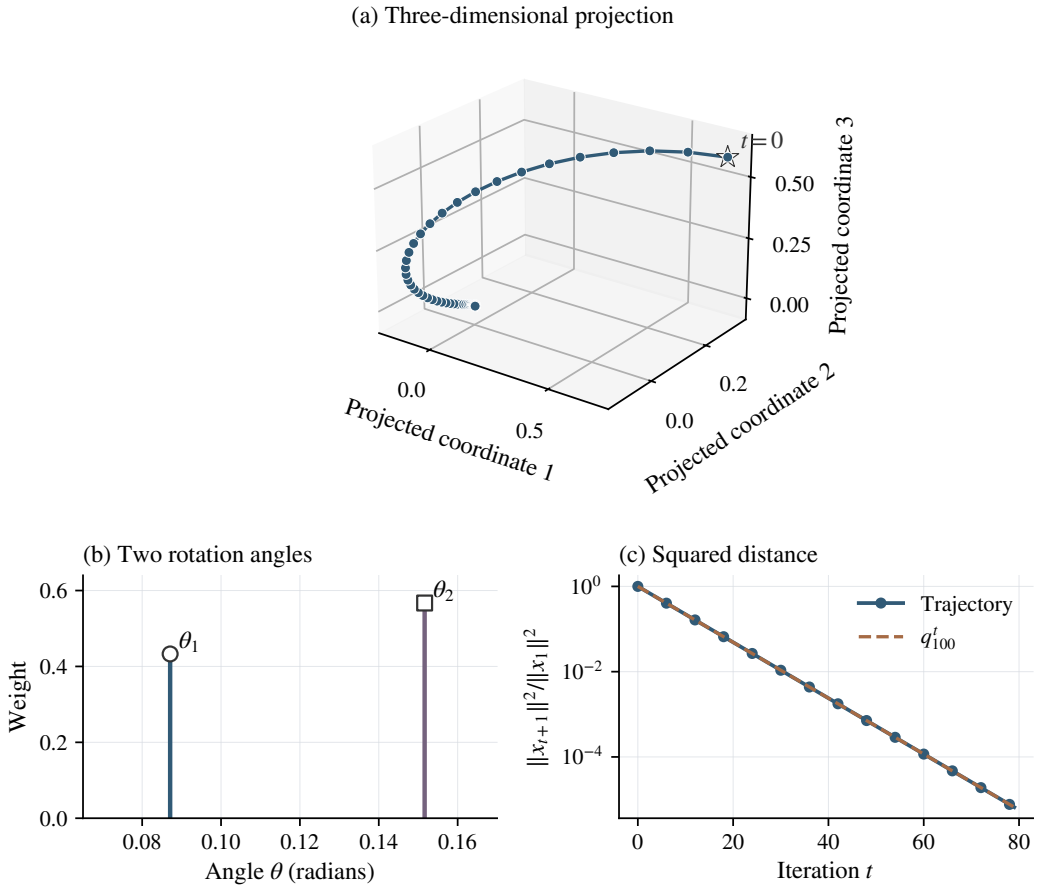


Figure 5: The $\kappa = 100$ example, certified by exact arithmetic. (a) An orthogonal three-dimensional projection of the four-dimensional trajectory after initialization. The star marks $t = 0$. (b) The two rotation angles and their positive weights. (c) The normalized squared norm in four dimensions equals q_{100}^t. The construction with standard initialization has a fifth, faster scalar coordinate, which is not shown. The exact certificate and trajectory data are included in the supplement.

G ALGEBRAIC DEGREE AND FORMULAS BY RADICALS

We prove Remark 4.2 using another exact example with one contracting and rotating component. Here the interpolation comparisons $(0,1)$, $(5,0)$, and $(6,0)$ are active, meaning that they hold with equality. We first identify the optimal squared-distance factor, then show that it has an algebraic specification but cannot be expressed by radicals over the rationals. This rules out typical descriptions of convergence rates expressed as radicals but we stress that it does not rule out other notions of closed form.

G.1 TWO ACTIVE COMPARISON CYCLES AND AN EXACT RATE

Write $z = \sqrt{q}\, e^{i\theta}$, where $0 < q < 1$ and $\theta \in [0, \pi]$, and put $\chi = z + \bar{z} = 2\sqrt{q}\cos\theta$. Thus q is the squared contraction factor, while the real coordinate χ encodes the rotation angle, as in Appendix E. For $s = \sqrt{\kappa} > 1$, write $\alpha_s \doteq 1 - 1/s$, so $sz - (s-1) = s(z - \alpha_s)$. At fixed s, the inverse transfer in (28) is $p_s(z) = G(z) - (s^2-1)^{-1}$: it maps σ to r, while G maps σ to y. The examples in Appendices E and F have κ fixed and as such do not need α_s.

We again sum interpolation comparisons around a cycle. For an integer $m \geq 1$, take the consecutive comparisons $(k, k+1)$ for $k = 0, \ldots, m-1$ and the closing comparison $(m, 0)$. The values on the left-hand sides cancel, leaving a constraint associated with $(0, 1, \ldots, m, 0)$. To express it in real coordinates, as in Appendix E we introduce the sums of conjugate powers $T_j \doteq z^j + \bar{z}^j$, which satisfy

$$T_0 = 2, \qquad T_1 = \chi, \qquad T_{j+1} = \chi T_j - q T_{j-1},$$

for $j \geq 1$. In the polynomial family below, α is a placeholder for a real number that we later set to α_s. Clearing the transfer-function denominator in the cycle expansion produces real parts involving $z^2(\bar{z} - \alpha)^2$ and its product with $\bar{z}$. We abbreviate these by

$$\begin{aligned} D &\doteq 2\,\mathrm{Re}\big[z^2(\bar{z}-\alpha)^2\big] = 2q^2 - 2\alpha q\chi + \alpha^2(\chi^2 - 2q), \\ E &\doteq 2\,\mathrm{Re}\big[z^2(\bar{z}-\alpha)^2\bar{z}\big] = (q^2 + \alpha^2 q)\chi - 4\alpha q^2, \\ H_m &\doteq \sum_{j=0}^{m-1} q^j. \end{aligned}$$

The polynomial expressions follow from $z\bar{z} = q$ and $z + \bar{z} = \chi$. In particular, $D - E$ collects the contribution with the factor $1 - \bar{z}$. For $m \geq 1$, define the polynomial

$$P_m(\alpha, q, \chi) \doteq \frac{(D-E)H_m + q^m D - q^2 T_m + 2\alpha q T_{m+1} - \alpha^2 T_{m+2}}{\chi - 1 - q}. \tag{63}$$

The numerator is divisible by $\chi - 1 - q$, so the quotient is a polynomial. To see this, substitute $\chi = 1 + q$. The recurrence gives $T_j = 1 + q^j$, and the numerator becomes

$$(q-\alpha)^2\big[(1-q)H_m + q^m - 1\big] = 0,$$

where the last equality uses the finite geometric-sum identity $(1-q)H_m = 1 - q^m$. Expanding the cycle cost from (11) gives the identity

$$\mathcal{C}^s_{(0,1,\ldots,m,0)}(z) = \frac{1 + q - \chi}{2} P_m(\alpha_s, q, \chi). \tag{64}$$

The prefactor is positive on $|\chi| \leq 2\sqrt{q}$, $0 < q < 1$. We will use $m = 5$ and $m = 6$, corresponding to the two cycles $(0, 1, \ldots, 5, 0)$ and $(0, 1, \ldots, 6, 0)$. These are two cycles of interpolation comparisons for a single rotating component, not two components as in Appendix F, nor periodic trajectories.

Selecting an exact candidate. We follow the same approach as before in Appendix E. Set

$$s_\dagger \doteq \frac{1421}{1250}, \qquad \alpha_\dagger \doteq 1 - \frac{1}{s_\dagger} = \frac{171}{1421}, \qquad \kappa_\dagger \doteq s_\dagger^2 = \frac{2019241}{1562500}.$$

Here $s_\dagger$, and therefore $\alpha_\dagger$ and $\kappa_\dagger$, are exact rational numbers. Consequently,

$$P_5(\alpha_\dagger, q, \chi),\ P_6(\alpha_\dagger, q, \chi) \in \mathbb{Q}[q, \chi].$$

In contrast, the polynomial we used for $\kappa = 10$ has coefficients involving $\sqrt{10}$ and belongs to $\mathbb{Q}(\sqrt{10})[q, \chi]$. We use this separate construction so that the elimination and Galois argument can be carried out directly over $\mathbb{Q}$. It is the rational choice of $s_\dagger$, not just of the condition number, that ensures rational coefficients. The rational data are the fixed parameters and polynomial coefficients; the unknown roots need not be rational.

Let $(q_\dagger, \chi_\dagger)$ be the common zero

$$P_5(\alpha_\dagger, q, \chi) = P_6(\alpha_\dagger, q, \chi) = 0 \tag{65}$$

in the rational isolating box of radius 10^{-72} around

$$q_\dagger = 0.0259177578383525961624\ldots, \qquad \chi_\dagger = 0.2515525327377247329537\ldots.$$

As before interval Newton inclusion proves existence and uniqueness there, as well as $q_0(s_\dagger) < q_\dagger < 1$ and $\chi_\dagger^2 < 4q_\dagger$.

Lemma G.1 (Identification of the exact rate)**.** *The isolated root satisfies $q_\star(s_\dagger) = q_\dagger$.*

Proof. Interpolation inequalities and the lower bound. We use the normalization of Appendix C.3, with p_s defined in (28). For $z = \sqrt{q}\, e^{i\theta}$, $s = s_\dagger$, and $q > q_0(s)$, define

$$K_m(z) \doteq (1 - \bar{z})\frac{1 - q^m}{1 - q} + q^m - z^m, \qquad C_m^\sigma(q, \chi) \doteq \mathrm{Re}[p_s(z)K_m(z)].$$

Here C_m^σ is the sum of the interpolation terms around $(0, 1, \ldots, m, 0)$ for $\sigma_t = z^t$, $r_t = p_s(z)z^t$. Unlike $C_m^\pm$ in Appendix E, it is a whole-cycle cost, not the right-hand side of an individual comparison. These cycle costs are rational functions of (q, χ), and

$$\mathcal{C}^s_{(0,1,\ldots,m,0)}(z) = (s^2 - 1)|z - \alpha_s|^4 C_m^\sigma(q, \chi).$$

Since $s = s_\dagger$ is now fixed, $\alpha_s = \alpha_\dagger$ and $|z - \alpha_\dagger|^2 = q - \alpha_\dagger\chi + \alpha_\dagger^2$. Combining this identity with (64) gives the explicit relation to the polynomials:

$$C_m^\sigma(q, \chi) = \frac{1 + q - \chi}{2(s_\dagger^2 - 1)(q - \alpha_\dagger\chi + \alpha_\dagger^2)^2} P_m(\alpha_\dagger, q, \chi). \tag{66}$$

The common factor is independent of m and strictly positive for $q_0(s_\dagger) < q < 1$ and $|\chi| \le 2\sqrt{q}$. In particular, the two active polynomial equations are exactly $C_5^\sigma = C_6^\sigma = 0$.

The isolated pair $(q_\dagger, \chi_\dagger)$ gives the complex multiplier $z_\dagger \doteq \chi_\dagger/2 + i\sqrt{q_\dagger - \chi_\dagger^2/4}$ for the contracting and rotating component. Choose the value at index zero of the convex interpolation function φ from (7) to be

$$v \doteq \frac{\mathrm{Re}[p_s(z_\dagger)(1 - \bar{z}_\dagger)]}{1 - q_\dagger}.$$

With $\varphi_t \doteq q_\dagger^t v$, the interpolation inequalities for $(0, m)$ and $(m, 0)$ from (10), expressed in this normalization, are equivalent to

$$0 \le C_m^\sigma(q_\dagger, \chi_\dagger) \le \mathrm{Re}\, p_s(z_\dagger)|1 - z_\dagger^m|^2, \qquad m \ge 1. \tag{67}$$

The lower equalities at $m = 5, 6$ and the upper equality at $m = 1$ follow algebraically. They correspond to the active comparisons $(5, 0)$, $(6, 0)$, and $(0, 1)$, respectively. The upper inequalities at $m = 5, 6$ are strict because $\mathrm{Re}\, p_s(z_\dagger) > 0$ and $|z_\dagger| < 1$. Outward-rounded interval arithmetic verifies the positive real part and all remaining inequalities for $m = 1, \ldots, 30$. Here m is the difference between the two iteration indices, not a bound on the iteration count. With $K_\infty \doteq (1 - \bar{z})/(1 - q)$ and $C_\infty^\sigma \doteq \mathrm{Re}(p_s(z)K_\infty)$, the identity

$$C_m^\sigma = C_\infty^\sigma + q^m\,\mathrm{Re}[p_s(z)(1 - K_\infty)] - \mathrm{Re}(p_s(z)z^m)$$

gives geometric bounds for every $m \ge 31$. The interval calculation shows that both limiting margins are strictly larger than the corresponding bounds on the remaining terms. Thus both inequalities hold for every $m \ge 1$. The positive rescaling in (30) gives an angular pair satisfying (10) at $q = q_\dagger$. Hence $q_\dagger \in \mathcal{Q}(s_\dagger)$, establishing the lower bound.

Ruling out larger asymptotic factors. For the upper bound, we need a positive weighted sum of cycle costs that is strictly negative at every allowed angle when q is just above $q_\dagger$. We construct this function from the two active costs C_5^σ and C_6^σ. Both vanish at the candidate; we choose their weight so that the derivative of the combination with respect to χ vanishes there too. We therefore take the weight

$$\lambda_\dagger \doteq \frac{\partial_\chi C_6^\sigma(q_\dagger, \chi_\dagger)}{\partial_\chi C_6^\sigma(q_\dagger, \chi_\dagger) - \partial_\chi C_5^\sigma(q_\dagger, \chi_\dagger)} = 0.41560302368050125 84\ldots \in (0,1)$$

and define the weighted sum, which in this case is a convex combination

$$\Phi_\dagger(q,\chi) \doteq \lambda_\dagger C_5^\sigma(q,\chi) + (1-\lambda_\dagger) C_6^\sigma(q,\chi). \tag{68}$$

By (66), this is equivalently

$$\Phi_\dagger(q,\chi) = \frac{(1+q-\chi)\big[\lambda_\dagger P_5(\alpha_\dagger, q, \chi) + (1-\lambda_\dagger) P_6(\alpha_\dagger, q, \chi)\big]}{2(s_\dagger^2-1)(q-\alpha_\dagger\chi+\alpha_\dagger^2)^2}. \tag{69}$$

Thus $\Phi_\dagger$ is an auxiliary rational function built from the two defining polynomials, with the same sign as their weighted combination on the allowed interval, which we use for the upper bound. The elimination below uses the equations $P_5 = P_6 = 0$ which only involve rational coefficients; note, we do not require the weight $\lambda_\dagger$ to be rational. We keep $\lambda_\dagger$ fixed as q varies. The algebraic definitions give $\Phi_\dagger = \partial_\chi \Phi_\dagger = 0$ at $(q_\dagger, \chi_\dagger)$. Outward-rounded interval arithmetic establishes $\partial_{\chi\chi}\Phi_\dagger < -100$ on a neighborhood of $\chi_\dagger$ at $q = q_\dagger$, and $\partial_q \Phi_\dagger(q_\dagger, \chi_\dagger) < -300$. The same method establishes a strict negative bound away from $\chi_\dagger$ using finitely many intervals with rational endpoints in $[-161/500, 161/500]$, where $2\sqrt{q_\dagger} < 161/500$. An exact rational check confirms that these intervals and the neighborhood used for the concavity bound together cover the entire interval; no part of the allowed frequency interval is omitted. Thus $\chi_\dagger$ is a zero-valued maximum of $\chi \mapsto \Phi_\dagger(q_\dagger, \chi)$. This is the contact with the horizontal axis; the maximum, rather than a minimum, reflects the sign of the cycle costs used here.

Local strict concavity and the negative partial derivative with respect to q, together with the strict negative bound away from $\chi_\dagger$ on a compact interval, make this same weighted sum strictly negative for every allowed rotation angle when $q > q_\dagger$ is sufficiently close to $q_\dagger$. Its coefficients are positive weights on finite cycle inequalities. The positive normalization factor transfers it to the costs of Lemma C.1. Proposition C.6 therefore gives $\rho_{\mathrm{hor}}(\kappa_\dagger) \le q$ for all such q. Letting $q \downarrow q_\dagger$ and using Theorem 4.1 gives $q_\star(s_\dagger) \le q_\dagger$. Feasibility supplies the reverse inequality. As before, we can ensure an initial segment that enforces $x_{-1} = x_0$ and realizes this factor in $\mathbf{R}^2$. □

Figure 6 plots the weighted sum (68), keeping both $q = q_\dagger$ and $\lambda_\dagger$ fixed. Its contact corresponds to the rotation angle of the single component.

G.2 The minimal polynomial and its Galois group

We now show why the exact rate $q_\dagger = q_\star(s_\dagger)$ derived above cannot be written using rational numbers, arithmetic operations, and finitely many nested nth roots. Recall the parameters of the preceding example:

$$s_\dagger = \frac{1421}{1250}, \qquad \kappa_\dagger = s_\dagger^2 = \frac{2019241}{1562500}, \qquad \alpha_\dagger = 1 - \frac{1}{s_\dagger} = \frac{171}{1421}.$$

Lemma G.1 identifies its exact squared-distance factor as $q_\star(s_\dagger) = q_\dagger = 0.0259177578383525961624\ldots$. We will give a polynomial equation specifying this number and then show why no radical expression can specify it. The second step is essential: degree alone does not rule out radicals, as the root $2^{1/20}$ of $q^{20} - 2$ illustrates. For background on the algebra used here, see Stewart (2022); precise versions of the two Galois-theoretic results invoked below are given in Milne (2022).

The polynomial. Define

$$F(q) \doteq \sum_{j=0}^{20} a_j q^j, \tag{70}$$

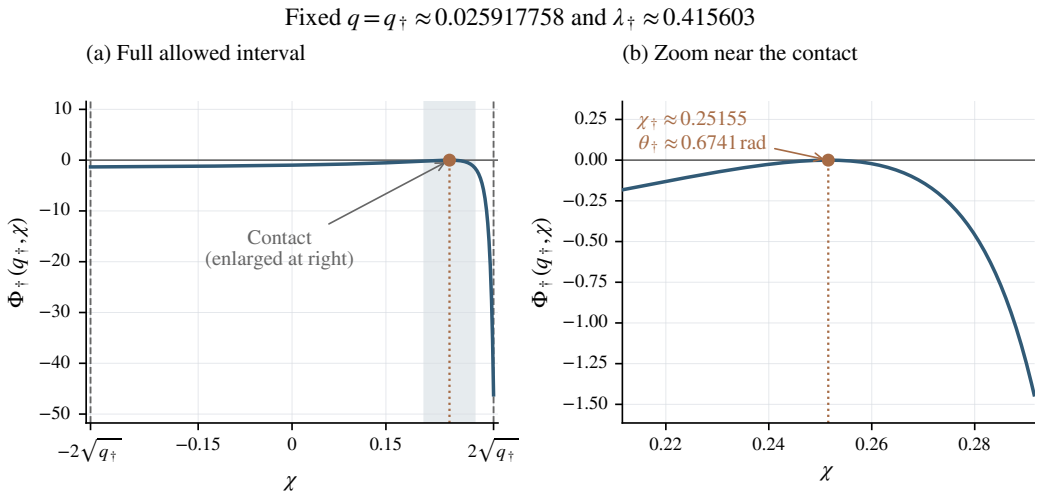


Figure 6: Numerical illustration of $\Phi_\dagger(q_\dagger, \chi)$ with q and the weight $\lambda_\dagger$ fixed. Left: the full allowed interval $[-2\sqrt{q_\dagger}, 2\sqrt{q_\dagger}]$. Right: enlargement of the shaded region, showing the zero-valued maximum at $\chi_\dagger$, labeled by the rotation angle $\theta_\dagger = \arccos(\chi_\dagger/(2\sqrt{q_\dagger}))$.

with all coefficients specified below. They are integers with no common factor greater than one, and $a_{20} > 0$. This normalization avoids fractional coefficients. Once irreducibility is proved, F/a_{20} is the *minimal polynomial*: the polynomial of least degree with rational coefficients that vanishes at $q_\dagger$, normalized to have leading coefficient one. This last normalization is called *monic*.

j	a_j	j	a_j
20	265994468953997804783976976	9	8189861320519662192293565908
19	807144202358229947702699124	8	634995018125129444105843789
18	1612480889879987742825308665	7	419734035354869237246982409
17	2946256217869261342444311626	6	236436111666155310881413139 8
16	4820345226837491495536334889	5	132922403764839996720208 1163
15	6720011736645743438628341186	4	525722768550046707732034423
14	8515544960311467977164916002	3	−279374126039493226882434 09
13	10093745217537977492178754770	2	401307772246057263077361
12	10921704964320530460147150207	1	−1444247861804047301319
11	10853408849668692979789731558	0	731086699811838561
10	97521735592181329786196456 98		

Proposition G.2 (A rate not expressible by radicals)**.** *The factor $q_\star(s_\dagger)$ is the unique real zero of F in the rational interval*

$$I \doteq (0.025917757838352596,\ 0.025917757838352597). \tag{71}$$

The polynomial F is irreducible over $\mathbb{Q}$ and its splitting-field Galois group is S_{20}. In particular, $q_\star(s_\dagger)$ has algebraic degree 20 and is not expressible by radicals over $\mathbb{Q}$.

Proof. Eliminating the angle. The preceding example is determined by two active cycle constraints, $P_5(\alpha_\dagger, q, \chi) = P_6(\alpha_\dagger, q, \chi) = 0$. Here $z = \sqrt{q}\, e^{i\theta}$, with $\theta \in [0, \pi]$, specifies the contracting and rotating component: $q = |z|^2$ is its squared contraction factor and $\chi = z + \bar{z} = 2\sqrt{q}\cos\theta$ encodes its rotation angle. The needed specialization of (63) is

$$T_0 = 2, \qquad T_1 = \chi, \qquad T_{j+1} = \chi T_j - q T_{j-1},$$
$$H_m = \sum_{j=0}^{m-1} q^j,$$
$$D_\dagger \doteq 2q^2 - 2\alpha_\dagger q\chi + \alpha_\dagger^2(\chi^2 - 2q),$$
$$E_\dagger \doteq (q^2 + \alpha_\dagger^2 q)\chi - 4\alpha_\dagger q^2,$$

and, for $m = 5, 6$,

$$P_m(\alpha_\dagger, q, \chi) = \frac{1}{\chi - 1 - q}\Big[(D_\dagger - E_\dagger)H_m + q^m D_\dagger - q^2 T_m + 2\alpha_\dagger q T_{m+1} - \alpha_\dagger^2 T_{m+2}\Big].$$

As before, the numerator is divisible by $\chi - 1 - q$, so these are polynomials. We eliminate χ from the two polynomial equations by taking their *resultant*. This is a polynomial in q alone that vanishes whenever the equations have a common χ solution. Thus it gives a necessary condition on q, algebraically projecting out χ. It may also vanish at values that do not correspond to an admissible trajectory, so we must identify the relevant factor. Exact elimination gives, up to a nonzero rational scalar,

$$\mathrm{Res}_\chi(P_5, P_6) = q^7(q+1)^2(q-\alpha_\dagger^2)^3(q^2-q+1)^2(q^2+q+1)^3 F(q),$$

where P_m here denotes $P_m(\alpha_\dagger, q, \chi)$. None of the factors preceding F vanishes in I, so the common solution from the preceding lemma satisfies $F(q_\dagger) = 0$. Sturm's rule, which counts real roots in a rational interval using exact polynomial arithmetic, gives one root in I and one in the smaller q interval isolating the common solution. The smaller interval lies in I. This identifies the specified root of F with $q_\star(s_\dagger)$.

Reading the modular factorizations. A polynomial is *irreducible over* $\mathbb{Q}$ if it cannot be factored into two positive-degree polynomials with rational coefficients. Rather than factor F over $\mathbb{Q}$ directly, we reduce its integer coefficients modulo a prime p and factor using arithmetic modulo p. Exact computation gives

p	degrees of the irreducible factors of $F \bmod p$
13	$1, 19$
107	$1, 1, 2, 3, 13$
223	20

.

The leading coefficients modulo these primes are 9, 102, and 124, respectively, so the degree does not drop. Each reduction is *squarefree*, meaning that no irreducible factor is repeated.

The *Galois group* $\mathfrak{G}$ consists of the permutations of the complex roots that preserve their algebraic relations over $\mathbb{Q}$. Dedekind's theorem turns each row of the table into information about such permutations: $\mathfrak{G}$ contains a permutation whose cycle lengths are precisely the listed factor degrees (Milne, 2022, Theorem 4.28).[1] Here cycles refer to permutations of polynomial roots, not to the iteration-index comparisons used to construct F. A cycle of length k sends each of k roots to the next and the last back to the first, fixing all other roots. The three rows give the following argument.

1. *Irreducibility and transitivity.* Modulo 223, there is just one irreducible factor, of degree 20. Because the integer coefficients of F have no common factor greater than one and its leading coefficient is nonzero modulo 223, this proves irreducibility over $\mathbb{Q}$. Consequently $\mathfrak{G}$ is *transitive*: any root can be sent to any other root by some permutation in $\mathfrak{G}$.

2. *Moving pairs of roots.* The row for $p = 13$ supplies a 19-cycle fixing the remaining root. Its powers can send any of the other nineteen roots to any other while keeping that root fixed. Transitivity gives the same property for any choice of fixed root. Thus $\mathfrak{G}$ is 2-*transitive*: it can send any ordered pair of distinct roots to any other such pair.

3. *Swapping two roots.* The row for $p = 107$ supplies disjoint cycles of lengths $1, 1, 2, 3, 13$. Applying that permutation $39 = 3 \cdot 13$ times makes the 3- and 13-cycles return to their starting points, while the 2-cycle still swaps its two roots. The result is a *transposition*, a permutation swapping just two roots. By 2-transitivity, relabelling this swap using elements of $\mathfrak{G}$ gives a swap of any chosen pair. Every permutation is a product of such swaps, so $\mathfrak{G} = S_{20}$, the full symmetric group on the twenty roots.

[1] The cited theorem is stated for monic integer polynomials. With $a \doteq a_{20}$, apply it to $a^{19}F(u/a)$, which is monic with integer coefficients and has the same splitting field as F. At the three primes, a is invertible, so the change of variable and the scalar factor preserve factor degrees and squarefreeness.

Excluding radicals. The needed consequence of the Galois criterion for solvability by radicals is this: an irreducible polynomial over $\mathbb{Q}$ whose Galois group is S_n, with $n \geq 5$, has no root expressible by radicals (Milne, 2022, Theorem 3.28 and its following discussion; proof in Theorem 5.34). Irreducibility matters here: a radical expression for one root would give such expressions for all the other roots as well. Since F is irreducible and its group is S_{20}, this excludes a radical expression for $q_\dagger$. Evaluating an everywhere-defined radical formula with rational coefficients at the rational input $s_\dagger$ would contradict this conclusion, proving Remark 4.2. This does not preclude specifying the rate as the unique root of a polynomial in an explicitly given interval, as we have done here. □

G.3 Robustness of the radical obstruction

Can additional constants help express the single value $q_\dagger$, or the rate as a function of s near $s_\dagger$? A radical expression permits field arithmetic and finitely many nested nth roots ($n \geq 2$), including complex intermediate values.

Proposition G.3 (Extensions at the certified parameter)**.** *The factor $q_\dagger$ is not expressible by radicals over $\mathbb{Q}(\tau)$ for any number τ transcendental over $\mathbb{Q}$, nor over any finite extension $E/\mathbb{Q}$ with $[E : \mathbb{Q}] < 20$. Adjoining the answer itself attains the bound: $\mathbb{Q}(q_\dagger)$ has degree 20.*

Proof. Let L be the splitting field of F, so $\mathrm{Gal}(L/\mathbb{Q}) = S_{20}$. A transcendental τ remains transcendental over L. Clearing denominators and comparing coefficients of powers of τ shows that a basis of L over $\mathbb{Q}$ stays linearly independent over $\mathbb{Q}(\tau)$; likewise for $\mathbb{Q}(q_\dagger)$. Thus adjoining τ preserves the irreducibility and group S_{20} of F, excluding radical expressions for its roots.

For a finite extension E, the splitting group over E identifies with $\mathfrak{G}_E = \mathrm{Gal}(L/(L\cap E))$, a subgroup of S_{20} of index $d = [L \cap E : \mathbb{Q}] < 20$. Acting on the cosets maps S_{20} to S_d. Since $20!/2 > d!$, the simple alternating group A_{20} of even permutations cannot inject into S_d. Its action is therefore trivial, so $\mathfrak{G}_E$ contains A_{20}. This makes $\mathfrak{G}_E$ transitive and nonsolvable: F stays irreducible over E, and the radical criterion again applies. □

We now seek a formula near $s_\dagger$, with constants fixed independently of s. These may be *arbitrary complex numbers*, even $q_\dagger$ itself.

Keeping α indeterminate in (63) gives a parameter-dependent version of F, the polynomial $H \in \mathbb{Z}[\alpha, q]$ defined by

$$\begin{aligned}\mathrm{Res}_\chi(P_5, P_6) = \; &-2\alpha^{12}q^7(q+1)^2(q-\alpha^2)^3 \times \\ &(q^2-q+1)^2(q^2+q+1)^3 H(\alpha, q).\end{aligned} \tag{72}$$

It has degree 20 in q, degree 8 in α, leading q coefficient 16, and $H(\alpha_\dagger, q) = 16F(q)/a_{20}$.

Proposition G.4 (No local radical formula with arbitrary fixed constants)**.** *There is an open interval $J \subset (1, \infty)$ containing $s_\dagger$ on which $q_\star(s)$ is a real-analytic root of $H(1 - 1/s, q)$. The splitting-field Galois group of H over $\mathbb{C}(\alpha)$ is S_{20}. Consequently no formula built from rational functions of s with arbitrary fixed complex coefficients and finitely many radicals agrees with $q_\star(s)$ on any nonempty open subinterval of J.*

Proof. Persistence of the optimal branch. At the isolated common zero, outward-rounded evaluation gives

$$4.5 \cdot 10^{-8} < \det \frac{\partial(P_5, P_6)}{\partial(q, \chi)}(\alpha_\dagger, q_\dagger, \chi_\dagger) < 4.6 \cdot 10^{-8}.$$

The analytic implicit-function theorem continues (q, χ) locally in s, preserving $q > \alpha_s^2$ and $\chi^2 < 4q$. The three active comparisons remain equalities. All other finite inequalities in Lemma G.1 have strict margins; its geometric tail bounds remain uniform on a small compact neighborhood with $q < 1$. Thus interpolation feasibility persists. The dual weight remains in $(0, 1)$; strict concavity near contact, the negative test-q derivative, and uniform negativity away from contact also persist. For each continued pair, hold its dual weight fixed as the test q varies. The same lower and strict-separator upper arguments identify the continued factor with $q_\star(s)$. Shrink J to exclude the other resultant factors.

Galois group over $\mathbb{Q}(\alpha)$. At $\alpha_\dagger$, the roots are distinct and their group is S_{20}. Each root lifts uniquely to a formal power-series root in $\alpha - \alpha_\dagger$, with coefficients in the splitting field L of F. Acting coefficientwise, $\mathrm{Gal}(L/\mathbb{Q})$ permutes these lifts faithfully because their constant terms are distinct. Thus the group over $\mathbb{Q}(\alpha)$ contains, and hence equals, S_{20}.

Allowing arbitrary complex constants. The discriminant $\mathcal{D}(\alpha) \doteq \mathrm{Disc}_q H(\alpha, q) \in \mathbb{Z}[\alpha]$ detects collisions of roots in q. Exact computation gives

$$\mathcal{D}(8) \equiv 0, \qquad \mathcal{D}'(8) \equiv 25 \pmod{29}. \tag{73}$$

These congruences imply that $\mathcal{D}$ has a simple complex zero: if every characteristic-zero irreducible factor were repeated, reduction modulo 29 could not produce a simple root. Since the leading q coefficient is the nonzero constant 16, a simple discriminant zero is a simple collision of two roots; continuation around it exchanges those roots and fixes the others. This transposition lies in the subgroup fixing the algebraic constant field, called the geometric group. It is normal in S_{20}, so contains all conjugate transpositions and equals S_{20}. Extending the algebraically closed constants from $\overline{\mathbb{Q}}$ to $\mathbb{C}$ preserves this group.

Hence H is irreducible over $\mathbb{C}(\alpha)$ with nonsolvable group. A radical expression agreeing with a root on an open interval would give a radical algebraic branch over that field, contradicting the radical criterion. The rational change $\alpha = 1 - 1/s$ proves the claim. □

Remark G.5 (Adjoining the contact coordinate). The contact coordinate and the rate generate the same degree-20 field: $\mathbb{Q}(\chi_\dagger) = \mathbb{Q}(q_\dagger)$. Exact elimination of the specialized P_5, P_6 gives rational-coefficient relations $U(q)\chi + V(q) = 0$ and $\widetilde{U}(\chi)q + \widetilde{V}(\chi) = 0$. Exact gcd checks exclude zeros of both leading coefficients at the isolated pair, so each coordinate is recovered rationally from the other. These obstructions do not exclude isolated-root descriptions or other nonradical notions of closed form.

H Rate values and their numerical evaluation

While expressing the optimal contraction factor as a formula with radicals might not be possible in general as we have seen in Section G, we can still use our characterization to compute contraction values for varying values of κ in Table 2. To this end, we evaluate $q_\star(\sqrt{\kappa})$ for selected condition numbers and compare it with Gannot's upper bound denoted by $q_{\mathrm{G}}(\sqrt{\kappa})$. The values are obtained by optimizing over mode weights and angles in the equivalent angular formulation. As can be seen, Gannot's upper bound is quite tight with largest gaps for moderate κ.

κ	$q_\star$	q_{G}	$q_{\mathrm{G}} - q_\star$	status
1	0	–	–	exact
1.001	$4.99001617487858 \times 10^{-7}$	$4.99001622504 \times 10^{-7}$	5.01595×10^{-15}	numerical
1.01	$4.90160022279601 \times 10^{-5}$	$4.90160037565 \times 10^{-5}$	1.52852×10^{-12}	numerical
1.1	0.00414080335569045	0.00414080555192	2.19623×10^{-9}	numerical
2	0.133970856851373	0.133998720281	2.78634×10^{-5}	numerical
5	0.400092644577688	0.40047760383	3.84959×10^{-4}	numerical
10	0.564423277972893	0.565235163135	8.11885×10^{-4}	exact CA
20	0.688283179210141	0.689465036059	0.00118186	numerical
50	0.801693369348438	0.803120025898	0.00142666	numerical
100	0.859899778052548	0.861059858423	0.00116008	exact CA
200	0.901252726206213	0.902086880416	8.34154×10^{-4}	numerical
500	0.937861932828071	0.938397069587	5.35137×10^{-4}	numerical
1000	0.95626894031021	0.956612260497	3.4332×10^{-4}	numerical

Table 2: Selected squared-distance optimal factor $q_\star$ and the Gannot upper bound q_{G}. For $\kappa = 1$ the values are analytically exact (and trivial). *Exact CA* (exact computer-assisted) refers to separately established algebraic values with matching lower and upper rate proofs via Appendices E and F; the values in the table are rounded. Other rate entries and all gaps are numerical approximations via the procedure described in the section.

Obtaining numerical rates. Basically the numerical rates are obtained by means of bisection. More precisely, for a given $\kappa > 1$, we range over test squared factors q. For each such q we want to assess numerically whether it is above or below $q_\star(\sqrt{\kappa})$ and to this end we choose a finite grid of angles in $[0, \pi]$. The unknowns are nonnegative mode weights summing to one and a nonnegative common potential value. We impose the two interpolation inequalities (10) for $t = 1, \dots, M$, using the equivalent transformed-gradient normalization of Appendix C.3. With q and the angles fixed, these constraints are linear in the unknowns. The linear program maximizes a common slack in the inequalities: a nonnegative optimal slack indicates feasibility of this finite system, subject to numerical accuracy. If we considered all angles and all interpolation inequalities and solved the system exactly, then, within our search range, infeasibility would imply $q > q_\star(\sqrt{\kappa})$. We are, however, dealing with three approximations: floating-point arithmetic, a finite angular grid, and a finite subset of the interpolation inequalities. We assess numerical accuracy through angular refinement and checks at additional iteration indices, nonetheless we stress that the obtained rates are only approximations.

The bisection in q starts with the initial interval $[(1 - 1/\sqrt{\kappa})^2, q_{\mathrm{G}}]$, whose endpoints are the quadratic factor and Gannot's upper bound q_{G} defined below. If the computed optimal slack is negative, we take q as the new upper endpoint of the numerical search; otherwise, we take it as the new lower endpoint. These updates do not constitute certified bounds on $q_\star(\sqrt{\kappa})$. We first solve on a grid covering the angular interval, with additional resolution near zero. We then add grid points around the angles receiving positive weight and repeat the bisection on this finer grid. This refines the locations of the modes further. For each condition number, the index limit for the LP is fixed at

$$M = \max\{80, \lceil 8\sqrt{\kappa} \rceil\}.$$

After the angular refinement, we check the recovered weights and potential against the two interpolation inequalities for $t = 1, \dots, K$, where

$$K = \max\{2000, \lceil 40\sqrt{\kappa} \rceil\}.$$

Thus the returned measure is tested against a large number of constraints that were not used to construct it. We stress that these are only finite, floating-point checks, i.e., an approximate test which does not prove feasibility for every integer $t \geq 1$. The exact characterization of Theorem 4.1 would require using all iteration indices.

Exact values. While the aforementioned numerical computation is only approximate, we can also do exact calculations for specific values. To illustrate this, we did this for $\kappa = 10$ and $\kappa = 100$ in Appendices E and F and report these values in Table 2 instead of the numerical approximations via the procedure described above. These algebraic solutions and matching lower and upper rate bounds are obtained from algebraic computations, exact identities and interval arithmetic. The infinitely many interpolation inequalities are handled by finite checks together with analytic bounds for the remaining iteration indices. At $\kappa = 1$, the rate is zero because the first gradient step reaches the minimizer.

Computing the comparison bound. For $\kappa > 1$, Gannot's upper bound is the smaller positive root r_- of

$$\begin{aligned} h_\kappa(r) = 8(1 - \sqrt{\kappa})^4 - (1 - \sqrt{\kappa})^2(15\kappa - 10\sqrt{\kappa} - 1)r \\ + 2\kappa(3\kappa - 4\sqrt{\kappa} + 1)r^2 + \kappa^2 r^3, \end{aligned}$$

as stated in Gannot (2022, Proposition 2.13). That proposition bounds the norm factor by $\sqrt{r_-}$, so the corresponding squared-distance bound is $q_{\mathrm{G}} = r_-$. We evaluate this cubic for the table and also use it to provide the upper endpoint of the initial LP search interval. The dash at $\kappa = 1$ in Table 2 records the proposition's assumption $\kappa > 1$; the plotted comparison curve in Figure 1 uses the continuous limit zero.