ELLIPTIC SAMPLING: OPTIMAL INTERPOLANT SCHEDULES FOR GAUSSIANS AND BEYOND

Anonymous authors
Paper under double-blind review

ABSTRACT

Generative modeling based on stochastic interpolants requires specifying two key choices: an *interpolant schedule* bridging the initial and target distributions, and an efficient *discretization* of the resulting probability-flow ODE. In this work, we introduce a new framework for joint schedule/discretization design based on the analyticity properties of the (complexified) ODE trajectories, which in turn ensure good numerical behavior. Our formulation can be solved exactly via conformal mapping, yielding a natural family of variance-preserving *elliptic schedules*, parametrized by Jacobi elliptic functions. Pairing an elliptic schedule with a tailored discretization, we obtain *Zoloflow*, a new sampling algorithm closely connected to Zolotarev's optimal rational function approximants. Crucially, both the elliptic schedule and Zoloflow explicitly incorporate the condition number κ of the target distribution and reduce to the classical trig schedule in the limit $\kappa \to 1$. For Gaussian targets, Zoloflow achieves *minimax optimality among all oblivious score-span algorithms*, attaining total-variation error ε using $\Theta(\log \kappa \, \log(\sqrt{d}/\varepsilon))$ (exact) score queries. For the log-concave case, the elliptic schedule allows sampling with at most ε 2-Wasserstein-error by requiring merely $\widetilde{O}(\sqrt{d}\,\mathrm{polylog}(\kappa, 1/\varepsilon))$ score evaluations. Complementing these theoretical results, we empirically evaluate the elliptic schedule on general (non-log-concave) target distributions, including detailed comparisons with alternative schedule proposals on both synthetic benchmarks and existing image generation datasets.

1 INTRODUCTION

Generative modeling seeks to produce novel random samples $X \sim \pi$ from a complicated target distribution π on $\mathbb{R}^d$ using a representation learned from a finite training dataset $\mathcal{D}_{\mathrm{train}} = \{X^{(i)}\}$ of iid samples $X^{(i)} \sim \pi$. Many modern approaches learn an ordinary or stochastic differential equation (ODE or SDE) which progressively transports a simple initial reference distribution, typically Gaussian, to the target π. Stochastic interpolant (SI) framework (Albergo et al., 2025) unifies these flow-based (ODE) and diffusion-based (SDE) approaches and offers flexibility in shaping the transport dynamics through *interpolant schedules*.

Although different valid schedules all bridge the same initial and target distributions under exact continuous-time transport and are related through time and space reparameterizations, they can induce trajectories with substantially different regularity properties, including Lipschitz continuity and complex analyticity. These properties, in turn, govern how accurately and efficiently the resulting dynamics, most notably the probability-flow ODE (PF-ODE), can be solved numerically. Sampling performance, therefore, depends not only on the choice of interpolant schedule, but also on the *discretization scheme* used to integrate the induced dynamics with a finite number evaluations of the learned score/velocity field, typically represented by a deep neural network. Despite this intrinsic coupling, interpolant schedules and the discretization schemes are usually designed separately and chosen largely heuristically. This raises the following central question:

Given a fixed budget of evaluations of the learned score/velocity field, how should one choose both the interpolant schedule and the numerical procedure used to integrate the resulting PF-ODE?

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.

1.1 CONTRIBUTIONS

We address this question by developing a principled theory of joint schedule and discretization design, which seeks the best achievable sampling accuracy over a class of target distributions under a finite budget of score/velocity-field evaluations. Our contributions are as follows:

1. Minimax-optimal Gaussian Sampling with Zoloflow. We formulate sampling from a Gaussian target with condition number at most κ as a minimax problem over oblivious score-span algorithms with a fixed budget of exact score evaluations, and show that it reduces to a scalar rational approximation problem solved by Zolotarev's optimal approximants. The resulting algorithm attains d_{TV} error ε with $\Theta(\log\kappa\,\log(\sqrt{d}/\varepsilon))$ score evaluations, which is minimax optimal. It admits a dynamical realization, *Zoloflow*, which pairs a variance-preserving schedule parametrized by Jacobi elliptic functions with a modified Euler-type discretization; both depend on κ and reduce to the trig schedule as $\kappa \to 1$.

2. Complex-Analytic Optimality of Elliptic Schedules. We propose a complementary design criterion: the width of the complex strip around real time to which PF-ODE trajectories extend holomorphically, which governs the convergence of high-order numerical integrators. For Gaussian targets, maximizing this width is solved exactly by conformal mapping and recovers the elliptic schedule of the Zolotarev construction. For strongly log-concave targets, we establish explicit lower bounds on the width attained by the elliptic schedule.

3. Efficient Log-Concave Sampling with Elliptic Schedules. Combining these width bounds with collocation methods for analytic ODEs, we obtain a sampler for strongly log-concave targets that attains W_2 error ε with $O(\sqrt{d}\,\log^{3.5}(\kappa d/\varepsilon))$ score evaluations. Because it relies on a generic high-order solver rather than the Gaussian-specific Zolotarev structure, this result isolates the benefit of the elliptic schedule itself.

4. Empirical Evaluations of Schedule and Discretization. On Gaussian targets, we empirically confirm the gains of Zoloflow over existing discretizations predicted by the minimax theory. Beyond Gaussians, we pair the elliptic schedule with state-of-the-art numerical solvers, and compare it against alternative schedules at matched evaluation budgets on synthetic and image-generation benchmarks. Across solvers, the elliptic schedule outperforms alternative schedules, including on non-log-concave targets beyond the scope of our theory.

Throughout the rest of the paper, we assume the score/velocity-field function evaluations are exact. All proofs and additional results are deferred to the Appendix.

1.2 PRIOR WORK

Interpolant schedules are typically chosen by hand, as with the trig schedule (Nichol & Dhariwal, 2021) and EDM (Karras et al., 2022), or optimized for a fixed solver (Sabour et al., 2024) or for drift regularity (Chen et al., 2025). Fast solvers such as DDIM (Song et al., 2022) and DPM-Solver++ (Lu et al., 2022b) instead improve the discretization for a fixed schedule. We design the two jointly and choose the schedule by the analyticity of the flow in complex time. For Gaussian targets, the closest work is Liu (2026), which combines smoothed-score queries through a sinc-quadrature rational approximation of $x^{-1/2}$. Zoloflow instead realizes the optimal Zolotarev approximant, which improves the query complexity and matches a lower bound for oblivious score-span algorithms. For log-concave targets, we build on collocation methods for analytic ODEs (Lee et al., 2018; Gatmiry et al., 2026), with the elliptic schedule controlling the domain of analyticity. Section L gives an extended discussion.

2 PRELIMINARIES

2.1 STOCHASTIC INTERPOLANTS

Stochastic interpolant provide a general framework for constructing a continuous path between a simple reference distribution and a target (Albergo et al., 2025), and defined as the stochastic process

$$x_t = \alpha_t z + \beta_t x_1, \qquad t \in [0,1], \tag{1}$$

where $z \sim \mathcal{N}(0, I_d)$ is the reference distribution, and $x \sim \pi$ is the target, and (α_t, β_t) is the *interpolant schedule*. Stochastic interpolants induce a deterministic probability-flow ODE with drift (velocity field) given by

$$b_t(x) \triangleq \mathbb{E}[\dot{x}_t \mid x_t = x] \tag{2}$$

so that, under standard regularity conditions, the solution of

$$\dot{X}_t = b_t(X_t), \qquad X_0 \sim \mathcal{N}(0, I_d), \tag{3}$$

has marginal law $X_t \sim p_t$ where $p_t = \mathcal{L}(x_t)$ is the law of the interpolant at time t. For the linear interpolant (1), the velocity field can be expressed using the score function or the conditional mean of the target sample. Indeed, using Tweedie's formula $\beta_t \mathbb{E}[x_1 \mid x_t = x] = x + \alpha_t^2 \nabla \log p_t(x)$,

$$b_t(x) = \frac{\dot{\alpha}_t}{\alpha_t} x + \beta_t \left(\frac{\dot{\beta}_t}{\beta_t} - \frac{\dot{\alpha}_t}{\alpha_t} \right) \mathbb{E}[x_1 \mid x_t = x] = \frac{\dot{\beta}_t}{\beta_t} x + \alpha_t^2 \left(\frac{\dot{\beta}_t}{\beta_t} - \frac{\dot{\alpha}_t}{\alpha_t} \right) s_t(x). \tag{4}$$

Definition 2.1 (VP schedules)**.** A schedule (α, β) is *variance preserving* (VP) if $\alpha_t^2 + \beta_t^2 = 1$ on $[0, 1]$, with α strictly decreasing from 1 to 0 and β strictly increasing from 0 to 1. Equivalently, $(\alpha_t, \beta_t) = (\cos \varphi(t), \sin \varphi(t))$ for a unique *phase* $\varphi = \arccos \alpha$ in

$$\mathcal{C} \triangleq \big\{ \varphi : [0,1] \to [0, \tfrac{\pi}{2}] \text{ strictly increasing} : \ \varphi(0) = 0, \ \varphi(1) = \tfrac{\pi}{2} \big\}.$$

Denote $\mathcal{S}$ as the class of VP schedules. Choosing a schedule in $\mathcal{S}$ amounts to choosing a phase in $\mathcal{C}$. For VP schedules, $\alpha_t \dot{\alpha}_t + \beta_t \dot{\beta}_t = 0$, so equation 4 simplifies to

$$b_t(x) = \frac{\dot{\beta}_t}{\beta_t} \big(x + s_t(x) \big). \tag{5}$$

2.2 Problem Setup

In this paper we study the following algorithm design problem. Given a class A_r of sampling algorithms that use at most r score evaluations and a class P of target distributions, we seek the algorithm whose output law $\hat{\pi}_{\mathrm{alg}}(\pi)$ is uniformly closest to the target over the class,

$$\inf_{\mathrm{alg} \in \mathsf{A}_r} \sup_{\pi \in \mathsf{P}} \mathrm{dist}\big(\hat{\pi}_{\mathrm{alg}}(\pi), \pi \big), \tag{6}$$

for a distance dist between probability measures. Throughout, we take:

1. Algorithms. A_r consists of all pairings of (i) a stochastic interpolant schedule (α_t, β_t) with (ii) a numerical discretization of the probability-flow ODE equation 3 (e.g. Euler) that makes r evaluations of the exact score $s_t = \nabla \log p_t$ of the corresponding interpolant. The schedule and discretization are the designed variables, and may depend on the class parameter κ below but not on the particular target $\pi \in \mathsf{P}$.
2. Targets. P is one of two nested classes of centered distributions with condition number at most $\kappa \geq 1$: Gaussians with precision spectrum in $[1, \kappa]$, and strongly log-concave distributions with curvature in $[1, \kappa]$,

$$\mathsf{G}_\kappa = \big\{ \mathcal{N}(0, \Lambda^{-1}) : I_d \preceq \Lambda \preceq \kappa I_d \big\} \ \subset \ \mathsf{P}_\kappa = \big\{ \pi \propto e^{-V} \text{ centered} : I_d \preceq \nabla^2 V \preceq \kappa I_d \big\}.$$

In Section 3 we show that for $\mathsf{P} = \mathsf{G}_\kappa$ and $\mathrm{dist} = d_{\mathrm{TV}}$, we can solve equation 6 exactly. In this regime, every algorithm in A_r outputs $q(\Lambda) z$ for a rational function q of degree r, so equation 6 collapses to a scalar problem, the best rational approximation of $\lambda^{-1/2}$ on $[1, \kappa]$ and can solved in closed form. The solution *Zoloflow*, pairs a schedule built from elliptic functions and a custom finite-difference discretization attaining the optimal $\Theta(\log \kappa \, \log(\sqrt{d}/\varepsilon))$ sample complexity. For $\mathsf{P} = \mathsf{P}_\kappa$ in Section 4 we show that, combined with collocation methods for analytic ODEs, this elliptic schedule samples with W_2 error ε using $O\big(\sqrt{d}\, \mathrm{polylog}(\kappa, d, 1/\varepsilon)\big)$ score evaluations.

For simplicity, we assume all target distributions are zero-mean, without of generality.

Algorithm 1 Zoloflow - sample from a general distribution from score queries

Require: Step count r, parameter κ ("condition number")
1: Initial sample $x_0 \sim \mathcal{N}(0, I_d)$
2: Set step size $h = 2/(2r+1)$ and $t_j = (2j-1)/(2r+1)$ for $j = 1, \dots, r$
3: Set parameter $m = 1 - \kappa^{-1}$, $K = K(m)$, $\alpha(t) = \mathrm{cn}(Kt \mid m)$, and $\beta(t) = \mathrm{sn}(Kt \mid m)$
4: **for** $j = 1, \dots, r$ **do**
5: Query score: $s_{t_j}(x_{j-1}) = \nabla \log p_{t_j}(x_{j-1})$
6: Update: $x_j \leftarrow x_{j-1} + \dfrac{\beta^2(t_j + h/2) - \beta^2(t_j)}{\beta^2(t_j)} \left(x_{j-1} + s_{t_j}(x_{j-1})\right)$
7: **end for**
8: **return** x_r

3 Gaussian Sampling: The Elliptic Schedule and Zoloflow

3.1 Elliptic Stochastic Interpolant

Motivation As discussed above, a VP schedule is *equivalent* to a phase function $\varphi \in \mathcal{C}$. Our primary criterion for choosing φ will concern the analyticity properties it induces in the ODE solution X_t, when extended to the complex plane.

For a Gaussian target $\mathcal{N}(0, \Lambda^{-1})$, the ODE (3) can be solved as $X_t = (\alpha_t^2 I_d + \beta_t^2 \Lambda^{-1})^{\frac{1}{2}} X_0$, and thus the eigenmodes are given by

$$x_\lambda(t) = \sqrt{\alpha(t)^2 + \lambda^{-1}\beta(t)^2}\, x_\lambda(0).$$

This has branch point singularities whenever the expression inside the square root vanishes, i.e., whenever $\cos(\varphi(t))^2 + \lambda^{-1}\sin(\varphi(t))^2 = 1 - (1-\lambda^{-1})\sin^2\varphi(t) = 0$. Thus, our design problem is to find φ such that the roots t of this equation are as far as possible from the real axis, whenever $\lambda \in [1, \kappa]$. This ensures that the ODE solution X_t is analytic on a wide strip. The optimal such φ can be nicely characterized as follows:

Proposition 3.1 (Jacobi amplitude as optimal phase)**.** *Consider the phase (angular clock) given by*

$$\varphi_\star(t) \triangleq \mathrm{am}(Kt \mid m), \tag{7}$$

where $m = 1 - \kappa^{-1}$, $K = K(m)$, $K' = K(1-m)$, and $\mathrm{am}()$ is the Jacobi amplitude function and $K()$ the complete elliptic integral of first kind (see Appendix C). Let $S_\gamma := \{t \in \mathbb{C} : |\operatorname{Im} t| < \gamma\}$ be a strip around the real axis of width 2γ. This choice of φ maximizes γ (analyticity strip width), subject to the constraints i) φ is holomorphic on S_γ, ii) $\varphi(0) = 0$, $\varphi(1) = \pi/2$, iii) $1 - (1 - \lambda^{-1})\sin^2\varphi(t) \neq 0$, $\forall t \in S_\gamma$, $\forall \lambda \in [1, \kappa]$. In particular, for this choice of $\varphi_\star$ we have $\gamma_\star = K'/K$.

A proof is provided in Appendix I.3.

The optimality of the Jacobi amplitude leads directly to the following definition:

Definition 3.2 (Elliptic schedule)**.** For a given $\kappa \geq 1$, let

$$m = 1 - \kappa^{-1}, \qquad K = K(m), \qquad K' = K(1-m).$$

The *elliptic stochastic interpolant schedule* is

$$\alpha_t = \mathrm{cn}(Kt \mid m), \qquad \beta_t = \mathrm{sn}(Kt \mid m), \qquad t \in [0, 1], \tag{8}$$

where cn and sn are Jacobi elliptic functions.

The elliptic schedule is variance-preserving, and satisfies all properties in Definition 2.1. Furthermore, both α_t and β_t are holomorphic on the complex strip $|\operatorname{Im} t| < K'/K$. Notice that the elliptic schedule depends explicitly on the condition number κ. Also, as $\kappa \to 1$, this reduces to trigonometric schedule from Albergo et al. (2025).

3.2 ZOLOFLOW

The elliptic schedule given in Definition 3.2 can in principle be paired with any numerical method (e.g., Euler, Heun, etc.) to solve the probability flow ODE (3). However, if instead we use a non-standard (but natural) discretization we obtain *Zoloflow*, a sampling algorithm that is *essentially optimal on Gaussians*.

Definition 3.3 (Elliptic finite-difference discretization)**.** Fix a step count r and an interpolant schedule (α, β). The *elliptic finite-difference* update is

$$x_j = x_{j-1} + h\left[\frac{\beta_{t_j+h/2}^2 - \beta_{t_j}^2}{h\,\beta_{t_j}^2}\right]\left(x_{j-1} + s_{t_j}(x_{j-1})\right), \qquad j = 1, \dots, r. \tag{9}$$

on the grid $h = \frac{2}{2r+1}$, $t_j \in \left\{\frac{1}{2r+1}, \frac{3}{2r+1}, \dots, \frac{2r-1}{2r+1}\right\}$.

A derivation of this approach, and a comparison with standard Euler and deterministic DDIM discretizations using the same iterate and time-grid notation are presented in Appendix J.

Definition 3.4 (Zoloflow)**.** Fix $\kappa \geq 1$ and $r \in \mathbb{N}$, a target distribution π with $x_1 \sim \pi$, and set $m \triangleq 1 - \kappa^{-1}$. Zoloflow is the r-step algorithm obtained by applying the elliptic finite-difference discretization (Def. 3.3) to the flow ODE (3), using the elliptic schedule (α_t, β_t) given in Def. 3.2. We use a step size $h = 2/(2r+1)$ and time grid $t_j = (2j-1)/(2r+1)$, $j = 1, \dots, r$. A formal description is presented as Algorithm 1 above.

Zoloflow can be interpreted as the optimal Zolotarev approximant $\widehat{q}_{\kappa,r}(\lambda)$ to $\lambda^{-1/2}$ given in (55), which has been implemented through an elliptic interpolant/discretization pair; see Lemma E.1. Although primarily motivated by Gaussian targets (where they are optimal), both the elliptic schedule and Zoloflow can be used to sample from *any* distribution. As we show in later sections, they will perform quite well from both theoretical (in the Gaussian/log-concave case) and empirical (for general distributions) viewpoints, often outperforming competing techniques.

3.3 SCORE-ORACLE COMPLEXITY OF GAUSSIAN SAMPLING

Gaussian Sampling as Rational Function Approximation. Before getting into the explicit complexity results for Zoloflow, we describe the core intuition behind our results. Let's introduce a class of algorithms which serve as a proxy for different discretizations of the probability-flow ODE (3).

Definition 3.5 (Oblivious score-span algorithms $\mathcal{A}_r$)**.** We say an algorithm **alg** belongs to an oblivious (non-adaptive) score span class $\mathcal{A}_r$ of degree r if its iterates satisfy

$$x_j \in \text{span}\left\{x_0, \dots, x_{j-1}, s_{t_0}, \dots s_{t_{j-1}}(x_{t_{j-1}})\right\},\ j = 1, 2, \dots r, \qquad x_0 = z \sim \mathcal{N}(0, I_d) \tag{10}$$

for pre-determined times $t_0, \dots, t_{r-1} \in (0, 1)$ and predetermined scalar coefficients in each linear combination. Equivalently, **alg** $\in \mathcal{A}_r$ if for predetermined $(a_{ij})_{0\leq i\leq j<r}$, $(b_{ij})_{0\leq i\leq j<r}, (t_j)_{0\leq j<r}$ the updates can be written as

$$x_j = \sum_{i=0}^{j-1} a_{ij} x_i + \sum_{i=0}^{j-1} b_{ij} s_{t_i}(x_i), \qquad j = 1, \dots, r \tag{11}$$

where $s_t(x) = \nabla \log p_t(x)$ depends on the interpolant schedule (α_t, β_t) and the target distribution $\pi = \mathcal{L}(x_1)$. Given a fixed schedule (α_t, β_t), a target distribution π, and an instance **alg** $\in \mathcal{A}_r$, we denote the distribution of the final iterate by $\mathcal{L}(x_r^{\textbf{alg}})$.

Having this definition in mind, the key point to make is that every algorithm **alg** $\in \mathcal{A}_r$ has a rational function representation. We formalize what we mean by this in the following lemma.

Lemma 3.6 (All **algs** $\in \mathcal{A}_r$ are rational functions on Gaussians)**.** *Suppose the target distribution $\pi = \mathcal{N}(0, \Lambda^{-1})$ is zero-mean Gaussian, with precision matrix $\Lambda = U\text{diag}(\lambda_1, \dots, \lambda_d)U^T$. For any $a, b \in \mathbb{N}$ define*

$$\mathcal{R}_{a,b} \triangleq \left\{ r(x) = \frac{p(x)}{q(x)} \;\middle|\; p, q \in \mathbb{R}[x],\ \deg p \leq a,\ \deg q \leq b,\ q \not\equiv 0 \right\}. \tag{12}$$

as the class of all rational functions whose numerator is a polynomial degree at most a and denominator a polynomial with degree at most b. Then for any algorithm in $\mathcal{A}_r$ we can express its final iterate as

$$x_r^{\textbf{alg}} = q_r^{\textbf{alg}}(\Lambda)z = U\mathrm{diag}(q_r^{\textbf{alg}}(\lambda_1), \dots, q_r^{\textbf{alg}}(\lambda_d))U^T z \tag{13}$$

where $q_r^{\textbf{alg}}(\lambda) \in \mathcal{R}_{r,r}$ is a rational function that can be obtained from the algorithm coefficients $((\alpha_{t_j}), (\beta_{t_j}), (a_{ij}), (b_{ij}))$. We delay a proof of this lemma to Appendix F

The key idea is that for Gaussian distributions, all algorithms in $\mathcal{A}_r$ (and more specifically every interpolant and discretization pair) admit a rational function $q(\cdot)$ which acts on the eigenvalues of the precision matrix Λ. In this way, the performance of any such algorithm is nothing more than a question about the quality of the approximation of $\lambda^{-1/2}$. In particular

$$q(\Lambda) \approx \Lambda^{-1/2} \implies q(\Lambda)z \approx \Lambda^{-1/2}z \overset{d}{=} \mathcal{N}(0, \Lambda^{-1}) \tag{14}$$

In this way, one can understand Zoloflow as the optimal rational function approximation to $\lambda^{-1/2}$, implemented through an elliptic interpolant/discretization pair.

To be more precise, let $x_r^{\mathrm{alg},\Lambda}$ denote the output of $\mathrm{alg} \in \mathcal{A}_r$ when the target is $\mathcal{N}(0, \Lambda^{-1})$ and the algorithm is given exact interpolant scores. Define the minimax query complexity

$$r^*_{\mathrm{obl}}(d, \kappa, \varepsilon) \triangleq \min\left\{ r \in \mathbb{Z}_{\geq 1} : \begin{array}{c} \exists\, \mathrm{alg} \in \mathcal{A}_r \text{ such that} \\ \sup\limits_{I_d \preceq \Lambda \preceq \kappa I_d} d_{\mathrm{TV}}\big(\mathcal{L}(x_r^{\mathrm{alg},\Lambda}), \mathcal{N}(0, \Lambda^{-1})\big) \leq \varepsilon \end{array} \right\}. \tag{15}$$

Here each algorithm, including its admissible interpolant schedule, query times, and scalar update coefficients, is chosen independently of Λ, the initialization, and the oracle responses.

Theorem 3.7 (Zoloflow is minimax optimal)**.** *Fix $\kappa > 1$, $d \geq 1$, and a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $I_d \preceq \Lambda \preceq \kappa I_d$. Run Zoloflow with parameters (κ, r), exact interpolant scores, and initialization $x_0 = z \sim \mathcal{N}(0, I_d)$, and let $\widehat{\pi} \triangleq \mathcal{L}(x_r)$. Then for any fixed $\varepsilon \in (0, \varepsilon_0)$ for a universal constant ε_0, Zoloflow attains the minimax-optimal oracle complexity with*

$$r^*_{\mathrm{obl}}(d, \kappa, \varepsilon) = \Theta(\log \kappa \, \log(\sqrt{d}/\varepsilon))$$

score queries, guaranteeing $d_{\mathrm{TV}}(\hat{\pi}, \pi) \leq \varepsilon$.

4 Log-Concave Sampling

We now consider the class P_κ of centered strongly log-concave targets (Section 2.2). Since the score is nonlinear, the rational-approximation reduction of Section 3.3 no longer applies. We instead build on the complex-analytic criterion of Proposition 3.1: after a mild Gaussian smoothing, the elliptic schedule keeps all PF-ODE trajectories holomorphic and bounded on a strip around $\mathbb{R}$ whose width degrades only polylogarithmically in κ and $1/\varepsilon$ (Theorem 4.1). Collocation methods for analytic ODEs (Lee et al., 2018) then yield W_2-error ε with $O(\sqrt{d}\,\log^{3.5}(\kappa d/\varepsilon))$ score evaluations (Theorem 4.2).

4.1 Holomorphic Width under the Elliptic Schedule

Holomorphic width. For an ODE $\dot{X}_t = b_t(X_t)$ with $X_0 = z$, a radius $B > 0$, and the strip $S_\omega \triangleq \{t \in \mathbb{C} : |\operatorname{Im} t| < \omega\}$, define the holomorphic width

$$\Omega_B(b) \triangleq \sup\big\{\omega > 0 :\ t \mapsto X_t \text{ extends holomorphically to } S_\omega \text{ for all } \|z\| \leq B\big\}. \tag{16}$$

If, in addition, all such trajectories are bounded by M on S_ω and b_t is L-Lipschitz in space, then collocation approximates X_1 to accuracy ζ uniformly over $\|z\| \leq B$ with

$$N \lesssim \max(L, \omega^{-1})\Big(1 + L + \log \frac{M \max(L, \omega^{-1})}{\zeta}\Big)^2 \tag{17}$$

drift evaluations (Theorem K.8; Lee et al., 2018, Thm. 2.3). For VP schedules, $b_t(x) = \frac{\dot{\beta}_t}{\beta_t}(x + s_t(x))$, so each drift evaluation is a single score query.

Smoothing as early stopping. For Gaussian targets, the elliptic schedule makes every trajectory holomorphic on S_ω with $\omega = K(1-m)/K(m) \approx \pi/\log(16\kappa)$, independently of d and z (Proposition 3.1). On P_κ, analyticity can fail at $t = 1$ regardless of the schedule (Example K.1), although it holds at every $t \in (0,1)$ for analytic schedules (Section K.1). We therefore target a smoothed distribution: for $\widetilde{\sigma} \in (0,1)$, $z \sim N(0, I_d)$, and $x_1 \sim \pi$ (Lemma K.6),

$$\widetilde{\pi} \triangleq \mathcal{L}\big(\widetilde{\sigma} z + \sqrt{1-\widetilde{\sigma}^2}\, x_1\big) \in \mathsf{P}_{\widetilde{\kappa}}, \qquad \widetilde{\kappa} \triangleq \kappa\big(1 + \widetilde{\sigma}^2(\kappa-1)\big)^{-1} \le \kappa, \tag{18}$$

with smoothing error $W_2(\widetilde{\pi}, \pi) \le \sqrt{d/\kappa}\,\big(\sqrt{1+\widetilde{\sigma}^2(\kappa-1)} - 1\big)$. For a VP schedule (α, β), the PF-ODE transporting $N(0, I_d)$ to $\widetilde{\pi}$ and its score are

$$\dot{\widetilde{X}}_t = \widetilde{b}_t(\widetilde{X}_t) \triangleq \frac{\dot{\beta}_t}{\beta_t}\big(\widetilde{X}_t + \widetilde{s}_t(\widetilde{X}_t)\big), \qquad \widetilde{s}_t = s_{\tau(t)}, \tau(t) \triangleq \beta^{-1}\big(\sqrt{1-\widetilde{\sigma}^2}\,\beta_t\big) \in [0,1), \tag{19}$$

since the smoothed interpolant satisfies $\alpha_t z + \beta_t \widetilde{x}_1 \overset{(d)}{=} \alpha_{\tau(t)} z + \beta_{\tau(t)} x_1$ with $\widetilde{x}_1 \sim \widetilde{\pi}$. Hence each evaluation of $\widetilde{b}_t$ costs one query of the original score, and $\widetilde{X}_t = X_{\tau(t)}$, so integrating equation 19 over $[0,1]$ amounts to stopping the original PF-ODE early at $\tau(1) < 1$ (Section K). We seek the schedule maximizing the worst-case width $\omega_\star(\kappa, d, B, \widetilde{\sigma}) \triangleq \sup_{(\alpha,\beta)} \inf_{\pi \in \mathsf{P}_\kappa} \Omega_B(\widetilde{b})$.

Theorem 4.1 (Holomorphic width under the elliptic schedule)**.** *Let $\kappa, \widehat{\kappa} > 1$, $\widetilde{\sigma} \in (0,1)$, $B > 0$, and $D \triangleq \sqrt{d} + \sqrt{B^2 + d\log\kappa}$. Under the elliptic schedule (Definition 3.2) with parameter $\widehat{m} \triangleq 1 - \widehat{\kappa}^{-1}$, for every $\pi \in \mathsf{P}_\kappa$ and $\|z\| \le B$, the trajectory of equation 19 is holomorphic on $S_{\omega_{\widehat{\kappa}}}$ and*

$$\sup_{t \in S_{\omega_{\widehat{\kappa}}}} \|\widetilde{X}_t\| \le B + \sqrt{d} + 1, \qquad \omega_\star(\kappa, d, B, \widetilde{\sigma}) \;\ge\; \omega_{\widehat{\kappa}} \triangleq \frac{\min\big(1, \widetilde{\sigma}\sqrt{\widehat{\kappa}}\big)}{K(\widehat{m})\big(1 + 30\sqrt{1-\widetilde{\sigma}^2}\, D\big)}. \tag{20}$$

Choosing $\widehat{\kappa} = \widetilde{\sigma}^{-2}$, where $\min(1, \widetilde{\sigma}\sqrt{\widehat{\kappa}})$ saturates, tunes the schedule to the smoothing level rather than to κ. Restricting the infimum to Gaussian targets and applying Proposition 3.1 then gives

$$\frac{1}{D\log(4/\widetilde{\sigma})} \;\lesssim\; \omega_\star(\kappa, d, B, \widetilde{\sigma}) \;\le\; \frac{K(\widetilde{\kappa}^{-1})}{K(1-\widetilde{\kappa}^{-1})} \xrightarrow{\widetilde{\sigma} \to 0} \frac{K(\kappa^{-1})}{K(1-\kappa^{-1})} \approx \frac{\pi}{\log(16\kappa)}, \tag{21}$$

so the elliptic schedule is optimal up to the factor $D \asymp \sqrt{d\log\kappa} + B$ and logarithmic terms.

4.2 Score-Oracle Complexity of Log-Concave Sampling

With $\widehat{m} = 1 - \widetilde{\sigma}^2$, the smoothed drift satisfies $-K(\widehat{m}) I_d \preceq \nabla \widetilde{b}_t \preceq 0$ (Lemma K.7), so $L \le K(\widehat{m})$ and the flow $z \mapsto \widetilde{X}_t$ is non-expansive. Combining this with Theorem 4.1 and equation 17 gives the following result.

Theorem 4.2 (Score-oracle complexity for log-concave targets)**.** *Let $\pi \in \mathsf{P}_\kappa$ with $\kappa > 1$, let $\varepsilon \in (0, \bar{\varepsilon}]$ with $\bar{\varepsilon} \triangleq (\kappa-1)\sqrt{d/\kappa}$, and set $\widetilde{\sigma}^2 = \widehat{\kappa}^{-1} = \frac{\varepsilon}{2\bar{\varepsilon}}$, $B = \sqrt{d} + \sqrt{2\log(32/\varepsilon^2)}$, $\zeta = \frac{\varepsilon}{4}$. Integrate equation 19 under the elliptic schedule with parameter $\widehat{m} = 1 - \widehat{\kappa}^{-1}$ by collocation with tolerance ζ, starting from $\widetilde{X}_0 = \Pi_B(Z) \triangleq \min(1, B/\|Z\|)\, Z$ with $Z \sim N(0, I_d)$. The output $\widehat{X}$ satisfies $W_2(\mathcal{L}(\widehat{X}), \pi) \le \varepsilon$ using at most*

$$N \le C \log^3\!\Big(\frac{e\kappa(d+1)}{\varepsilon}\Big) \sqrt{d\log(\kappa+1) + \log\Big(\frac{e\kappa(d+1)}{\varepsilon}\Big)} = O\Big(\sqrt{d}\,\log^{3.5}\frac{\kappa d}{\varepsilon}\Big) \tag{22}$$

exact score evaluations, where C is a universal constant.

The error splits into the smoothing bias, the projection of the initialization, and the collocation error, each at most a constant fraction of ε. The inverse width contributes the square-root factor and one logarithm, and the accuracy term contributes the remaining $\log^2$ (Section K). Unlike Zoloflow, this result uses a generic solver and thus isolates the benefit of the elliptic schedule. The $\sqrt{d}$ factor stems from D and is absent from the Gaussian width in equation 21; whether it is intrinsic to P_κ remains open.

5 Experiments

The following experiments are meant to (i) visualize the performance optimality of the Zoloflow method for Gaussian distributions and to (ii) assess how well the elliptic schedule performs on more complex distributions under higher order solvers. We note that the elliptic finite difference discretization from Zoloflow is specific to gaussian distributions.

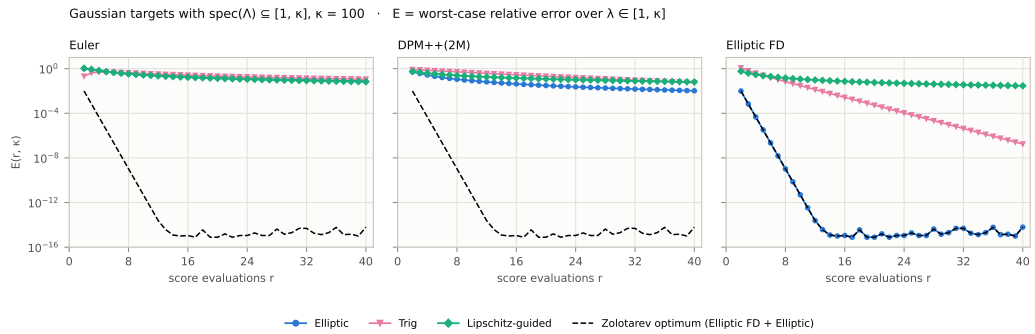

Figure 1: **Worst-case relative error E against number of score evaluations r on a Gaussian target with spectrum in $[1, 100]$.** Panels: solvers; colours: schedules, drawn only at the r valid for each pair. The black dashed line, the elliptic FD rule with the Elliptic schedule ($\widehat{E}_{100,r}$), is repeated in every panel.

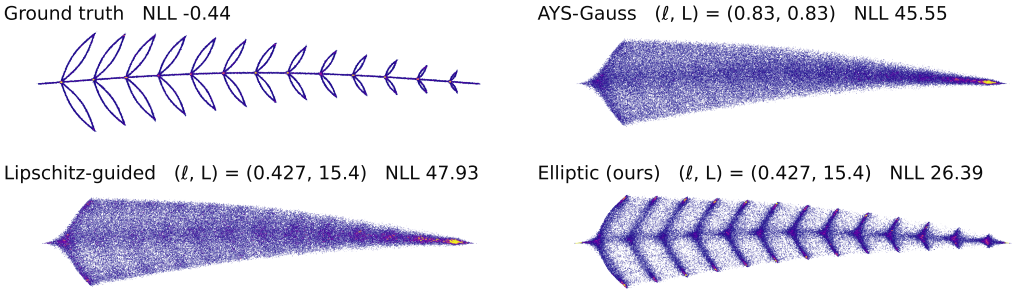

Figure 2: **Frond target,** DPM-Solver++(2M) with $r = 8$ score evaluations; 10^5 samples from the same initial noise in every panel. Titles give each schedule's (ℓ, L) and the NLL of its samples under the target. Details in Section A.3.

Gaussian distributions. Figure 1 compares solver–schedule pairs on a Gaussian target with spectrum $[1, \kappa]$, $\kappa = 100$. By Lemma 3.6, each pair alg builds a rational function $q_r^{\mathrm{alg}}(\lambda)$ acting along eigenmodes, and we plot its worst-case relative error $E_{\kappa,r}(\mathrm{alg}) = \max_{\lambda\in[1,\kappa]} |\sqrt{\lambda}\, q_r^{\mathrm{alg}}(\lambda) - 1|$, which controls the sampling error, $D_{\mathrm{KL}}(\widehat{\pi} \,\|\, \pi) \lesssim d\, E_{\kappa,r}(\mathrm{alg})^2$ by equation 80. No pair can fall below the Zolotarev error $E_{\kappa,r}$ of equation 49, which underlies the lower bound of Theorem 3.7, and Zoloflow, the elliptic schedule with the elliptic finite-difference rule, attains $\widehat{E}_{\kappa,r} = 2E_{\kappa,r}/(1-E_{\kappa,r})$ (Theorem 3.7) (Section A.1). More details and log-concave experiments are given in Appendices A.1 and A.2.

2D Toy experiments. Figure 2 tests the covariance choice of (ℓ, L) (Section H) on a strongly non-Gaussian target concentrated on the strokes of a fern frond, using DPM-Solver++(2M) with exact scores and $r = 8$ score evaluations. Only the elliptic schedule recovers the leaflets, cutting the negative log-likelihood $\mathrm{NLL} = -\frac{1}{N}\sum_{i=1}^{N} \log p_1(x_i)$ of the samples from about 46 to 26. The Lipschitz-guided schedule uses the same (ℓ, L), so the gain comes from the elliptic schedule itself. Section A.3 gives the setup and four more targets.

Image-generation We sample from the pretrained SiT-XL model (675M parameters; Ma et al., 2024)[1] on ImageNet 256×256 with classifier-free guidance scale 1.5, using DPM-Solver++(2M) throughout. SiT-XL learns the velocity field of the linear interpolant, which we convert to other schedules via Chen et al. (2025, Prop. 3.1). Since this target does not have a precision matrix spectrum in $[1, \kappa]$, we use the target-rescaled elliptic schedule (Section H) with hyperparameters ℓ and κ_{param}. Setup details, the list of schedules, and further results are in Sections A.4, A.4.2 and I.1.

[1] https://github.com/willisma/SiT

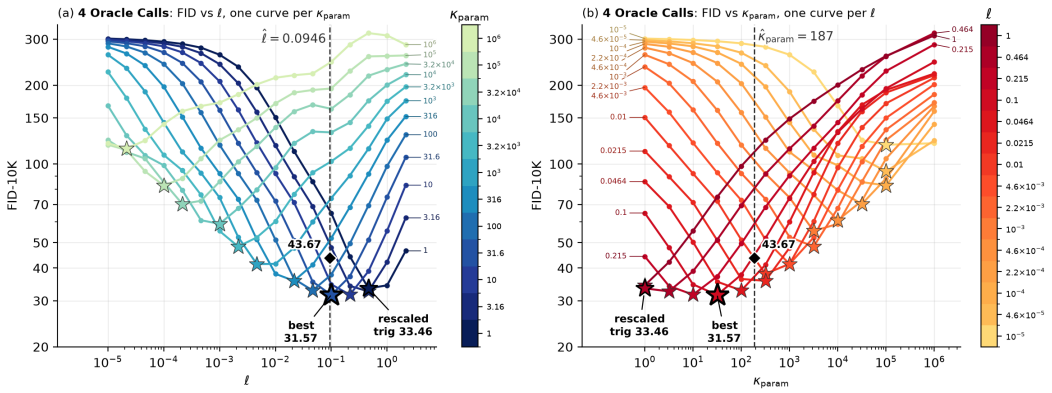

Figure 3: **FID-10K over the $(\ell, \kappa_{\text{param}})$ grid at 4 oracle calls**, using the target-rescaled elliptic schedule with DPM-Solver++(2M). (a) FID-10K vs. ℓ, one curve per κ_{param}; (b) FID-10K vs. κ_{param}, one curve per ℓ. Stars mark each curve's minimum; bold stars mark the best target-rescaled elliptic and best target-rescaled trigonometric configurations. Dashed lines mark the heuristic $\hat{\ell}$ and $\hat{\kappa}_{\text{param}}$, and the diamond is a run at $(\hat{\ell}, \hat{\kappa}_{\text{param}})$, achieving FID-10K 43.67.

Table 1: **FID-50k and IS by schedule and number of oracle calls.**

	4 calls		*8 calls*		*32 calls*	
Schedule	FID↓	IS↑	FID↓	IS↑	FID↓	IS↑
Linear	40.28	67.9	14.08	139.5	2.75	233.5
Target-rescaled trig. (AYS)	27.34	101.8	4.52	201.9	**2.07**	**250.3**
Target-rescaled elliptic (**ours**)	**25.51**	**106.8**	**3.89**	**210.4**	**2.07**	**250.3**

Figure 3 shows FID-10K for a sweep over the ℓ and κ_{param} hyperparameters of the target-rescaled elliptic schedule with DPM-Solver++(2M). The figure shows that the sampling performance is highly sensitive to both ℓ and κ_{param}. We take $\hat{\ell}$ and $\hat{L}$ as the 0.5th and 99.5th percentiles of the eigenvalues of the empirical precision matrix, and we set $\hat{\kappa}_{\text{param}} = \hat{L}/\hat{\ell}$. This gives $\hat{\ell} = 0.0946$ and $\hat{\kappa}_{\text{param}} = 187$, which attains an FID of 43.67.

Sweeping over ℓ and κ_{param}, we see that Performance is sensitive to both hyperparameters (Figures 3 and 15). A simple heuristic takes $\hat{\ell}$ and $\hat{L} := \hat{\ell}\hat{\kappa}_{\text{param}}$ as the 0.5th and 99.5th percentiles of the empirical precision spectrum. This lands in the low-FID region of the sweep, making it a sensible starting point for further tuning. Table 1 compares the best target-rescaled elliptic and target-rescaled trigonometric configurations from the sweep against the linear and trigonometric schedules. The choice of schedule matters most at low budgets, where the target-rescaled elliptic schedule outperforms the alternatives; as the budget grows, all schedules converge to the same exact ODE solution and the differences vanish.

6 Discussion

We designed interpolant schedules and discretizations jointly through two complementary lenses, minimax rational approximation and complex-analytic regularity, which lead to the same elliptic schedule. Several limitations remain. Our analysis assumes exact scores, and extending it to approximate scores is left open. For Gaussian targets, Zoloflow is minimax optimal, but its discretization is tailored to Gaussian structure, and we do not identify an optimal discretization for log-concave targets. Our log-concave guarantees are upper bounds without matching lower bounds, so it remains open whether the $\sqrt{d}$ factor is necessary. Finally, for non-log-concave targets, the schedule parameter κ must be chosen heuristically, since no curvature bounds are available.

7 AI USAGE

We did not use generative AI tools for ideation. In particular, the method for constructing an optimal schedule and solver for a Gaussian distribution was developed without AI assistance. We used AI to assist in writing proofs and, in some cases, to suggest mathematical tools for proving claims. We used AI as a search-engine for discovering papers in the field. We also used generative AI tools to generate synthetic data distributions, and to write and proofread code. All code was verified for correctness through manual review and unit testing. Additionally, we used generative AI tools to edit grammar and wording, and we evaluated each suggestion before deciding whether to incorporate it. We reviewed all AI-assisted work and take full responsibility for the final content of this paper, including any text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Naman Agarwal, Surbhi Goel, and Cyril Zhang. Acceleration via Fractal Learning Rate Schedules, June 2021. URL http://arxiv.org/abs/2103.01338. arXiv:2103.01338 [cs.LG].

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Jason M. Altschuler and Pablo A. Parrilo. Acceleration by Stepsize Hedging I: Multi-Step Descent and the Silver Stepsize Schedule. *Journal of the ACM*, 72(2):1–38, April 2025. ISSN 0004-5411, 1557-735X. doi: 10.1145/3708502. URL http://arxiv.org/abs/2309.07879. arXiv:2309.07879 [math.OC].

Jason M. Altschuler and Pablo A. Parrilo. Stepsize Hedging: an Alternative Mechanism for Accelerating Gradient Descent, June 2026. URL http://arxiv.org/abs/2605.31386. arXiv:2605.31386 [math.OC].

Santiago Aranguri, Giulio Biroli, Marc Mezard, and Eric Vanden-Eijnden. Optimizing Noise Schedules of Generative Models in High Dimensionss, January 2025. URL http://arxiv.org/abs/2501.00988. arXiv:2501.00988 [cs.LG].

Bernhard Beckermann and Alex Townsend. On the singular values of matrices with displacement structure. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1227–1248, 2017.

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow ODE is provably fast, 2023a. URL https://arxiv.org/abs/2305.11798.

Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru R. Zhang. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions, 2023b. URL https://arxiv.org/abs/2209.11215.

Yifan Chen, Eric Vanden-Eijnden, and Jiawei Xu. Lipschitz-guided design of interpolation schedules in generative models. *arXiv preprint arXiv:2509.01629*, 2025.

Sinho Chewi. *Log-concave sampling*. Forthcoming, 2026. Available online at https://chewisinho.github.io/.

Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional Gaussians with the same mean, 2023. URL https://arxiv.org/abs/1810.08693.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021.

Sander Dieleman. Noise schedules considered harmful, June 2024. URL https://sander.ai/2024/06/14/noise-schedules.html. Blog post.

DLMF. NIST Digital Library of Mathematical Functions, Sections 22.13, 22.6. https://dlmf.nist.gov/, 2026. Release 1.2.7 of 2026-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

Yuan Gao, Jian Huang, and Yuling Jiao. Gaussian Interpolation Flows, July 2024. URL http://arxiv.org/abs/2311.11475. arXiv:2311.11475 [stat.ML].

Khashayar Gatmiry, Sitan Chen, and Adil Salim. High-accuracy and dimension-free sampling with diffusions, January 2026. URL http://arxiv.org/abs/2601.10708. arXiv:2601.10708 [cs.LG].

Evan S Gawlik. Zolotarev iterations for the matrix square root. *SIAM Journal on Matrix Analysis and Applications*, 40(2):696–719, 2019.

Nicholas Hale, Nicholas J. Higham, and Lloyd N. Trefethen. Computing A^α, $\log(A)$, and related matrix functions by contour integrals. *SIAM Journal on Numerical Analysis*, 46(5):2505–2523, 2008. doi: 10.1137/070700607. URL https://doi.org/10.1137/070700607.

Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models, 2022. URL https://arxiv.org/abs/2206.00364.

Diederik P. Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models, 2023. URL https://arxiv.org/abs/2107.00630.

Derek F Lawden. *Elliptic functions and applications*. Springer Science & Business Media, 2013.

Yin Tat Lee, Zhao Song, and Santosh S. Vempala. Algorithmic Theory of ODEs and Sampling from Well-conditioned Logconcave Densities, December 2018. URL http://arxiv.org/abs/1812.06243. arXiv:1812.06243 [cs.DS].

Jingbo Liu. Smoothed score queries and the complexity of sampling, 2026. URL https://arxiv.org/abs/2605.27769.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps, 2022a. URL https://arxiv.org/abs/2206.00927.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models, 2022b. URL https://arxiv.org/abs/2211.01095.

Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pp. 23–40. Springer, 2024.

Masatake Mori. Discovery of the Double Exponential Transformation and Its Developments. *Publications of the Research Institute for Mathematical Sciences*, 41(4):897–935, December 2005. ISSN 0034-5318, 1663-4926. doi: 10.2977/prims/1145474600. URL https://ems.press/doi/10.2977/prims/1145474600.

Yuji Nakatsukasa and Roland W. Freund. Computing fundamental matrix decompositions accurately via the matrix sign function in two iterations: The power of Zolotarev's functions. *SIAM Review*, 58(3):461–493, 2016. doi: 10.1137/140990334.

Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models, 2021. URL https://arxiv.org/abs/2102.09672.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, 2022. doi: 10.1109/CVPR52688.2022.01042.

Amirmojtaba Sabour, Sanja Fidler, and Karsten Kreis. Align your steps: optimizing sampling schedules in diffusion models. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Maximilian Seitzer. pytorch-fid: FID Score for PyTorch. https://github.com/mseitzer/pytorch-fid, August 2020. Version 0.3.0.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022. URL https://arxiv.org/abs/2010.02502.

Frank Stenger. Summary of sinc numerical methods. *Journal of Computational and Applied Mathematics*, 121(1-2):379–420, September 2000. ISSN 03770427. doi: 10.1016/S0377-0427(00)00348-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0377042700003484.

Hidetosi Takahasi and Masatake Mori. Double Exponential Formulas for Numerical Integration. *Publications of the Research Institute for Mathematical Sciences*, 9(3):721–741, December 1973. ISSN 0034-5318, 1663-4926. doi: 10.2977/prims/1195192451. URL https://ems.press/doi/10.2977/prims/1195192451.

Gerald Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.

Lloyd N. Trefethen. *Approximation Theory and Approximation Practice, Extended Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019. doi: 10.1137/1.9781611975949. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611975949.

Lloyd N Trefethen and Heather D Wilber. Computation of Zolotarev rational functions. *SIAM Journal on Scientific Computing*, 47(4):A2205–A2220, 2025.

Qinsheng Zhang and Yongxin Chen. Fast sampling of diffusion models with exponential integrator, 2023. URL http://arxiv.org/abs/2204.13902.

A EXPERIMENTS

A.1 GAUSSIAN EXPERIMENTS

Summary. On a Gaussian target, every solver we test acts on each precision mode λ as a rational function $M(\lambda)$ (Section F), so we compute its worst-case relative error over the spectrum, $E = \max_\lambda \left|\sqrt{\lambda}\, M(\lambda) - 1\right|$, exactly. E controls the KL divergence, $D_{\mathrm{KL}} \lesssim d\, E^2$ (Section E). Two findings:

- *Zoloflow wins by orders of magnitude.* The elliptic schedule with the elliptic finite-difference (FD) rule (Definition 3.4) reaches round-off with 12 score evaluations at $\kappa = 100$, where every other solver–schedule pair is still above 10^{-2} (Section A.1.1).
- *Both parameters must match the spectrum.* On an unnormalized spectrum (Section G), widening the schedule without rescaling, or rescaling without widening, stays orders of magnitude above the optimum (Section A.1.2).

A.1.1 SOLVERS AND SCHEDULES ON A NORMALIZED SPECTRUM

Figures 4 and 5 plot E on the normalized spectrum $[1, \kappa]$ for five standard solvers and the elliptic FD rule (panels), each with three schedules (colours). Compare the black dashed line, the elliptic FD rule with the Elliptic schedule, with every coloured line. The dashed line is the endpoint-normalized Zolotarev error $\widehat{E}_{\kappa,r}$ equation 56 and falls geometrically in r. At $\kappa = 100$ (Figure 4) it reaches round-off at $r = 12$, where the best coloured line is at 1.3×10^{-2}. At $r = 16$ (Figure 5) it stays below 2.4×10^{-8} up to $\kappa = 10^6$, where the best coloured line is at 0.15. Neither half works alone: with a standard solver, the Elliptic schedule (blue, first five panels) stays above 10^{-4} for every $r \le 40$, and the FD rule with Trig (pink, last panel) is still at 10^{-2} at $r = 12$.

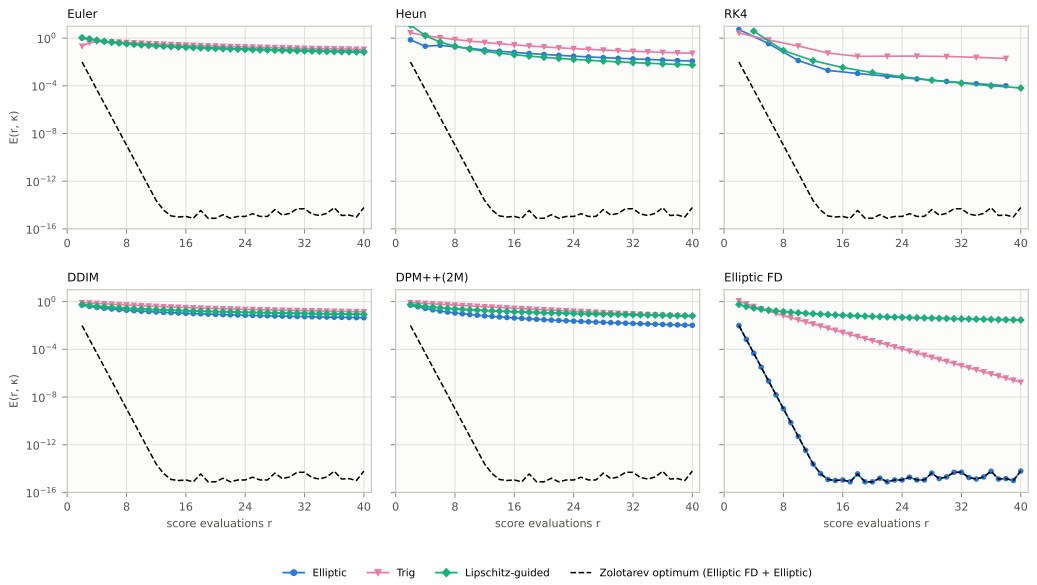

Figure 4: **Worst-case relative error E against the number of score evaluations r on the spectrum $[1, 100]$.** Panels: solvers; colours: schedules, drawn only at the r valid for each pair. The black dashed line, the elliptic FD rule with the Elliptic schedule ($\widehat{E}_{100,r}$), is repeated in every panel.

A.1.2 BOTH ENDPOINTS MUST MATCH THE SPECTRUM

Figure 6 takes a Gaussian whose spectrum $[\lambda_{\min}, \lambda_{\max}] = [0.01, 1]$ is not normalized. It runs Generalized Zoloflow (Definition G.6), the exponential-gauge rescaling of Section G.2.2, with $r = 12$ at every $(\ell, \kappa_{\mathrm{param}})$ on a grid, where ℓ takes the place of $\lambda_{\min}$ and $\kappa_{\mathrm{param}} = L/\ell$ that of κ.

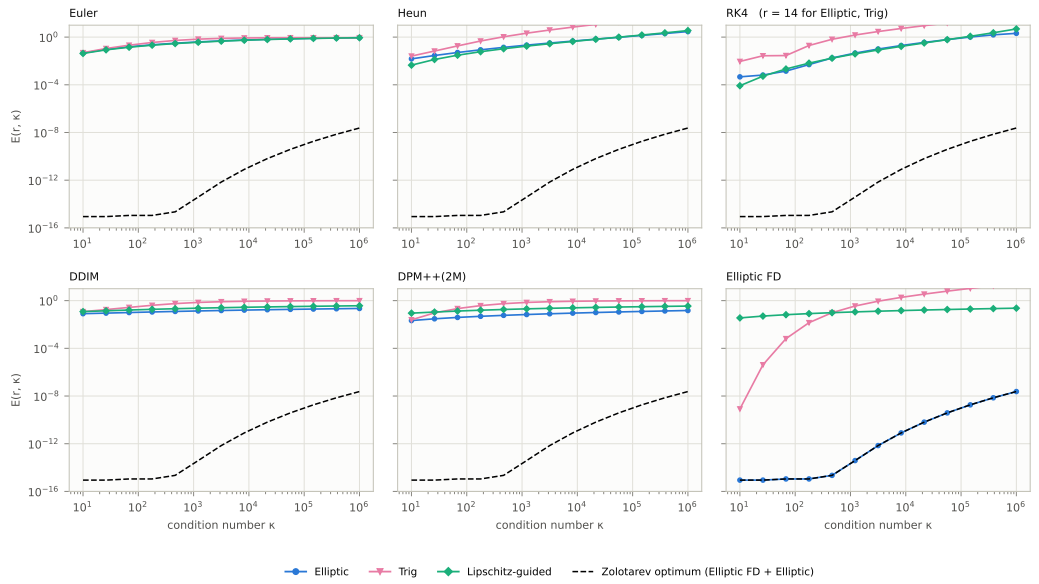

Figure 5: Worst-case relative error E against κ on the spectrum $[1, \kappa]$ with $r = 16$ ($r = 14$ for RK4 with Elliptic and Trig, where 16 is not valid). Layout as in Figure 1.

By equation 137, $M(\lambda) = \ell^{-1/2}\, \widehat{q}_{\kappa_{\text{param}},r}(\lambda/\ell)$ per mode, so $(\ell, L) = (\lambda_{\min}, \lambda_{\max})$ gives $E = \widehat{E}_{100,12}$.

Each curve fixes κ_{param}, and its star marks the curve's minimum over ℓ, placed in the middle of its flat segment where there is one. Only the $\kappa_{\text{param}} = 100$ curve reaches $\widehat{E}_{100,12}$, at $\ell = \lambda_{\min}$ (blue circle). Without rescaling ($\ell = 1$), the lowest curve is $\kappa_{\text{param}} = 1$, which is Trig (pink triangle, $E = 1.3 \times 10^{-2}$), so widening the schedule alone does not help. Without widening ($\kappa_{\text{param}} = 1$), the best scale is the geometric centre $\ell = \sqrt{\lambda_{\min}\lambda_{\max}} = 0.1$ (orange square, $E = 1.6 \times 10^{-7}$), about seven orders of magnitude above the endpoints (2.4×10^{-14}). The flat segments are exact: once $[\ell, L]$ contains the spectrum, $E = \widehat{E}_{\kappa_{\text{param}},r}$ for every such ℓ, so a schedule wider than the spectrum pays for its own, larger κ_{param}.

A.1.3 Setup

Solvers and schedules. Euler, Heun, RK4, DDIM and DPM-Solver++(2M) (Section J), and the elliptic FD rule of Definition 3.4 on its shared grid equation 186. The schedules are Elliptic equation 156 with $(\ell, L) = (1, \kappa)$, Trig ($\kappa = 1$) and Lipschitz-guided equation 157 (Chen et al., 2025).

Initialization and budget. Every solver but FD starts from $x(0) = z$ at $t = 0$ and steps on a uniform grid to $t = 1$. The budget r counts score evaluations; evaluations that do not depend on the target are not counted: the denoiser at $t = 0$, and the drift at $t = 0$ and $t = 1$ for Elliptic and Trig, where $\frac{d}{dt}\beta_t^2 = 0$. A budget that gives no whole number of steps is not valid for that pair.

Computing E. Each solver is run per mode on a fine log-spaced λ grid with local refinement; values below 10^{-14} are round-off.

A.2 Log-concave experiments

Summary. The Gaussian theory sets the elliptic schedule parameters from the endpoints of the precision spectrum. For non-Gaussian targets we set it from the covariance, $(\ell, L) = \big(\lambda_{\min}(\Sigma^{-1}), \lambda_{\max}(\Sigma^{-1})\big)$ (Elliptic (ours), Section H), rather than from the curvature bounds $\lambda_{\min} I \preceq \nabla^2 V \preceq \lambda_{\max} I$ (Elliptic, curvature endpoints). The schedule should widen when scales differ *across* directions, which Σ measures, but not when curvature varies *within* a direction, which

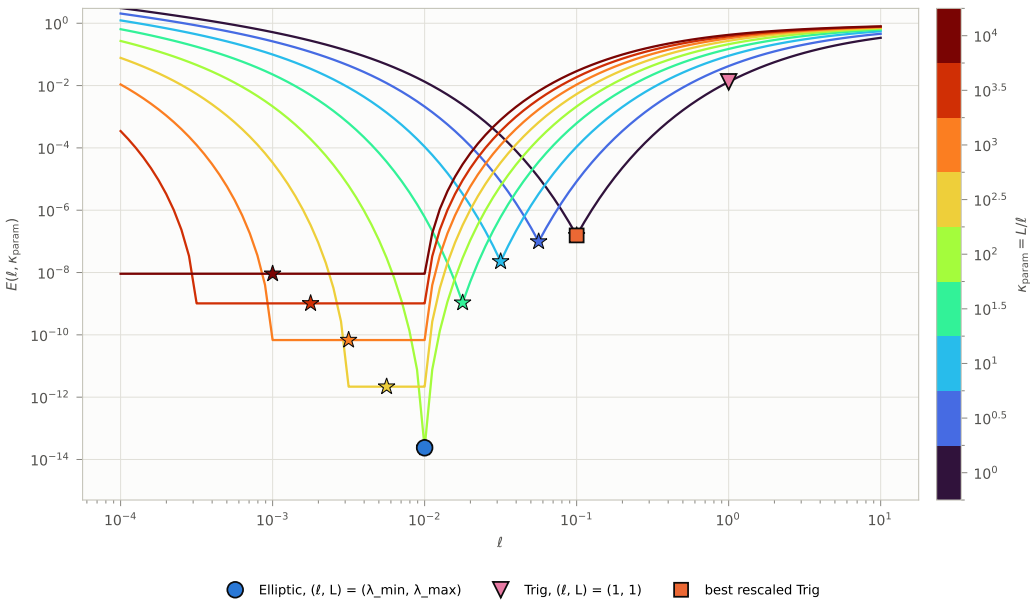

Figure 6: Gaussian target with spectrum $[0.01, 1]$, $r = 12$: E against ℓ, one curve per $\kappa_{\text{param}} \triangleq \frac{L}{\ell}$; stars mark each curve's minimum over ℓ, in the middle of its flat segment where there is one, on the grid $\log_{10} \ell \in [-4, 1]$ (step 0.05), $\log_{10} \kappa_{\text{param}} \in [0, 4]$ (step 0.125). Blue circle: $(\ell, L) = (\lambda_{\min}, \lambda_{\max})$; pink triangle: Trig, $(\ell, L) = (1, 1)$; orange square: the best rescaled Trig, $\kappa_{\text{param}} = 1$.

widens the curvature bounds without changing Σ. We test this on log-concave product targets with exact scores, with KL computed by quadrature, so every difference is discretization error. On two-dimensional targets that separate the two effects (Figure 7), the covariance rule is the only rule that is never far from the best. In $d = 32$ (Figure 8) it lies within a factor 1.5–3.5 of the best (ℓ, L) on a full grid for DPM-Solver++(2M), where AYS-Gauss is 10–21 times worse. Schedules, targets and the metric are defined at the end (Sections A.2.3 to A.2.5).

A.2.1 Two-dimensional targets: which rule to use

Figure 7 plots KL against the budget r on four two-dimensional targets (rows, Table 2) for four solvers (columns). Compare the purple line (Elliptic, ours) with orange (AYS-Gauss) and blue (curvature endpoints). The rows cover three regimes:

- *A, B: scales differ across directions.* Purple lies below orange, by a median factor of 2 to 4, for every solver except Euler. Blue is close to purple for DDIM and DPM-Solver++(2M) and worse for Heun and Euler.
- *E1: $\Sigma = I$, curvature varies only within a direction.* Purple coincides with orange and pink, while blue widens the schedule to $\kappa = 100$ and lies far above (a median factor of 71).
- *G: Gaussian.* Purple coincides with blue, the Gaussian optimum, and lies far below orange (a median factor of 19).

Purple is never far from the best line in any regime. The exception is Euler, for which a near-trig schedule is better.

A.2.2 Validation in $d = 32$: distance to the best (ℓ, L)

Figure 8 evaluates DPM-Solver++(2M) with $r = 8$ at every (ℓ, L) on a grid, on the 32-dimensional targets H1 and H2. Each curve is a slice of the KL$/d$ landscape and its star marks the slice's minimum, so the lowest stars locate the best settings. The question is where the purple cross (Elliptic, ours) sits relative to them. It lies within a factor of 1.5 (H1) and 3.5 (H2) of the best grid setting, while AYS-Gauss (orange), Trig (pink) and Lipschitz-guided (dashed line) are 7 to 131 times above

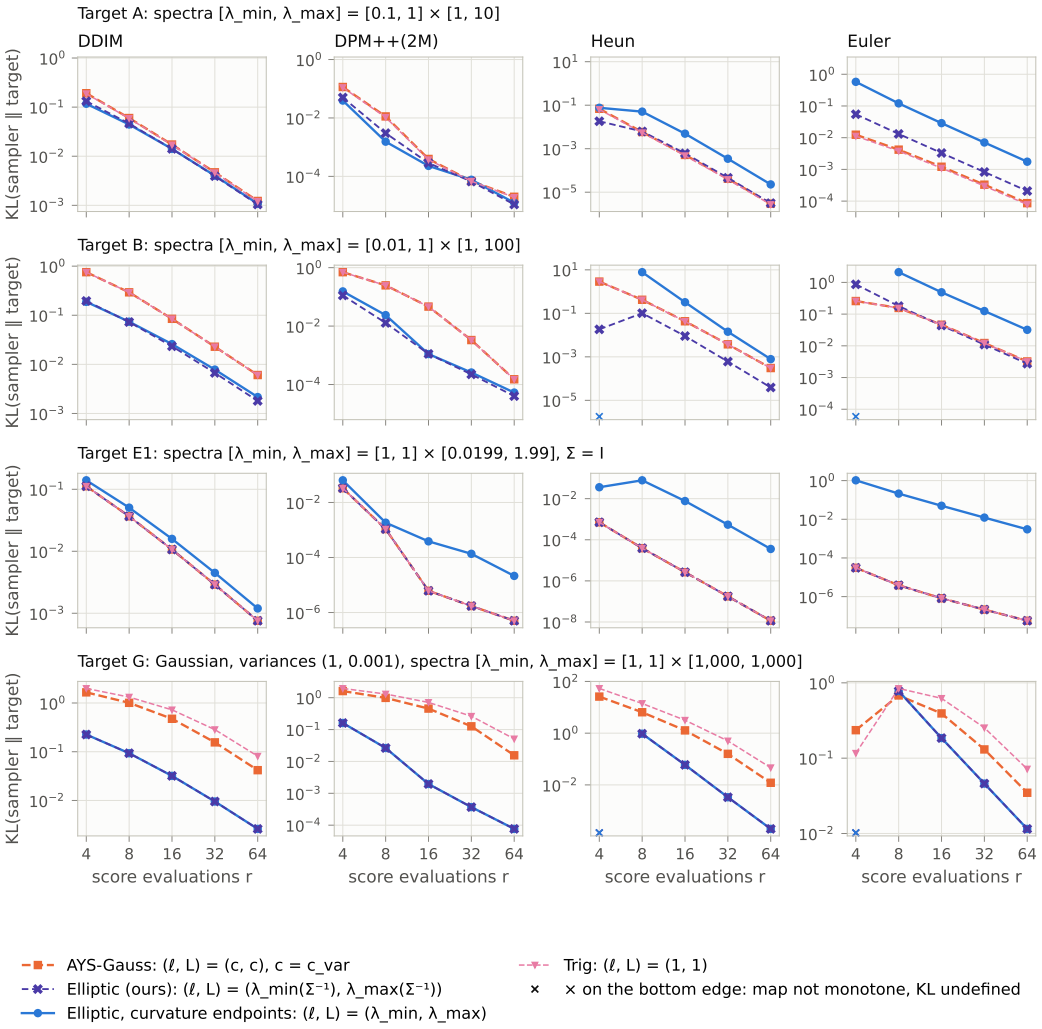

Figure 7: $\mathrm{KL}(F_{\#}\gamma \,\|\, p_1)$ against the number of score evaluations r. Rows: the two-dimensional targets of Table 2, grouped by regime (A, B; E1; G); columns: solvers. Compare purple (Elliptic, ours) with orange (AYS-Gauss) and blue (curvature endpoints); Trig (pink) coincides with AYS-Gauss wherever $c \approx 1$ (A, B, E1). A $\times$ on the bottom edge marks a folded map (KL undefined).

it. The curvature endpoints (blue) are near-optimal on H1 but 7 times above on H2, whose coordinates are flat. For Euler and Heun (not shown) the best setting has $\kappa \leq 6$ and AYS-Gauss is the better choice.

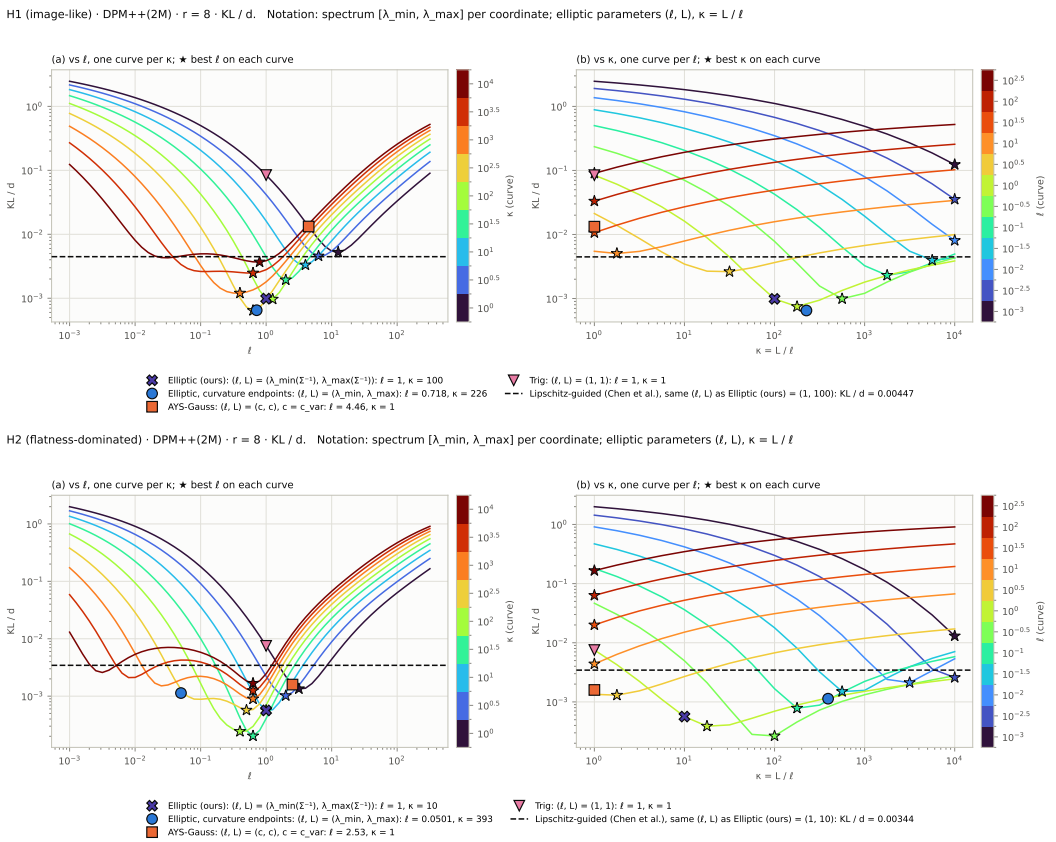

Figure 8: KL$/d$ of DPM-Solver++(2M) with $r = 8$ over the grid $\log_{10}\ell \in [-3, 2.5]$, $\log_{10}\kappa \in [0, 4]$; top H1, bottom H2. (a) Against ℓ, one curve per $\kappa = L/\ell$; (b) against κ, one curve per ℓ; stars mark each curve's minimum. The purple cross (Elliptic, ours) sits near the lowest stars; the dashed line is Lipschitz-guided at the same (ℓ, L).

A.2.3 Schedules under comparison

All schedules are variance preserving, $\alpha_t^2 + \beta_t^2 = 1$, with $(\alpha_0, \beta_0) = (1, 0)$ and $(\alpha_1, \beta_1) = (0, 1)$. The elliptic schedule with condition number $\kappa \geq 1$ is

$$\alpha_t = \mathrm{cn}\big(K(m)\, t \mid m\big), \qquad \beta_t = \mathrm{sn}\big(K(m)\, t \mid m\big), \qquad m = 1 - \kappa^{-1}, \tag{23}$$

which reduces to $(\cos\frac{\pi t}{2}, \sin\frac{\pi t}{2})$ at $\kappa = 1$.

A *schedule setting* is a pair (ℓ, L) with $\ell \leq L$. It is realized by *target rescaling* (Section G.1): the elliptic schedule equation 23 with $\kappa = L/\ell$ is run on the rescaled target $\sqrt{\ell}\, x_1$, and the output is divided by $\sqrt{\ell}$. Thus ℓ sets the *scale* and $\kappa = L/\ell$ sets the *shape*. We compare four settings of this family:

schedule	setting (ℓ, L)
Trigonometric	$(1, 1)$
AYS-Gauss	(c, c), $c = d/\mathbb{E}\|x_1\|^2$
Elliptic (ours)	$\big(\lambda_{\min}(\Sigma^{-1}),\ \lambda_{\max}(\Sigma^{-1})\big)$, Section H
Elliptic, curvature endpoints	$(\lambda_{\min}, \lambda_{\max})$

AYS-Gauss is the Align Your Steps schedule for isotropic Gaussian data with the target's average variance (Sabour et al., 2024, Theorem 3.1). The sweep also shows the Lipschitz-guided schedule equation 157 of Chen et al. (2025) at the same (ℓ, L) as Elliptic (ours).

A.2.4 Target family: products of one-dimensional Gaussian mixtures

Definition. The target is a product of d independent, symmetric, two-component mixtures,

$$p_1(y) = \prod_{i=1}^{d} p_i(y_i), \qquad p_i = \tfrac{1}{2}\,\mathcal{N}(\mu_i, \sigma_i^2) + \tfrac{1}{2}\,\mathcal{N}(-\mu_i, \sigma_i^2), \qquad 0 \le \mu_i < \sigma_i. \tag{24}$$

Writing $p_i = e^{-V_i}$, a direct computation gives

$$V_i(y) = \frac{y^2 + \mu_i^2}{2\sigma_i^2} - \log\cosh\Big(\frac{\mu_i y}{\sigma_i^2}\Big) + \tfrac{1}{2}\log(2\pi\sigma_i^2), \qquad V_i''(y) = \frac{1}{\sigma_i^2} - \frac{\mu_i^2}{\sigma_i^4}\operatorname{sech}^2\Big(\frac{\mu_i y}{\sigma_i^2}\Big). \tag{25}$$

The target has mean zero, covariance $\mathrm{Cov}(x_1) = \mathrm{diag}(\sigma_i^2 + \mu_i^2)$, and second moment $\mathbb{E}\|x_1\|^2 = \sum_i(\sigma_i^2 + \mu_i^2)$.

Curvature. The Hessian $\nabla^2 V(y) = \mathrm{diag}\,\big(V_1''(y_1), \ldots, V_d''(y_d)\big)$ is diagonal at every point: its eigenbasis is fixed and its eigenvalues vary with position. With $\rho_i = \mu_i/\sigma_i$,

$$\begin{aligned} &V_i''(y) \in \Big[\lambda_{\min,i},\ \lambda_{\max,i}\Big), \\ &\lambda_{\min,i} = \frac{1-\rho_i^2}{\sigma_i^2} \text{ (attained at } y = 0), \\ &\lambda_{\max,i} = \frac{1}{\sigma_i^2} \text{ (approached as } |y| \to \infty), \end{aligned} \tag{26}$$

so p_1 is $\lambda_{\min}$-strongly log-concave and $\lambda_{\max}$-smooth with $\lambda_{\min} = \min_i \lambda_{\min,i}$, $\lambda_{\max} = \max_i \lambda_{\max,i}$. Strong log-concavity requires $\rho_i < 1$ strictly; at $\rho_i = 1$ the mixture is exactly at the threshold where it would become bimodal and $\lambda_{\min,i} = 0$. To prescribe a per-coordinate curvature range $[\lambda_{\min,i}, \lambda_{\max,i}]$ we invert equation 26:

$$\sigma_i^2 = \frac{1}{\lambda_{\max,i}}, \qquad \mu_i^2 = \frac{\lambda_{\max,i} - \lambda_{\min,i}}{\lambda_{\max,i}^2}, \qquad \text{so that } \kappa_i \triangleq \frac{\lambda_{\max,i}}{\lambda_{\min,i}} = \frac{1}{1-\rho_i^2}. \tag{27}$$

The choice $\lambda_{\min,i} = \lambda_{\max,i}$ (i.e. $\mu_i = 0$) makes coordinate i Gaussian with precision $\lambda_{\max,i}$. Equivalently, a coordinate with variance v_i and $\rho_i^2 \in [0,1)$ has $\sigma_i^2 = v_i/(1+\rho_i^2)$, $\mu_i^2 = v_i\rho_i^2/(1+\rho_i^2)$ and curvature range $[(1-\rho_i^4)/v_i,\ (1+\rho_i^2)/v_i]$. We call coordinates with $\rho_i \approx 1$ *flat*: their curvature at the center, $\lambda_{\min,i}$, approaches 0.

Targets. Table 2 lists the targets. A and B have scales that differ across directions; E1 has $\Sigma = I$, so its curvature varies only within a direction; G is Gaussian. In $d = 32$, H1 spreads its variances over two decades with mild flattening, and H2 is dominated by flattening.

target	d	coordinates
A	2	curvature ranges $[0.1, 1]$, $[1, 10]$
B	2	curvature ranges $[0.01, 1]$, $[1, 100]$
E1	2	$v = (1, 1)$, $\rho^2 = (0, 0.99)$
G	2	Gaussian, $v = (1, 10^{-3})$
H1	32	$v_k = 10^{-2k/31}$, $\rho_k^2 = 0.8\, g_k$
H2	32	$v_k = 10^{-k/31}$, $\rho_k^2 = 0.9 + 0.09\, g_k$

Table 2: Log-concave targets. For H1 and H2, $k = 0, \ldots, 31$ and $g_k = (k+1)\omega - \lfloor (k+1)\omega \rfloor$ with $\omega = (\sqrt{5}-1)/2$.

Closed-form interpolated law. Because the interpolant noise is isotropic, $x_t = \alpha_t z + \beta_t x_1$ remains a product, and each factor is again a two-component mixture:

$$x_{t,i} \sim \tfrac{1}{2}\,\mathcal{N}(\beta_t\mu_i,\, v_{t,i}) + \tfrac{1}{2}\,\mathcal{N}(-\beta_t\mu_i,\, v_{t,i}), \qquad v_{t,i} = \alpha_t^2 + \beta_t^2\sigma_i^2. \tag{28}$$

Hence the score is available exactly for every t,

$$s_{t,i}(x) = \partial_x \log p_{t,i}(x) = -\frac{x}{v_{t,i}} + \frac{\beta_t\mu_i}{v_{t,i}}\tanh\Big(\frac{\beta_t\mu_i\, x}{v_{t,i}}\Big), \tag{29}$$

and $\mathrm{Var}(x_{t,i}) = \alpha_t^2 + \beta_t^2(\sigma_i^2 + \mu_i^2)$. Every solver is therefore run with the *exact* score, so the measured error is discretization error alone and is not confounded with score estimation.

A.2.5 Metrics

Let $F : \mathbb{R}^d \to \mathbb{R}^d$ denote the map $z \mapsto x_r$ implemented by a solver with r score queries, and let $\gamma = \mathcal{N}(0, I_d)$. The sampler's output law is the pushforward $F_{\#}\gamma$.

KL divergence by change of variables. If F is a C^1 bijection, then $(F_{\#}\gamma)(F(z)) = \gamma(z)/|\det \nabla F(z)|$, and substituting $x = F(z)$ in the definition of KL gives

$$\mathrm{KL}\big(F_{\#}\gamma \,\big\|\, p_1\big) = \mathbb{E}_{z\sim\gamma}\Big[\log\gamma(z) - \log|\det\nabla F(z)| - \log p_1\big(F(z)\big)\Big]. \tag{30}$$

Every term is available exactly at every z: $\log p_1$ is the closed-form mixture density, and ∇F follows from the chain rule through the solver. For a one-query step $x \mapsto a_j x + b_j\, w_{t_j}(x)$ the derivative is $a_j + b_j\,\partial_x w_{t_j}(x)$, with $\partial_x w = 1 + \partial_x s$ closed form from equation 29; multistage and multistep solvers propagate the running Jacobian through their stages. In the product setting F is coordinate-wise, $F(z) = (F_1(z_1), \ldots, F_d(z_d))$, so equation 30 splits into one-dimensional integrals against the standard normal,

$$\mathrm{KL} = \sum_{i=1}^{d} \mathrm{KL}_i, \qquad \mathrm{KL}_i = \mathbb{E}_{\zeta\sim\mathcal{N}(0,1)}\Big[\log\varphi(\zeta) - \log F_i'(\zeta) - \log p_i\big(F_i(\zeta)\big)\Big], \tag{31}$$

with $\mathbb{E}\log\varphi = -\frac{1}{2}(1 + \log 2\pi)$ known exactly. The remaining expectation is computed by 256-point Gauss–Hermite quadrature per coordinate, so no quantity involves Monte Carlo sampling; re-evaluating with 512 nodes changes the reported values by less than $\max(1\%, 10^{-14})$ in all but 2 of about 1,500 cases.

Validity. Formula equation 30 requires F_i to be strictly increasing. We check $F_i'(\zeta) > 0$ at every quadrature node and coordinate; if the check fails the map has folded and KL is undefined (marked '$\times$' in the figures). On the two-dimensional targets folding occurs only for Euler and Heun at $r = 4$ with $\kappa \geq 100$; DDIM and DPM-Solver++(2M) never fold.

Sanity check against the Gaussian theory. On a Gaussian coordinate with precision λ every solver's map is linear, $F_i(\zeta) = q\,\zeta$, and equation 31 evaluates to $\frac{1}{2}\big(\lambda q^2 - 1 - \log\lambda q^2\big)$, which is the per-mode identity equation 80 with $1 + e = \sqrt{\lambda}\,q$. For the elliptic finite-difference rule on Gaussian targets the pipeline reproduces the normalized Zolotarev multiplier $\hat{q}_{\kappa,r}(\lambda)$ to relative accuracy 10^{-14}.

A.3 2D Toy Experiments

Setup. Each target is a two-dimensional Gaussian mixture that traces a line drawing: the word "ZOLO", a double helix, an ECG trace, a fern frond, and a radial fractal tree. Each plot is made via a mixture of gaussian spikes. The interpolated law p_t of a Gaussian mixture has a closed-form score, so every run uses the exact score and all error comes from the discretization. We sample with DPM-Solver++(2M) (Section J) using $r = 8$ score evaluations on the grid equation 186, and draw $N = 10^5$ samples from the same initial noise z for every schedule. We compare four schedules, each realized by target rescaling (Section G.1) with setting (ℓ, L):

- *Trigonometric*: $(\ell, L) = (1, 1)$, the Trigonometric schedule;

- *AYS-Gauss*: $(\ell, L) = (c, c)$ with $c = 2/\operatorname{tr}\Sigma$, the closed-form schedule of Align Your Steps for isotropic Gaussian data (Sabour et al., 2024, Theorem 3.1) with the data variance set to the target's average variance.
- *Lipschitz-guided*: the schedule of Chen et al. (2025, Proposition 3.9), which minimizes the averaged squared Lipschitz constant of the drift for Gaussian targets; in our notation it is equation 157 (Section I.5), used with the same heuristic (ℓ, L) as Elliptic;
- *Elliptic* (ours): $(\ell, L) = \big(1/\lambda_{\max}(\Sigma),\ 1/\lambda_{\min}(\Sigma)\big)$, the covariance heuristic of Section H.

Here Σ is the exact covariance of the target mixture, so Lipschitz-guided and Elliptic use the same spectral information and differ only in the shape of the schedule. The tree has eight-fold rotational symmetry, so its covariance is a multiple of I_2: Elliptic coincides with AYS-Gauss and Lipschitz-guided reduces to the square-root schedule. It serves as a control. In each figure, the panel titles give the schedule, its setting (ℓ, L) and NLL.

Metric. *NLL* is the average negative log-density of the samples under the exact target mixture, $\mathrm{NLL} = -\frac{1}{N}\sum_{i=1}^{N} \log p_1(x_i)$, the metric used by Sabour et al. (2024). It is a cross-entropy, so it rewards over-concentrated samples and can fall below the target's own value; the ground-truth panel of each figure reports the NLL of 10^5 exact samples for reference.

A.4 Image generation experiments

A.4.1 Sampling procedure

Oracle calls. One oracle call is one evaluation of the transferred velocity field $b_t(x)$. With classifier-free guidance this is a single forward pass of SiT-XL/2 on a batch that stacks the conditional and unconditional inputs, and we count it as one call.

Noise, labels, encoding and decoding. Each sample i uses a noise vector $z^{(i)}$ and a class label $y^{(i)}$ drawn uniformly from the 1000 ImageNet classes. Both are generated deterministically from a fixed seed (0) and the sample index, so every schedule, solver, and oracle budget is evaluated on identical noise and labels.

SiT operates in the latent space of the Stable Diffusion VAE (Rombach et al., 2022). The encoder maps a $256 \times 256 \times 3$ RGB image, scaled to $[-1, 1]$, to a Gaussian posterior over $4 \times 32 \times 32$ latents. A sample from this posterior, multiplied by the scale factor 0.18215, gives a data latent x_1. This is how SiT's training data were produced. Sampling never uses the encoder, because the ODE is solved entirely in latent space starting from noise. To obtain images, we divide the final latents by 0.18215, decode them with the `sd-vae-ft-ema` decoder[2], clip the result to $[-1, 1]$, and quantize it to 8-bit 256×256 RGB images.

Compute. Sampling was run on NVIDIA L40S and H200 GPUs with TF32 matrix multiplications enabled and a batch size of 64.

A.4.2 Target-rescaled elliptic experiments

We implement the target-rescaled elliptic schedule with parameters $(\ell, \kappa_{\mathrm{param}})$ as described in Appendix G.1. We build the elliptic schedule $(\alpha_t^{ell}, \beta_t^{ell})$ for condition number κ_{param}, run the sampler with $(\alpha_t^{ell}, \sqrt{\ell}\,\beta_t^{ell})$, and divide the final state by $\sqrt{\ell}$ before decoding. Since $\alpha_t^{ell} z + \sqrt{\ell}\,\beta_t^{ell} x_1 = \alpha_t^{ell} z + \beta_t^{ell}(\sqrt{\ell}\,x_1)$, this is the elliptic schedule applied to the rescaled target $\sqrt{\ell}\,x_1$. For a Gaussian target, this rescaling maps a precision spectrum in $[\ell, \ell\kappa_{\mathrm{param}}]$ to $[1, \kappa_{\mathrm{param}}]$. At $\kappa_{\mathrm{param}} = 1$ the elliptic schedule reduces exactly to the trigonometric schedule $(\cos\frac{\pi t}{2}, \sin\frac{\pi t}{2})$, so the $\kappa_{\mathrm{param}} = 1$ slice of our sweep is the target-rescaled trigonometric baseline.

We sweep $(\ell, \kappa_{\mathrm{param}})$ over a log-spaced grid at budgets of 4, 8 and 32 oracle calls, sampling with DPM-Solver++(2M). We score each configuration by the Fréchet Inception on 10,000 samples (FID-10K). The FID-10K is computed with the `pytorch-fid` implementation Seitzer (2020).

[2] `https://huggingface.co/stabilityai/sd-vae-ft-ema`

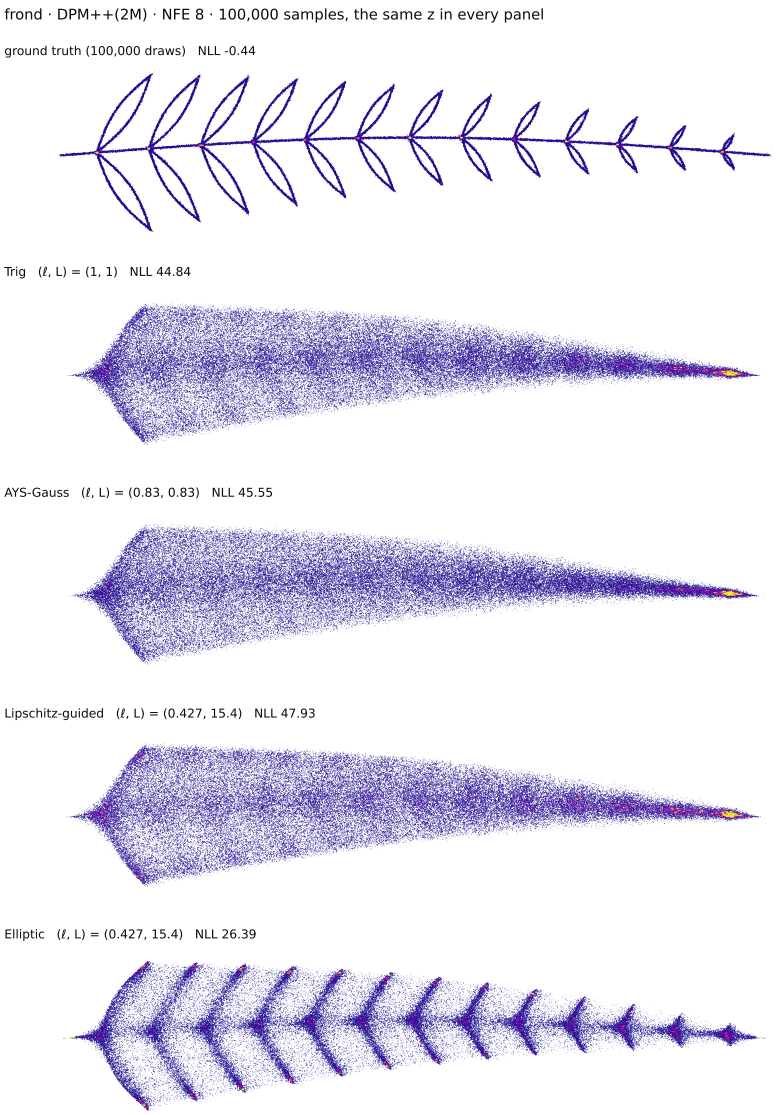

Figure 9: Frond target, DPM-Solver++(2M) with $r = 8$ score evaluations. Rows, top to bottom: exact samples, Trigonometric, AYS-Gauss, Lipschitz-guided, Elliptic.

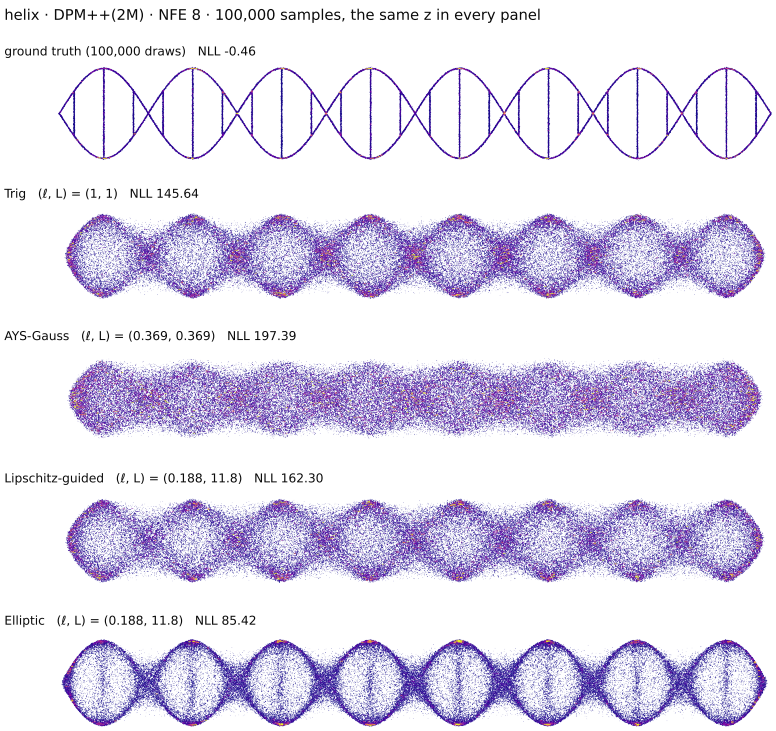

Figure 10: Helix target, DPM-Solver++(2M) with $r = 8$ score evaluations. Rows, top to bottom: exact samples, Trig, AYS-Gauss, Lipschitz-guided, Elliptic.

Figure 15 shows the results. At low budgets the choice of schedule matters a great deal, but the differences shrink as the number of oracle calls grows. Since schedules differ in discretization error, their differences vanish as the number of steps in a solver increase.

The dashed lines mark the heuristic estimates $\hat{\ell} = 0.0946$ and $\hat{\kappa}_{\mathrm{param}} = 187$. These are obtained from the percentiles of the latent covariance spectrum marked in Figure 14, following Appendix H. Without any tuning, the heuristic configuration reaches an FID of 43.67, 15.40 and 9.00 at 4, 8 and 32 calls, compared with 31.57, 10.27 and 8.59 for the best configuration in the sweep. It does not match the tuned optimum, but it lands near the low-FID region of the grid. That makes it a sensible default and a good center for a narrower sweep.

From Figure 15, we take the best target-rescaled elliptic and trigonometric configurations for each number of oracle calls. For these schedules and the untuned linear and trigonometric schedules, we generate $50{,}000$ samples from each with DPM-Solver++(2M). We compute FID and IS with the ADM evaluation suite (Dhariwal & Nichol, 2021)[3]. The results are in Table 1. Since the elliptic family contains the trigonometric one at $\kappa_{\mathrm{param}} = 1$, the two coincide at 32 calls.

B Gaussian Diagonalization and Eigenmodes

When both the initial and target distributions are centered Gaussians, the analysis of the SDEs and ODEs simplifies significantly, since the dynamics can be decoupled by working on an eigenbasis. We describe the details below.

Definition B.1 (Gaussian eigenmodes)**.** Let $\pi = \mathcal{N}(0, \Lambda^{-1})$, where $\Lambda \succ 0$, and fix an orthogonal eigendecomposition

$$\Lambda = U \operatorname{diag}(\lambda_1, \dots, \lambda_d) U^\top, \qquad U = [u_1, \dots, u_d], \qquad U^\top U = I_d.$$

[3] https://github.com/openai/guided-diffusion/tree/main/evaluations

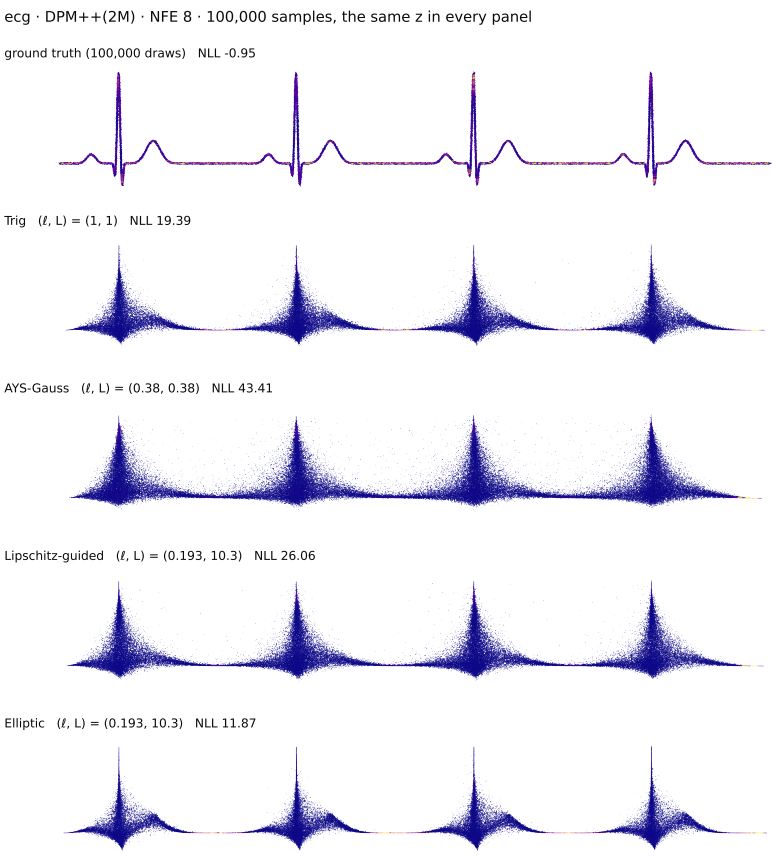

Figure 11: ECG target, DPM-Solver++(2M) with $r = 8$ score evaluations. Rows, top to bottom: exact samples, Trig, AYS-Gauss, Lipschitz-guided, Elliptic.

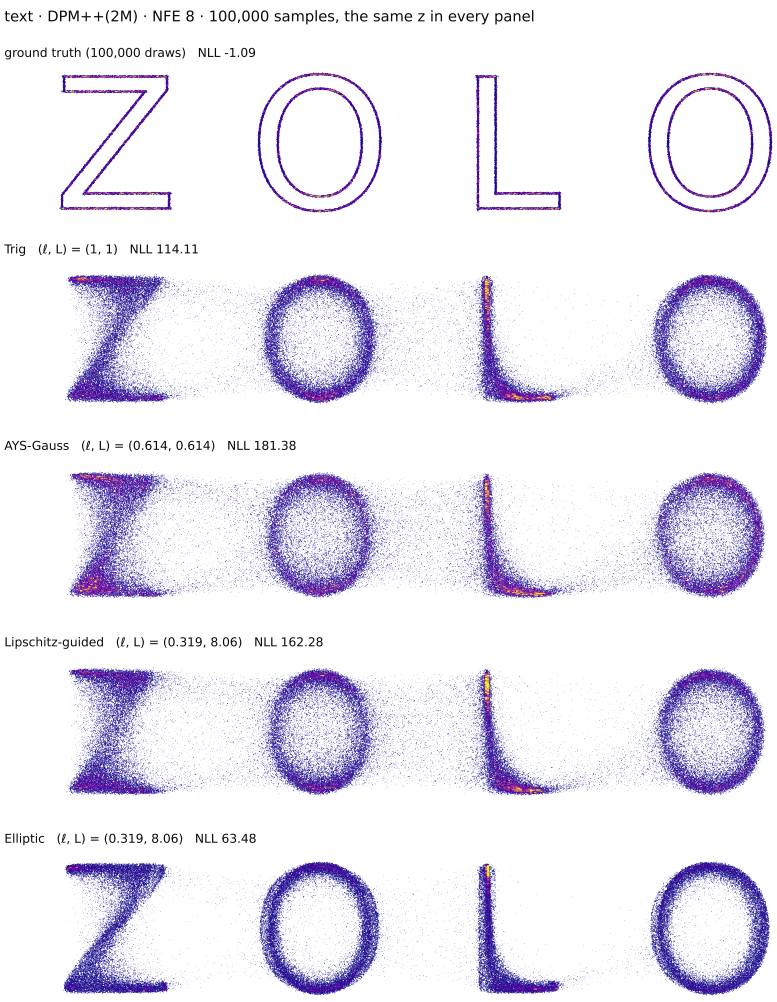

Figure 12: "ZOLO" text target, DPM-Solver++(2M) with $r = 8$ score evaluations. Rows, top to bottom: exact samples, Trig, AYS-Gauss, Lipschitz-guided, Elliptic.

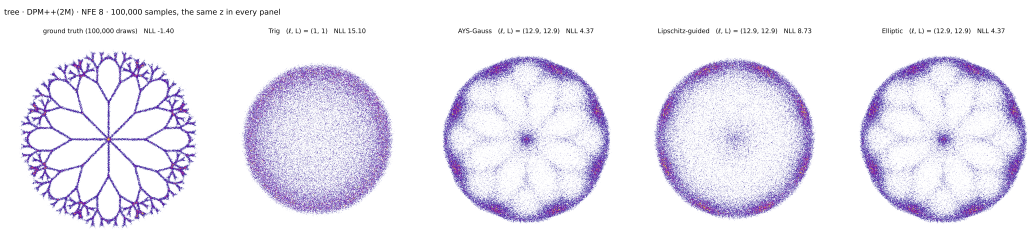

Figure 13: Radial tree target (isotropic covariance, so Elliptic coincides with AYS-Gauss), DPM-Solver++(2M) with $r = 8$ score evaluations. Panels, left to right: exact samples, Trig, AYS-Gauss, Lipschitz-guided, Elliptic.

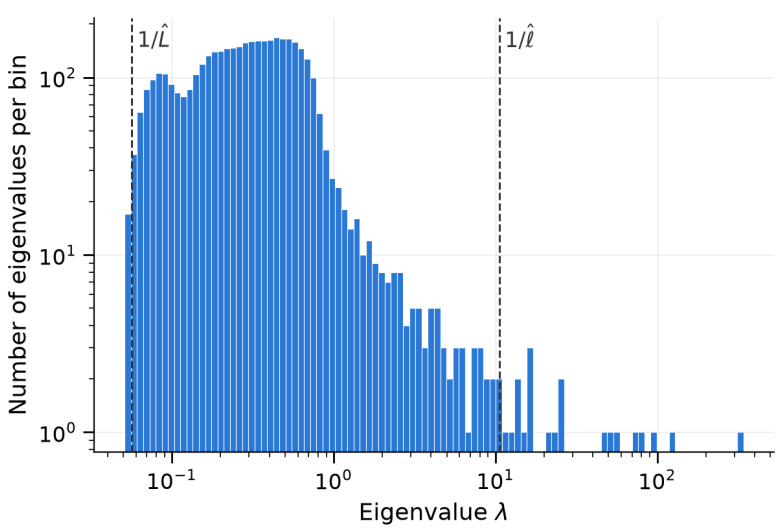

Figure 14: **Eigenvalue spectrum of the SiT training-data covariance.** Histogram of the eigenvalues λ of the empirical covariance $\hat{\Sigma}$ of the latents on which SiT is trained, using 100 bins of equal width in $\log_{10}\lambda$. We estimate $\hat{\Sigma}$ (unbiased) from $N = 50{,}000$ ImageNet 256×256 training images passed through the exact SiT training pipeline, so each sample is identical in distribution to what the network sees during training. Dashed lines mark $1/\hat{L} = 0.0566$ and $1/\hat{\ell} = 10.57$, which are the 0.5th and 99.5th percentiles of the spectrum, respectively.

For each $k \in \{1, \dots, d\}$, the *k-th eigenmode* of π is the line $\mathrm{span}\{u_k\}$, whose precision is λ_k. Any $x \in \mathbb{R}^d$ can be decomposed into eigenmodes as

$$x^{(k)} \triangleq u_k u_k^\top x, \qquad x = \sum_{k=1}^{d} x^{(k)}.$$

Given a scalar function f defined on the spectrum $\mathrm{spec}(\Lambda)$, there is a natural matrix extension given by $f(\Lambda) \triangleq U\,\mathrm{diag}\big(f(\lambda_1), \dots, f(\lambda_d)\big)U^\top$. As a linear map, $f(\Lambda)$ acts modewise, since

$$f(\Lambda)x = \sum_{k=1}^{d} f(\lambda_k)x^{(k)}.$$

Working along an eigenmode means projecting onto its eigendirection, so that multiplication by $f(\Lambda)$ becomes scalar multiplication by $f(\lambda_k)$. If the specific mode λ_k is understood from the context, we may write $\lambda = \lambda_k$, $x_\lambda = x^{(k)}$ and λ-eigenmode for short.

Gaussian scores and drifts. For a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$, the interpolant (1) has law $p_t = \mathcal{N}(0, \Sigma_t)$, where $\Sigma_t = \alpha_t^2 I_d + \beta_t^2 \Lambda^{-1}$. Its score and probability-flow drift are

$$s_t(x) = -\Sigma_t^{-1}x, \qquad b_t(x) = \big(\alpha_t\dot{\alpha}_t I_d + \beta_t\dot{\beta}_t\Lambda^{-1}\big)\Sigma_t^{-1}x, \tag{32}$$

and since the matrices commute, both maps preserve the eigenmodes. In the notation of Definition B.1,

$$s_t(x_\lambda) = -\frac{\lambda}{\lambda\alpha_t^2 + \beta_t^2}\,x_\lambda, \qquad b_t(x_\lambda) = g_\lambda(t)x_\lambda, \tag{33}$$

where the *eigenmode velocity coefficient* is

$$g_\lambda(t) \triangleq \frac{\lambda\alpha_t\dot{\alpha}_t + \beta_t\dot{\beta}_t}{\lambda\alpha_t^2 + \beta_t^2} = \frac{1}{2}\frac{d}{dt}\log\big(\lambda\alpha_t^2 + \beta_t^2\big)\,. \tag{34}$$

Consequently, each component of the probability-flow solution evolves independently according to $\dot{X}_\lambda(t) = g_\lambda(t)X_\lambda(t)$. For variance-preserving schedules, these expressions simplify to $g_\lambda(t) = -(\lambda - 1)\beta_t\dot{\beta}_t/(\lambda\alpha_t^2 + \beta_t^2)$ and $b_t(x_\lambda) = (\dot{\beta}_t/\beta_t)(x_\lambda + s_t(x_\lambda))$.

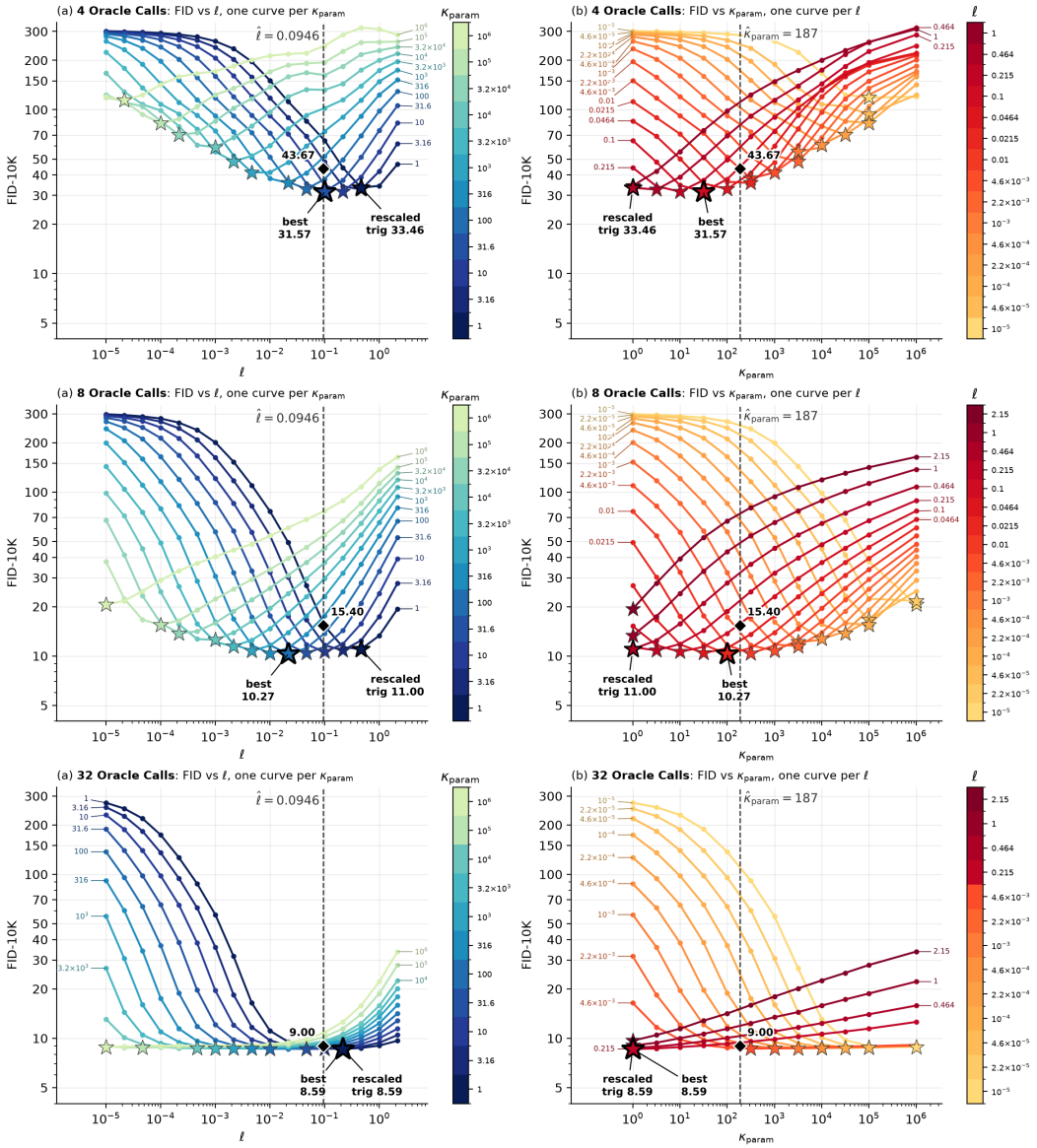

Figure 15: **FID-10K over the $(\ell, \kappa_{\text{param}})$ grid at varying amounts of oracle calls.** Target-rescaled elliptic schedule, sampled with DPM-Solver++(2M). Each row is one oracle-call budget: 4, 8 and 32 from top to bottom. (a) FID-10K vs. ℓ, one curve per κ_{param}; (b) FID-10K vs. κ_{param}, one curve per ℓ. Stars mark each curve's minimum; bold stars mark the best target-rescaled elliptic and best target-rescaled trigonometric configurations. Dashed lines mark the heuristic $\hat{\ell}$ and $\hat{\kappa}_{\text{param}}$, and the diamond is a run at $(\hat{\ell}, \hat{\kappa}_{\text{param}})$.

Gaussian sampling and the inverse square root. A natural approach to sampling from $\mathcal{N}(0, \Lambda^{-1})$ is to consider $\Lambda^{-\frac{1}{2}} z$, where $z \sim \mathcal{N}(0, I_d)$ is a standard Gaussian. By the arguments above, this leads to consideration of the scalar function $\lambda \mapsto 1/\sqrt{\lambda}$, where λ is in $\mathrm{spec}(\Lambda)$. The inverse square root has a singularity (branch point) at the origin, but it can be well-approximated (e.g., by polynomials or rational functions) on compact intervals on the positive real axis. We are particularly interested in optimal rational approximations (see e.g. Higham (2008)); see Section D.

C JACOBI ELLIPTIC FUNCTIONS

The Jacobi elliptic functions are elliptic generalizations of the trigonometric and hyperbolic functions. Like all elliptic functions, they are meromorphic and doubly periodic on the complex plane. On the real line, they have a natural interpretation as sines and cosines under a nonlinear reparameterization of angle, given by the *Jacobi amplitude* (inverse of the elliptic integral of the first kind). The Jacobi elliptic functions appear throughout mathematics and physics, particularly in classical mechanics (e.g., pendulum, rigid body rotations); see e.g. Lawden (2013) and references therein. They have been implemented in almost every computer language – in particular, in Python they can be computed with SciPy via `scipy.special.ellipj`, which simultaneously provides the functions $\mathrm{cn}, \mathrm{sn}, \mathrm{dn}$ and the Jacobi amplitude am.

C.1 BASIC NOTIONS

C.1.1 ELLIPTIC FUNCTIONS IN TERMS OF SINE AND COSINE

At their simplest level the Jacobi elliptic functions are an extension of the usual trigonometric functions, using the geometry of the ellipse. They are indexed by a *parameter*[4] $m \in [0, 1)$. For $0 \le m < 1$, the *incomplete elliptic integral of the first kind* is

$$F(\varphi \mid m) \triangleq \int_0^{\varphi} \frac{\mathrm{d}\theta}{\sqrt{1 - m \sin^2 \theta}}, \qquad K(m) \triangleq F(\pi/2 \mid m). \tag{35}$$

The *incomplete* integral $F : \varphi \mapsto u$ maps a variable angle φ on a circle to elliptic time u. The *complete* integral $K(m)$ is the time for a quarter-period; a full period takes $4K(m)$. Its complementary quarter-period is $K(1 - m)$. When $m = 0$, then $K(m) = \pi/2$, which is exactly the quarter period of trigonometric functions.

For every fixed m, since F is strictly increasing for real φ, its real inverse is well-defined: the *Jacobi amplitude*.

$$u = F(\varphi \mid m) \quad \Longleftrightarrow \quad \varphi = \mathrm{am}(u \mid m). \tag{36}$$

Here u is an elliptic time coordinate, while φ is an ordinary angular parameter. In particular, the amplitude is an angle, not the 'y'-coordinate of a curve. The main three Jacobi elliptic functions are defined as

$$\boxed{\mathrm{sn}(u \mid m) = \sin \varphi, \qquad \mathrm{cn}(u \mid m) = \cos \varphi, \qquad \mathrm{dn}(u \mid m) = \sqrt{1 - m \sin^2 \varphi}.} \tag{37}$$

where $\varphi \triangleq \mathrm{am}(u \mid m)$.

The square root is positive for real u. Differentiating the inverse relation gives

$$\frac{\mathrm{d}u}{\mathrm{d}\varphi} = \frac{1}{\sqrt{1 - m \sin^2 \varphi}}, \qquad \frac{\mathrm{d}\varphi}{\mathrm{d}u} = \mathrm{dn}(u \mid m) \in [\sqrt{1 - m}, 1]. \tag{38}$$

Thus the elliptic clock u stretches angular intervals most near $\varphi = \pi/2 \pmod{\pi}$. A trajectory moving uniformly in u advances slowly through those angles. When $m = 0$, there is no stretching: $u = \varphi$. Importantly also dn is exactly the angular speed $\frac{d\varphi}{du}$ of the angle as a function of the elliptic clock u. Below we give some plots of u and φ for varying parameters m. In particular note how for larger m close to the quarter period, the angle φ moves much slower relative to the elliptic clock u. Intuitively, for very flat ellipses, this ensures that movement towards the end of the quarter-period (around the top of the ellipse) is not too fast.

[4] There are two different conventions for Jacobi elliptic functions, either using the parameter m or the *elliptic modulus* $k = \sqrt{m}$. We use the former notation since it more common in modern software, and provides slightly cleaner notation.

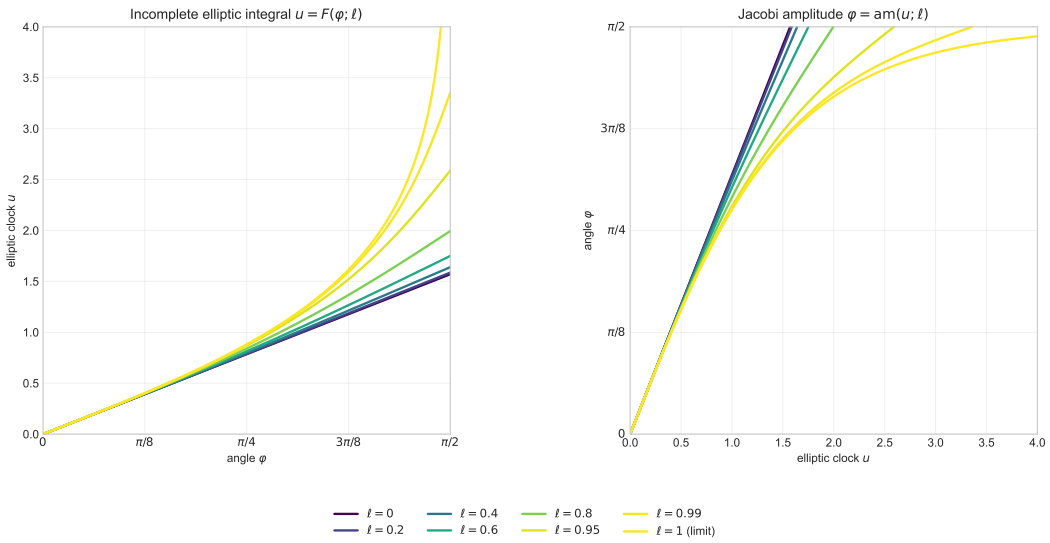

Figure 16: Angular and elliptic clocks (φ, u) for varying parameters m

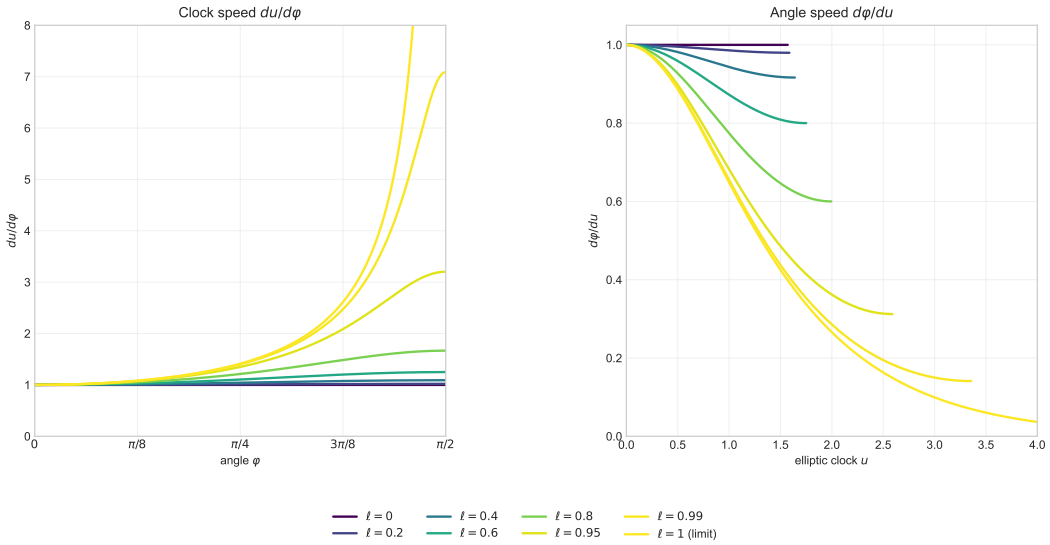

Figure 17: Angular and elliptic clock speeds (φ, u) for varying parameters m

C.1.2 LIMITING CASES $m \to 0,\ m \to 1$

At the $m = \{0, 1\}$ endpoints the Jacobi amplitude function $\mathrm{am}(u \mid m)$ evaluates to

$$\mathrm{am}(u \mid 0) = u, \text{ and } \mathrm{am}(u \mid 1) = \arcsin(\tanh u) \tag{39}$$

allowing us to reduce the Jacobi functions at these parameter endpoints in the corresponding table. Here $\tanh$ and sech denote the hyperbolic tangent and hyperbolic secant.

Function	$m = 0$	$m \to 1$
$\varphi(u) = \mathrm{am}\, u$	u	$\arcsin(\tanh u)$
$\mathrm{sn}\, u = \sin(\mathrm{am}\, u)$	$\sin u$	$\tanh u$
$\mathrm{cn}\, u = \cos(\mathrm{am}\, u)$	$\cos u$	$\mathrm{sech}\, u$
$\mathrm{dn}\, u = \sqrt{1 - m \sin^2(\mathrm{am}\, u)}$	1	$\mathrm{sech}\, u$

Table 3: Limiting cases of the Jacobi elliptic functions at $m = 0$ and $m = 1$.

We show the standard Jacobi elliptic functions for varying active parameter m along the quarter period. As the quarter period has infinite limit $\lim_{m \to 1} K(m) = \infty$, we plot these functions in both the real and normalized elliptic time u, and $t = u/K(m)$, making explicit these endpoint conditions. With the normalized clock parameter $t = u/K(m)$, these functions $\mathrm{cn}(t),\ \mathrm{sn}(t)$ are exactly our proposed elliptic schedules $\alpha_t,\ \beta_t$.

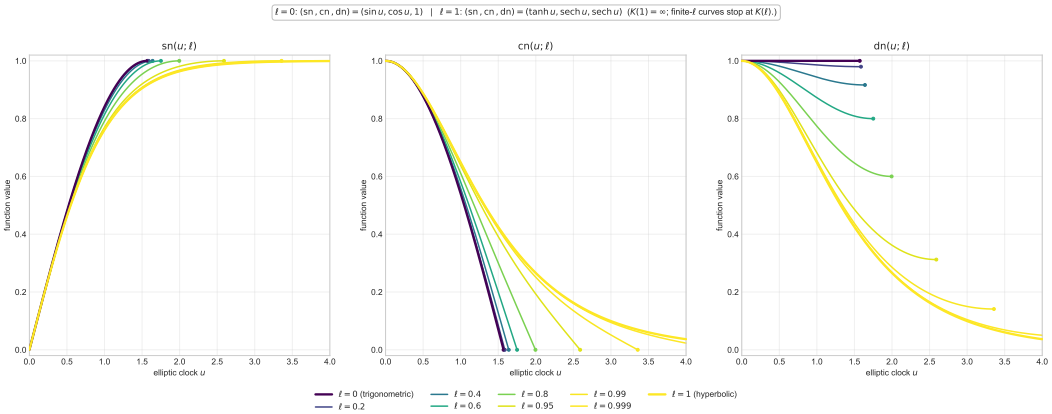

Figure 18: Elementary Jacobi Elliptic functions under raw clock u

C.1.3 THE ELLIPTIC FAMILY AND ITS ALGEBRAIC IDENTITIES

There are other Jacobi elliptic functions, which are ratios or reciprocals of the principal three. For example,

$$\mathrm{sc} = \frac{\mathrm{sn}}{\mathrm{cn}} = \tan\varphi, \quad \mathrm{cs} = \frac{\mathrm{cn}}{\mathrm{sn}} = \cot\varphi, \quad \mathrm{sd} = \frac{\mathrm{sn}}{\mathrm{dn}}, \quad \mathrm{cd} = \frac{\mathrm{cn}}{\mathrm{dn}}, \quad \mathrm{ns} = \frac{1}{\mathrm{sn}}. \tag{40}$$

There are twelve functions in total. The mnemonic is to regard the letters s, c, d, n as $\mathrm{sn}, \mathrm{cn}, \mathrm{dn}, 1$ and to interpret two distinct letters as a quotient. Their basic identities are

$$\boxed{\mathrm{sn}^2 + \mathrm{cn}^2 = 1, \qquad \mathrm{dn}^2 + m\,\mathrm{sn}^2 = 1, \qquad \mathrm{dn}^2 = (1 - m) + m\,\mathrm{cn}^2\,.} \tag{41}$$

Arguments $(u \mid m)$ are suppressed here and below. These are meromorphic identities, valid wherever the displayed expressions are finite.

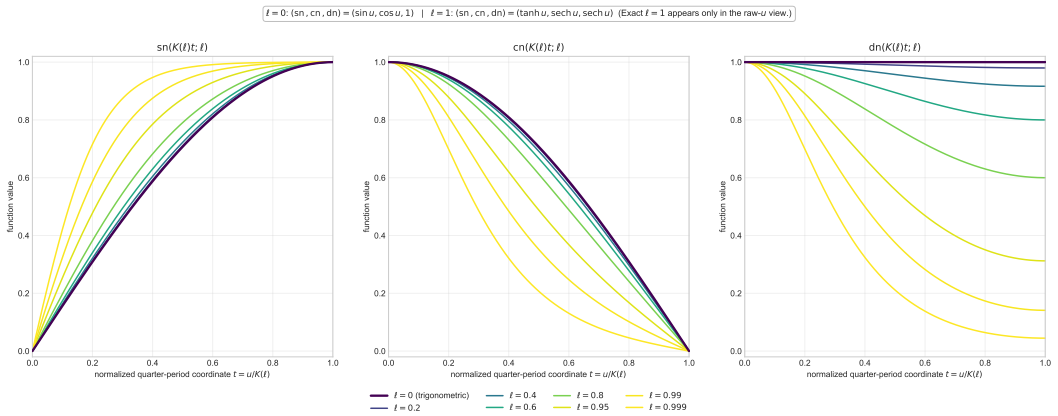

Figure 19: Elementary Jacobi Elliptic functions under normalized clock $t = u/K(m)$

C.2 Elliptic functions as differential equations

It is often common, and perhaps more natural to define the Jacobi elliptic functions through a system of first-order differential equations. Indeed $\mathrm{sn}, \mathrm{cn}, \mathrm{dn}$ may be defined as the unique solutions to the following initial value problem.

$$\begin{aligned} \dot{\mathrm{sn}} &= \mathrm{cn}\ \mathrm{dn}, & \mathrm{sn}(0) &= 0 \\ \dot{\mathrm{cn}} &= -\mathrm{sn}\ \mathrm{dn}, & \mathrm{cn}(0) &= 1 \\ \dot{\mathrm{dn}} &= -m\,\mathrm{sn}\ \mathrm{cn}, & \mathrm{dn}(0) &= 1 \end{aligned} \tag{42}$$

where the arguments $(u \mid m)$ are suppressed and dots denote u-derivatives. A nice consequence of this interpretation is one gets for free the algebraic identities (41).

C.3 Complex structure: poles, periods, and a torus

C.3.1 Complex extensions of elliptic functions and plots

With active quarter-period $K(m)$ and complementary quarter-period $K(1-m)$, generating periods and representative poles are

Function	Generators of its period lattice	Example pole
$\mathrm{sn}(u \mid m)$	$4K(m),\ 2iK(1-m)$	$iK(1-m)$
$\mathrm{cn}(u \mid m)$	$4K(m),\ 2K(m)+2iK(1-m)$	$iK(1-m)$
$\mathrm{dn}(u \mid m)$	$2K(m),\ 4iK(1-m)$	$iK(1-m)$

For example,

$$\mathrm{sn}(u + 4K(m) \mid m) = \mathrm{sn}(u \mid m), \qquad \mathrm{sn}(u + 2iK(1-m) \mid m) = \mathrm{sn}(u \mid m). \tag{43}$$

The principal three share the simple-pole locations

$$u = 2jK(m) + (2n+1)iK(1-m), \qquad j, n \in \mathbb{Z}. \tag{44}$$

Consequently, all three are holomorphic in the horizontal strip $|\operatorname{Im} u| < K(1-m)$. The functions have different period lattices: in particular, $2K(m)$ is a *sign-changing shift*, not a period, of sn and cn.

In normalized $t = \frac{u}{K(m)}$ coordinates. As in the paper, it may be more natural to visualize what happens. In 20 we plot $\mathrm{sn}(K(m)t \mid m)$, $\mathrm{cn}(K(m)t \mid m)$ for complex valued $t \in \mathbb{C}$ for $m = 0.8^2 = 0.64$. Again, the simple-pole locations are shared, and now both $\mathrm{sn}(K(m)t \mid m)$ and

$\mathrm{cn}(K(m)t \mid m)$ are holomorphic on the strip $|\operatorname{Im} t| < \frac{K(1-m)}{K(m)}$ sharing poles at

$$t = 2j + (2n+1)i\frac{K(1-m)}{K(m)}, \qquad j, n \in \mathbb{Z}. \tag{45}$$

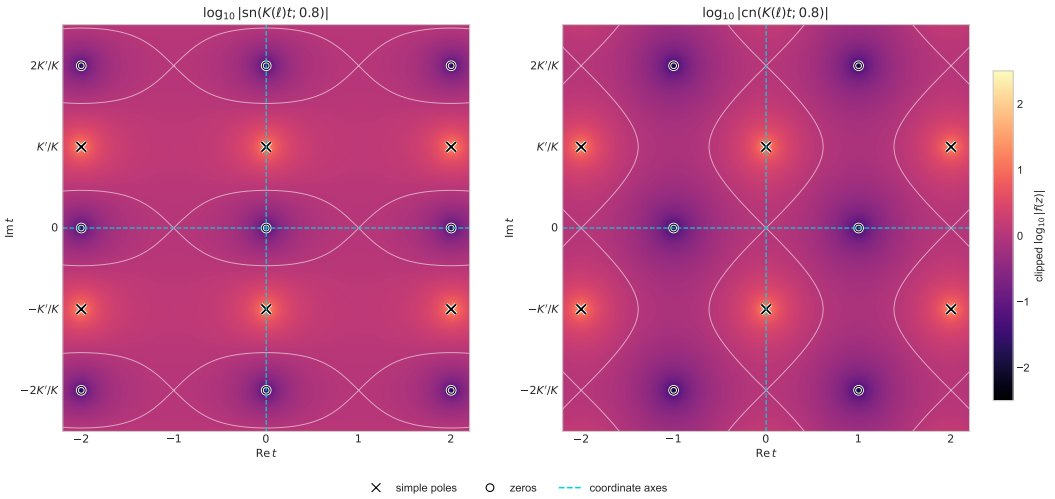

Figure 20: plots of holomorphic extensions to $|\,\mathrm{sn}\,|$ and $|\,\mathrm{cn}\,|$ in normalized time $t = \frac{u}{K(m)}$. Zeros are labeled with "$\circ$" and poles are labeled with "$\times$." The white lines represent $|f(z)| = 1$ level sets. Lighter indicates $|f(z)|$ large and darker indicates $|f(z)|$ small.

Paper notation

$$\begin{aligned} m &\triangleq 1 - \kappa^{-1}, \\ 1 - m &= \kappa^{-1}, \\ \alpha_t = \mathrm{cn}(K(m)t \mid m), &\quad \beta_t = \mathrm{sn}(K(m)t \mid m). \end{aligned} \tag{46}$$

The active parameter in these functions is $m = 1 - \kappa^{-1}$. Thus the associated ellipse has eccentricity $\sqrt{m}$ and semiaxis ratio $\sqrt{1-m}$. Equivalently, the schedule is a cosine–sine (trig) schedule with a nonlinear angular clock:

$$\varphi_t = \mathrm{am}(K(m)t \mid m), \qquad (\alpha_t, \beta_t) = (\cos\varphi_t, \sin\varphi_t), \qquad \dot{\varphi}_t = K(m)\,\mathrm{dn}(K(m)t \mid m). \tag{47}$$

It follows directly that $(\alpha_0, \beta_0) = (1, 0)$, $(\alpha_1, \beta_1) = (0, 1)$, and $\alpha_t^2 + \beta_t^2 = 1$. For $\kappa = 1$, $m = 0$ and the schedule is exactly $(\cos(\pi t/2), \sin(\pi t/2))$.

Its signal-to-noise ratio and the positive Zolotarev shift parameters are generated by the same function:

$$\frac{\beta_t^2}{\alpha_t^2} = \mathrm{sc}^2(K(m)t \mid m), \qquad \gamma_i = \mathrm{sc}^2\left(\frac{iK(m)}{2r+1} \mid m\right), \quad i = 1, \ldots, 2r. \tag{48}$$

In the inverse-square-root product, $-\gamma_{2j}$ and $-\gamma_{2j-1}$ are the zero and pole locations, respectively. The γ_i are *positive shifts*, not the locations themselves Nakatsukasa & Freund (2016).

D ZOLOTAREV RATIONAL FUNCTION APPROXIMATION RESULTS

Let $\mathcal{R}_{p,q}$ denote the real rational functions whose numerator and denominator have degrees at most p and q, respectively. We would like to approximate the function $1/\sqrt{\lambda}$ on the interval $[1, \kappa]$, with an element of $\mathcal{R}_{p,q}$. For $\kappa > 1$ and $r \in \mathbb{N}_0$, define the optimal relative approximation error

$$E_{\kappa,r} \triangleq \min_{q \in \mathcal{R}_{r,r}} \max_{\lambda \in [1,\kappa]} \left|\sqrt{\lambda}\, q(\lambda) - 1\right|, \tag{49}$$

where the minimization is restricted to functions without poles on $[1, \kappa]$. The unique minimizer, denoted by $q_{\kappa,r}$, is the *Zolotarev inverse-square-root approximant*. It is classically known (Zolotarev 1877) that there is an explicit expression for $q_{\kappa,r}$ in terms of Jacobi elliptic functions. To present this, we first discuss optimal rational approximations to the sign function.

Sign function. Zolotarev's theorem gives the best type-$(2r+1, 2r)$ rational approximation $Z(x)$ to $\operatorname{sign}(x)$ on $[-\sqrt{\kappa}, -1] \cup [1, \sqrt{\kappa}]$. Since $\operatorname{sign}(x)$ is odd, so is the optimal approximant, and thus it has the form

$$Z(x) = Cx\frac{P(x^2)}{Q(x^2)} = xq_{\kappa,r}(x^2), \qquad \deg P, \deg Q \le r.$$

The substitution $\lambda = x^2$ gives the exact error identity

$$\max_{1\le|x|\le\sqrt{\kappa}} |Z(x) - \operatorname{sign}(x)| = \max_{\lambda\in[1,\kappa]} \left|\sqrt{\lambda}\, q_{\kappa,r}(\lambda) - 1\right| = E_{\kappa,r}. \tag{50}$$

Thus *absolute-error* approximation of the sign function is equivalent to *relative-error* approximation of the inverse square root, and achieve the same optimal error. Rescaling x by $\sqrt{\kappa}$ gives the usual sign-approximation domain $[-1, -\sqrt{1-m}] \cup [\sqrt{1-m}, 1]$, where $m \triangleq 1 - \kappa^{-1}$; see Nakatsukasa & Freund (2016, Section 3.1). More generally, on any positive interval $[a, b]$, the approximant $a^{-1/2} q_{b/a,r}(\lambda/a)$ has optimal relative error $E_{b/a,r}$. Thus, the ratio $\kappa = b/a$ of the interval endpoints directly controls the optimal approximation error.

Zolotarev approximant and the elliptic grid. The Zolotarev optimal approximant $q_{\kappa,r}$ can be explicitly described using Jacobi elliptic functions. To this end, define the *elliptic parameter*

$$m \triangleq 1 - \kappa^{-1} \tag{51}$$

and notice that $m \in [0, 1)$. Let $K \triangleq K(m)$, where $K(\cdot)$ is the complete elliptic integral of the first kind. Define

$$u_i = \frac{i}{2r+1}, \qquad \gamma_i = \operatorname{sc}^2(Ku_i \mid m) = \frac{\operatorname{sn}^2(Ku_i \mid m)}{\operatorname{cn}^2(Ku_i \mid m)}, \qquad i = 1, \dots, 2r. \tag{52}$$

Notice that the u_i describe a uniform grid on $[0, 1]$ (excluding the endpoints), and the γ_i evaluate an elliptic function on this grid. Background on Jacobi elliptic functions is given in Appendix C.

The Zolotarev approximant is

$$q_{\kappa,r}(\lambda) = C_{\kappa,r} \prod_{j=1}^{r} \frac{\lambda + \gamma_{2j}}{\lambda + \gamma_{2j-1}}, \qquad q_{\kappa,r}(1) + \sqrt{\kappa}\, q_{\kappa,r}(\kappa) = 2, \tag{53}$$

where the endpoint condition uniquely determines $C_{\kappa,r} > 0$ (Nakatsukasa & Freund, 2016, Section 3.1). The zero and pole shifts are $\xi_j \triangleq \gamma_{2j}$ and $\zeta_j \triangleq \gamma_{2j-1}$, respectively; the actual zeros and poles are on the negative real axis, given by $-\xi_j$ and $-\zeta_j$. Thus they are generated by evaluating sc^2 at alternating even and odd points of the same uniform u-grid. In particular, it is the "elliptic times" Ku_i, not the shifts γ_i, that are uniformly spaced.

The relative error equioscillates between $\pm E_{\kappa,r}$, with endpoint values

$$q_{\kappa,r}(1) = 1 - E_{\kappa,r}, \qquad \sqrt{\kappa}\, q_{\kappa,r}(\kappa) = 1 + E_{\kappa,r}. \tag{54}$$

Endpoint normalization. For our sampling construction, we use a slightly modified approximant that is asymptotically equivalent. It differs only by a constant factor, since we will require the approximation to be exact at $\lambda = 1$. Concretely, we use

$$\widehat{q}_{\kappa,r}(\lambda) \triangleq \frac{q_{\kappa,r}(\lambda)}{q_{\kappa,r}(1)} = \prod_{j=1}^{r} \frac{1 + \gamma_{2j-1}}{1 + \gamma_{2j}} \frac{\lambda + \gamma_{2j}}{\lambda + \gamma_{2j-1}}. \tag{55}$$

This removes the implicit constant $C_{\kappa,r}$ and enforces $\widehat{q}_{\kappa,r}(1) = 1$, so the unit-precision eigenmode is preserved exactly. Its relative error is

$$\widehat{E}_{\kappa,r} \triangleq \max_{\lambda\in[1,\kappa]} \left|\sqrt{\lambda}\, \widehat{q}_{\kappa,r}(\lambda) - 1\right| = \frac{2E_{\kappa,r}}{1 - E_{\kappa,r}}. \tag{56}$$

Indeed, $\sqrt{\lambda}\, \widehat{q}_{\kappa,r}(\lambda)$ ranges from 1 to $(1 + E_{\kappa,r})/(1 - E_{\kappa,r})$, and when the error is small, $\widehat{E}_{\kappa,r} \approx 2E_{\kappa,r}$. Although endpoint normalization sacrifices exact minimax optimality in equation 49, it preserves the asymptotic convergence rate and makes all expressions considerably simpler.

Approximation rates. Set $\rho_\kappa \triangleq \exp(\pi K(1-m)/K(m)) > 1$. The minimax error $E_{\kappa,r}$ satisfies (Beckermann & Townsend, 2017, Appendix A)

$$\frac{4\rho_\kappa^{-(2r+1)}}{(1+\rho_\kappa^{-2(2r+1)})^4} \le E_{\kappa,r} \le 4\rho_\kappa^{-(2r+1)}. \tag{57}$$

Consequently, for fixed $\kappa > 1$ as $r \to \infty$,

$$E_{\kappa,r} \sim 4\rho_\kappa^{-(2r+1)}, \qquad \widehat{E}_{\kappa,r} \sim 8\rho_\kappa^{-(2r+1)}. \tag{58}$$

Moreover, the standard elliptic-integral asymptotics give

$$\log \rho_\kappa = \frac{\pi K(1-m)}{K(m)} \sim \frac{\pi^2}{\log(16\kappa)} \qquad (\kappa \to \infty).$$

Thus, for $\kappa \ge 2$ and sufficiently small $\varepsilon > 0$, the minimum degree required to achieve either error at most ε is $\Theta(\log \kappa \, \log(1/\varepsilon))$[5]. When $\kappa = 1$, the approximation interval is a singleton and the constant function 1 achieves zero error. These results and joining proofs are articulately in greater detail in D.

D.1 Proofs

Proposition D.1 (Equivalence of sign and inverse-square-root minimax errors)**.** *Fix $\kappa > 1$ and an integer $r \ge 0$, and let $m \triangleq 1 - \kappa^{-1}$ and $\delta_\kappa \triangleq \sqrt{1-m} = \kappa^{-1/2}$. Define*

$$E_{p,q}(\delta_\kappa) \triangleq \min_{R \in \mathcal{R}_{p,q}} \max_{x \in [-1,-\delta_\kappa] \cup [\delta_\kappa, 1]} |R(x) - \operatorname{sign}(x)|, \tag{59}$$

$$E_{\kappa,r} \triangleq \min_{q \in \mathcal{R}_{r,r}} \max_{\lambda \in [1,\kappa]} |\sqrt{\lambda}\, q(\lambda) - 1|, \tag{60}$$

where the minima are restricted to real rational functions without poles on the respective approximation domains. Then

$$E_{2r+1,2r+1}(\delta_\kappa) = E_{2r+1,2r}(\delta_\kappa) = E_{\kappa,r}.$$

In particular, allowing denominator degree $2r+1$ rather than $2r$ does not improve the optimal sign approximation error.

Proof. Zolotarev's explicit solution of the sign minimax problem (Beckermann & Townsend, 2017, proof of Theorem 3.3, equation (3.6)) states that a global minimizer over $\mathcal{R}_{2r+1,2r+1}$ has the form $R(x) = xP(x^2)/Q(x^2)$ with $\deg P, \deg Q \le r$. Consequently, restricting to this form does not change either sign minimax error. By oddness,

$$E_{2r+1,2r+1}(\delta_\kappa) = E_{2r+1,2r}(\delta_\kappa) = \min_{\deg P, \deg Q \le r} \max_{x \in [\delta_\kappa, 1]} \left| x \frac{P(x^2)}{Q(x^2)} - 1 \right|,$$

where Q has no zeros on $[\delta_\kappa^2, 1]$. Now, for any admissible P, Q with degree $\le r$, set

$$\lambda \triangleq \frac{x^2}{\delta_\kappa^2}, \qquad q(\lambda) \triangleq \delta_\kappa \frac{P(\delta_\kappa^2 \lambda)}{Q(\delta_\kappa^2 \lambda)}.$$

Since $\delta_\kappa^{-2} = \kappa$, this gives

$$\max_{x \in [\delta_\kappa, 1]} \left| x \frac{P(x^2)}{Q(x^2)} - 1 \right| = \max_{\frac{x^2}{\delta_\kappa^2} \in [1, 1/\delta_\kappa^2]} \left| x \frac{P(x^2)}{Q(x^2)} - 1 \right| = \max_{\lambda \in [1,\kappa]} \left| \sqrt{\lambda}\, q(\lambda) - 1 \right|.$$

As P, Q range over the admissible polynomnials of degree at most r, the rescaled ration $q(\lambda) = \delta_\kappa P(\delta_\kappa^2 \lambda) Q(\delta_\kappa^2 \lambda)$ ranges over all admissible functions in $\mathcal{R}_{r,r}$. Hence

$$\max_{\lambda \in [1,\kappa]} \left| \sqrt{\lambda} \delta_\kappa \frac{P(\delta_\kappa^2 \lambda)}{Q(\delta_\kappa^2 \lambda)} - 1 \right| = \min_{q \in \mathcal{R}_{r,r}} \max_{\lambda \in [1,\kappa]} \left| \sqrt{\lambda} q(\lambda) - 1 \right| \tag{61}$$

$$E_{2r+1,2r+1}(\delta_\kappa) = E_{2r+1,2r}(\delta_\kappa) = E_{\kappa,r}.$$

□

[5] Ignoring constant factors, see Appendix.

Proposition D.2 (Rates of $E_{\kappa,r}$)**.** *for all $\kappa > 1, r \geq 0$, with $m \triangleq 1 - \kappa^{-1}$,*

$$\frac{1}{4}\exp\left(-\frac{\pi^2(2r+1)}{\ln(\kappa)}\right) \leq E_{\kappa,r} \leq 4\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right) \tag{62}$$

more preciscely

$$\frac{1}{4}\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right) \leq E_{\kappa,r} \leq 4\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right) \tag{63}$$

so that for all $\kappa > 1, r \geq 0$

$$E_{\kappa,r} = \Theta\left(\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right)\right) \tag{64}$$

Proof. Fix $\kappa > 1$ and $r \in \mathbb{Z}_{\geq 0}$, and write

$$N \triangleq 2r+1, \qquad m \triangleq 1-\kappa^{-1}, \qquad \delta_\kappa \triangleq \sqrt{1-m} = \kappa^{-1/2}.$$

Following Beckermann & Townsend (2017, equation (3.2)), define

$$\mu(m) \triangleq \frac{\pi K(m)}{2K(1-m)}, \qquad \rho \triangleq \exp\left(\frac{\pi^2}{2\mu(m)}\right) = \exp\left(\pi\frac{K(1-m)}{K(m)}\right) > 1.$$

By Lemma D.1, $E_{\kappa,r} = E_{N,N}(\delta_\kappa)$. Applying the sign-approximation bounds of Beckermann & Townsend (2017, Appendix A, equation (A.5)) with $a = \delta_\kappa$ and $b = 1$ therefore gives

$$\frac{4\rho^{-N}}{(1+\rho^{-2N})^4} \leq E_{\kappa,r} \leq \frac{4\rho^{-N}}{(1+\rho^{-2N})^2} \leq 4\rho^{-N}. \tag{65}$$

Since $0 < \rho^{-2N} < 1$, we have $(1+\rho^{-2N})^4 \leq 16$. Consequently,

$$\frac{1}{4}\exp\left(-\pi N\frac{K(1-m)}{K(m)}\right) \leq E_{\kappa,r} \leq 4\exp\left(-\pi N\frac{K(1-m)}{K(m)}\right). \tag{66}$$

This proves the elliptic bounds and the Θ-statement, with constants independent of both κ and r.

To obtain the elementary bounds, we use the elliptic-integral inequalities (DLMF, equation (19.9.5))

$$\ln\left(\frac{(1+m^{1/4})^2}{\sqrt{1-m}}\right) < \mu(m) < \ln\left(\frac{2(1+\sqrt{m})}{\sqrt{1-m}}\right).$$

Because $0 < m < 1$, these imply

$$\ln\left(1/\sqrt{1-m}\right) < \mu(m) < \ln\left(4/\sqrt{1-m}\right).$$

Taking reciprocals and using $1 - m = \kappa^{-1}$ yields

$$\frac{\pi}{\ln(16\kappa)} < \frac{K(1-m)}{K(m)} = \frac{\pi}{2\mu(m)} < \frac{\pi}{\ln\kappa}. \tag{67}$$

Thus

$$\exp\left(-\frac{\pi^2 N}{\ln\kappa}\right) \leq \exp\left(-\pi N\frac{K(1-m)}{K(m)}\right) \leq \exp\left(-\frac{\pi^2 N}{\ln(16\kappa)}\right).$$

Combining this with equation 66 and substituting $N = 2r+1$ proves

$$\frac{1}{4}\exp\left(-\frac{\pi^2(2r+1)}{\ln\kappa}\right) \leq E_{\kappa,r} \leq 4\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right).$$

□

Corollary D.3 (Asymptotic Sharpness)**.** *From 65 we get that for any fixed $\kappa > 1$, with $m \triangleq 1-\kappa^{-1}$, the elliptic upper bound is asymptotically exact, i.e.*

$$E_{\kappa,r} \sim 4\rho^{-2r+1}, \quad r \to \infty \tag{68}$$

moreover from 67 we have

$$\ln\rho = \pi\frac{K(1-m)}{K(m)} \sim \frac{\pi^2}{\ln 16\kappa}, \quad \kappa \to \infty \tag{69}$$

It will also be helpful to analyze the error associated with the normalized zolotarev rational approximation to $\lambda^{-1/2}$. For

$$\hat{q}_{\kappa,r}(\lambda) \triangleq \frac{\bar{q}_{\kappa,r}(\lambda)}{\bar{q}_{\kappa,r}(1)} \tag{70}$$

we define

$$\hat{E}_{\kappa,r} \triangleq \max_{\lambda\in[1,\kappa]} \left|\lambda^{1/2}\hat{q}_{\kappa,r}(\lambda) - 1\right| = \frac{2E_{\kappa,r}}{1-E_{\kappa,r}} \tag{71}$$

Corollary D.4 (Rates of $\hat{E}_{\kappa,r}$)**.** *Let $m \triangleq 1-\kappa^{-1}$. Whenever*

$$\pi(2r+1)\frac{K(1-m)}{K(m)} \geq \ln 8 \approx 2.08^{6} \tag{72}$$

the previous results tell us that

$$\frac{1}{2}\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right) \leq \hat{E}_{\kappa,r} \leq 16\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right) \tag{73}$$

so that

$$\hat{E}_{\kappa,r} = \Theta\left(\exp\left[-\pi(2r+1)\frac{K(1-m)}{K(m)}\right]\right) \tag{74}$$

as a consequence, using the bounds on $\frac{K(1-m)}{K(m)}$ we get that

$$\frac{1}{2}\exp\left(-\frac{\pi^2(2r+1)}{\ln(\kappa)}\right) \leq \hat{E}_{\kappa,r} \leq 16\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right) \tag{75}$$

The two lower bounds remain valid for every $\kappa > 1$ and $r \in \mathbb{Z}_{\geq 0}$, without the displayed condition.

Proof. Fix $\kappa > 1$ and $r \in \mathbb{Z}_{\geq 0}$, and set

$$N \triangleq 2r+1, \qquad \eta \triangleq \rho^{-N} = \exp\left(-\pi N\frac{K(1-m)}{K(m)}\right).$$

By equation 66,

$$\frac{1}{4}\eta \leq E_{\kappa,r} \leq 4\eta.$$

First, $0 \leq E_{\kappa,r} < 1$ for all such parameters. Indeed, the constant approximant $q(\lambda) = 2/(1+\sqrt{\kappa})$ is admissible for every $r \geq 0$, so

$$E_{\kappa,r} \leq \max_{\lambda\in[1,\kappa]}\left|\frac{2\sqrt{\lambda}}{1+\sqrt{\kappa}} - 1\right| = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1} < 1.$$

Consequently,

$$\widehat{E}_{\kappa,r} = \frac{2E_{\kappa,r}}{1-E_{\kappa,r}} \geq 2E_{\kappa,r} \geq \frac{1}{2}\eta,$$

which proves the elliptic lower bound without any additional condition.

For the upper bound, suppose that $\pi N K(1-m)/K(m) \geq \ln 8$. Then $\eta \leq 1/8$, and hence $E_{\kappa,r} \leq 4\eta \leq 1/2$. Therefore,

$$\widehat{E}_{\kappa,r} = \frac{2E_{\kappa,r}}{1-E_{\kappa,r}} \leq 4E_{\kappa,r} \leq 16\eta.$$

Together, these inequalities establish the elliptic bounds and the Θ-statement, with universal constants throughout the stated regime.

Finally, equation 67 gives

$$\exp\left(-\frac{\pi^2 N}{\ln\kappa}\right) \leq \eta \leq \exp\left(-\frac{\pi^2 N}{\ln(16\kappa)}\right).$$

[6] a simple sufficient condition is $r \geq 2\ln 16\kappa$ by directly applying the lower bound on $K(1-m)/K(m)$

Combining these inequalities with the preceding lower and upper bounds proves the elementary bounds. The elementary lower bound is again unconditional, whereas the elementary upper bound holds under $\pi NK(1-m)/K(m) \geq \ln 8$. In particular, a sufficient condition for both upper bounds is

$$\frac{\pi^2(2r+1)}{\ln(16\kappa)} \geq \ln 8.$$

□

Corollary D.5 (Accuracy requirement for endpoint normalization)**.** *For every $\kappa > 1$, $r \in \mathbb{Z}_{\geq 0}$, and $\varepsilon > 0$,*

$$\widehat{E}_{\kappa,r} \leq \varepsilon \quad \Longleftrightarrow \quad E_{\kappa,r} \leq \frac{\varepsilon}{2+\varepsilon}. \tag{76}$$

Consequently, $\widehat{E}_{\kappa,r} \leq \varepsilon$ is guaranteed whenever

$$r \geq \frac{\ln(16\kappa)}{2\pi^2} \ln\left(\frac{4(2+\varepsilon)}{\varepsilon}\right) - \frac{1}{2}, \tag{77}$$

with r chosen as a nonnegative integer.

Proof. Since $0 \leq E_{\kappa,r} < 1$, multiplying by the positive denominator yields

$$\frac{2E_{\kappa,r}}{1-E_{\kappa,r}} \leq \varepsilon \quad \Longleftrightarrow \quad (2+\varepsilon)E_{\kappa,r} \leq \varepsilon,$$

which proves equation 76. Using the elementary upper bound on $E_{\kappa,r}$, it therefore suffices that

$$4\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right) \leq \frac{\varepsilon}{2+\varepsilon}.$$

Taking logarithms and rearranging gives equation 77. □

E Zoloflow Complexity Results on Gaussians

Lemma E.1 (Zoloflow realizes the endpoint-normalized Zolotarev approximation)**.** *Fix $\kappa > 1$, an integer $r \geq 1$, and a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $I_d \preceq \Lambda \preceq \kappa I_d$. Let m, u_i and γ_i be as in equation 52, and write $\mathrm{cn}_i \triangleq \mathrm{cn}(u_i \mid m)$ and $\mathrm{sn}_i \triangleq \mathrm{sn}(u_i \mid m)$, so that $\gamma_i = \mathrm{sn}_i^2 \,/\, \mathrm{cn}_i^2$. Let $q_{\kappa,r}$ be the Zolotarev approximant (53) and $\widehat{q}_{\kappa,r}$ its endpoint normalization (55).*

Run Zoloflow (Algorithm 1) with parameters (κ, r), exact interpolant scores, and initialization $x_0 = z \sim \mathcal{N}(0, I_d)$. For each precision eigenmode λ, let $x_{\lambda,j}$ denote the corresponding component of x_j. Then, for $j = 1, \ldots, r$,

$$\begin{aligned} x_{\lambda,j} &= \frac{\lambda\,\mathrm{cn}_{2j}^2 + \mathrm{sn}_{2j}^2}{\lambda\,\mathrm{cn}_{2j-1}^2 + \mathrm{sn}_{2j-1}^2}\, x_{\lambda,j-1} \\ &= \frac{1+\gamma_{2j-1}}{1+\gamma_{2j}} \frac{\lambda+\gamma_{2j}}{\lambda+\gamma_{2j-1}}\, x_{\lambda,j-1}. \end{aligned} \tag{78}$$

Consequently,

$$x_{\lambda,r} = \widehat{q}_{\kappa,r}(\lambda) x_{\lambda,0}, \qquad x_r = \widehat{q}_{\kappa,r}(\Lambda) z. \tag{79}$$

Thus Zoloflow implements the endpoint-normalized approximant exactly, without any additional scalar normalization of its initialization or output.

Proof. Recall that Zoloflow uses $h = 2/(2r+1)$ and $t_j = (2j-1)/(2r+1)$, so

$$K(m)t_j = u_{2j-1}, \qquad K(m)(t_j + h/2) = u_{2j}.$$

For a fixed precision eigenmode $\lambda > 0$, define

$$D_i(\lambda) \triangleq \lambda\,\mathrm{cn}_i^2 + \mathrm{sn}_i^2 = \lambda + (1-\lambda)\,\mathrm{sn}_i^2,$$

where the second equality follows from $\mathrm{cn}_i^2 + \mathrm{sn}_i^2 = 1$. The Gaussian score preserves this eigenmode and satisfies

$$s_{t_j}(x_{\lambda,j-1}) = -\frac{\lambda}{D_{2j-1}(\lambda)}\, x_{\lambda,j-1}.$$

Consequently, projecting the update in Algorithm 1 onto this eigenmode as described in F

$$\begin{aligned} x_{\lambda,j} &= \left[1 + \frac{\mathrm{sn}_{2j}^2 - \mathrm{sn}_{2j-1}^2}{\mathrm{sn}_{2j-1}^2}\left(1 - \frac{\lambda}{D_{2j-1}(\lambda)}\right)\right] x_{\lambda,j-1} \\ &= \left[1 + \frac{(1-\lambda)(\mathrm{sn}_{2j}^2 - \mathrm{sn}_{2j-1}^2)}{D_{2j-1}(\lambda)}\right] x_{\lambda,j-1} \\ &= \frac{D_{2j}(\lambda)}{D_{2j-1}(\lambda)}\, x_{\lambda,j-1}. \end{aligned}$$

Here we used $D_{2j-1}(\lambda) - \lambda = (1-\lambda)\,\mathrm{sn}_{2j-1}^2$ in the second equality.

Next, since $\gamma_i = \mathrm{sn}_i^2 \,/\, \mathrm{cn}_i^2$ and $\mathrm{cn}_i^2 + \mathrm{sn}_i^2 = 1$,

$$\mathrm{cn}_i^2 = \frac{1}{1+\gamma_i}, \qquad D_i(\lambda) = \mathrm{cn}_i^2(\lambda + \gamma_i) = \frac{\lambda + \gamma_i}{1 + \gamma_i}.$$

Thus each step satisfies

$$x_{\lambda,j} = \frac{1+\gamma_{2j-1}}{1+\gamma_{2j}} \frac{\lambda + \gamma_{2j}}{\lambda + \gamma_{2j-1}}\, x_{\lambda,j-1}.$$

All divisions are valid because $0 < u_i < K(m)$ and $\lambda > 0$.

Iterating from $x_0 = z$ yields

$$\begin{aligned} x_{\lambda,r} &= \left[\prod_{j=1}^{r} \frac{1+\gamma_{2j-1}}{1+\gamma_{2j}} \frac{\lambda + \gamma_{2j}}{\lambda + \gamma_{2j-1}}\right] x_{\lambda,0} \\ &= \frac{q_{\kappa,r}(\lambda)}{q_{\kappa,r}(1)}\, x_{\lambda,0} = \widehat{q}_{\kappa,r}(\lambda) x_{\lambda,0}, \end{aligned}$$

where the constant $C_{\kappa,r}$ cancels in the ratio. Finally, writing $\Lambda = U\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)U^\top$ and reassembling the eigenmode components gives

$$x_r = U\,\mathrm{diag}(\widehat{q}_{\kappa,r}(\lambda_1), \ldots, \widehat{q}_{\kappa,r}(\lambda_d))\, U^\top z = \widehat{q}_{\kappa,r}(\Lambda) z.$$

Hence the step factors already incorporate the endpoint normalization, and no additional scaling of the initialization or output is required. □

We are now ready to prove the upper bound to 3.7

Proof of Theorem 3.7. Write $\widehat{q} \triangleq \widehat{q}_{\kappa,r}$. By Lemma E.1,

$$x_r = \widehat{q}(\Lambda) z, \qquad z \sim \mathcal{N}(0, I_d).$$

Since $\widehat{q}(\Lambda)$ is symmetric, the output law is

$$\widehat{\pi} = \mathcal{N}\big(0, \widehat{q}(\Lambda)^2\big)\,.$$

KL divergence in terms of relative errors. Let $\lambda_1, \ldots, \lambda_d \in [1, \kappa]$ be the eigenvalues of Λ, and define

$$e_k \triangleq \sqrt{\lambda_k}\, \widehat{q}(\lambda_k) - 1, \qquad k = 1, \ldots, d.$$

The endpoint-normalization identity gives

$$0 \le e_k \le \widehat{E}_{\kappa,r}.$$

In particular, $\widehat{q}(\lambda_k) > 0$, so $\widehat{q}(\Lambda)^2 \succ 0$. Moreover, the eigenvalues of $\Lambda \widehat{q}(\Lambda)^2$ are

$$\lambda_k \widehat{q}(\lambda_k)^2 = (1 + e_k)^2.$$

Taking the expectation of the Gaussian log-density ratio therefore yields

$$\begin{aligned} D_{\mathrm{KL}}(\widehat{\pi}\|\pi) &= \frac{1}{2}\left(\mathrm{tr}\big(\Lambda\widehat{q}(\Lambda)^2\big) - d - \ln\det\big(\Lambda\widehat{q}(\Lambda)^2\big)\right) \\ &= \frac{1}{2}\sum_{k=1}^{d}\left((1+e_k)^2 - 1 - 2\ln(1+e_k)\right) \\ &= \sum_{k=1}^{d}\left(e_k + \frac{e_k^2}{2} - \ln(1+e_k)\right). \end{aligned} \tag{80}$$

For every $u \geq 0$,

$$\ln(1+u) \geq u - \frac{u^2}{2}.$$

Indeed, the difference between the two sides vanishes at zero and has derivative $u^2/(1+u) \geq 0$. Applying this inequality to each e_k gives

$$D_{\mathrm{KL}}(\widehat{\pi}\|\pi) \leq \sum_{k=1}^{d} e_k^2 \leq d\,\widehat{E}_{\kappa,r}^{\,2}. \tag{81}$$

From KL divergence to total variation. By Pinsker's inequality,

$$d_{\mathrm{TV}}(\widehat{\pi},\pi) \leq \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\widehat{\pi}\|\pi)} \leq \left(\frac{1}{2}\sum_{k=1}^{d} e_k^2\right)^{1/2} \leq \sqrt{\frac{d}{2}}\,\widehat{E}_{\kappa,r}. \tag{82}$$

Applying the rational-approximation rate. Setting $m \triangleq 1-\kappa^{-1}$ and $A_{\kappa,r} \triangleq \frac{\pi^2(2r+1)}{\ln(16\kappa)}$. we have from equation 67,

$$A_{\kappa,r} \leq \pi(2r+1)\frac{K(1-m)}{K(m)}.$$

Consequently, whenever $A_{\kappa,r} \geq \ln 8$ (sufficient if $r \geq 2\ln 16\kappa$), the condition of Corollary D.4 is satisfied , and

$$\widehat{E}_{\kappa,r} \leq 16\exp\left(-\pi(2r+1)\frac{K(1-m)}{K(m)}\right) \leq 16\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right)$$

Combining this with equation 82 yields

$$d_{\mathrm{TV}}(\widehat{\pi},\pi) \leq 8\sqrt{2d}\,\exp\left(-\frac{\pi^2(2r+1)}{\ln(16\kappa)}\right). \tag{83}$$

The query condition is equivalent to

$$A_{\kappa,r} \geq \ln\left(\frac{8\sqrt{2d}}{\varepsilon}\right).$$

Because $d \geq 1$ and $\varepsilon < 1$, the right-hand side exceeds $\ln 8$. Thus the preceding bound applies and gives $d_{\mathrm{TV}}(\widehat{\pi},\pi) \leq \varepsilon$.

Finally, choose the smallest positive integer satisfying

$$r \geq \frac{\ln(16\kappa)}{2\pi^2}\ln\left(\frac{8\sqrt{2d}}{\varepsilon}\right) - \frac{1}{2}. \tag{84}$$

Each Zoloflow step uses one score query, so the number of queries is

$$O\left(\ln(16\kappa)\,\ln\left(\frac{8\sqrt{2d}}{\varepsilon}\right)\right).$$

For $\kappa \geq 2$ and $0 < \varepsilon \leq 1/2$, this is $O(\ln\kappa\,\ln(\sqrt{d}/\varepsilon))$, with a universal implicit constant. □

Next we restate the lower bound. Recall the minimax query complexity

$$r^*_{\mathrm{obl}}(d,\kappa,\varepsilon) \triangleq \min\left\{ r \in \mathbb{Z}_{\geq 1} : \begin{array}{c} \exists\, \mathrm{alg} \in \mathcal{A}_r \text{ such that} \\ \sup\limits_{I_d \preceq \Lambda \preceq \kappa I_d} d_{\mathrm{TV}}\big(\mathcal{L}(x_r^{\mathrm{alg},\Lambda}), \mathcal{N}(0,\Lambda^{-1})\big) \leq \varepsilon \end{array} \right\}. \tag{15}$$

Here each algorithm, including its admissible interpolant schedule, query times, and scalar update coefficients, is chosen independently of Λ, the initialization, and the oracle responses.

Proof of Lower Bound in Theorem 3.7. Fix $d \geq 1$, $\kappa \geq 2$, an integer $r \geq 1$, and $\mathrm{alg} \in \mathcal{A}_r$. Write $\delta \triangleq \varepsilon$.

Rational representation. By Lemma 3.6, there is a real rational function $q \triangleq q_r^{\mathrm{alg}} \in \mathcal{R}_{r,r}$ such that

$$x_r^{\mathrm{alg},\Lambda} = q(\Lambda)z, \qquad z \sim \mathcal{N}(0, I_d).$$

Because the schedule, query times, and scalar coefficients are fixed independently of the target, the same function q applies to every $I_d \preceq \Lambda \preceq \kappa I_d$. Moreover, q has no poles on $[1,\kappa]$: the proof of the rationality lemma gives a denominator $\prod_{i=0}^{r-1}(\alpha_{t_i}^2\lambda + \beta_{t_i}^2)$, which is strictly positive for $\lambda > 0$. In particular, q is continuous on $[1,\kappa]$, and

$$\mathcal{L}(x_r^{\mathrm{alg},\Lambda}) = \mathcal{N}\big(0, q(\Lambda)^2\big).$$

A worst-case Gaussian target. We first prove

$$\sup_{I_d \preceq \Lambda \preceq \kappa I_d} d_{\mathrm{TV}}\big(\mathcal{L}(x_r^{\mathrm{alg},\Lambda}), \mathcal{N}(0,\Lambda^{-1})\big) \geq \frac{1}{100}\min\{1, \sqrt{d}\, E_{\kappa,r}\}. \tag{85}$$

If $q(\lambda_0) = 0$ for some $\lambda_0 \in [1,\kappa]$, choose $\Lambda = \lambda_0 I_d$. The algorithm then outputs zero almost surely, whereas $\mathcal{N}(0, \lambda_0^{-1} I_d)$ assigns zero probability to the origin. Their total variation distance is therefore 1, so equation 85 holds in this case.

Otherwise, continuity implies that q has constant sign on $[1,\kappa]$. Thus there is a single $\sigma \in \{-1,1\}$ such that $\sigma q(\lambda) = |q(\lambda)|$ throughout the interval. Since $\sigma q \in \mathcal{R}_{r,r}$ is admissible in the definition of $E_{\kappa,r}$,

$$\max_{\lambda\in[1,\kappa]} \Big|\sqrt{\lambda}\,|q(\lambda)| - 1\Big| = \max_{\lambda\in[1,\kappa]} \Big|\sqrt{\lambda}\,\sigma q(\lambda) - 1\Big| \geq E_{\kappa,r}.$$

Choose $\lambda_* \in [1,\kappa]$ attaining this maximum and set $\Lambda_* \triangleq \lambda_* I_d$. For this target,

$$\mathcal{L}(x_r^{\mathrm{alg},\Lambda_*}) = \mathcal{N}\big(0, q(\lambda_*)^2 I_d\big), \qquad \mathcal{N}(0,\Lambda_*^{-1}) = \mathcal{N}(0,\lambda_*^{-1} I_d).$$

Applying the same-mean Gaussian TV lower bound of Devroye et al. (2023, Theorem 1.1), whose relative covariance eigenvalues here are all $\lambda_* q(\lambda_*)^2$, gives

$$d_{\mathrm{TV}}\big(\mathcal{L}(x_r^{\mathrm{alg},\Lambda_*}), \mathcal{N}(0,\Lambda_*^{-1})\big) \geq \frac{1}{100}\min\Big\{1, \sqrt{d}\,\big|\lambda_* q(\lambda_*)^2 - 1\big|\Big\}. \tag{86}$$

Furthermore,

$$\begin{aligned} \big|\lambda_* q(\lambda_*)^2 - 1\big| &= \Big(\sqrt{\lambda_*}|q(\lambda_*)| + 1\Big)\Big|\sqrt{\lambda_*}|q(\lambda_*)| - 1\Big| \\ &\geq \Big|\sqrt{\lambda_*}|q(\lambda_*)| - 1\Big| \geq E_{\kappa,r}. \end{aligned}$$

Substituting into equation 86 proves equation 85.

From TV accuracy to query complexity. Suppose now that alg satisfies the uniform accuracy guarantee

$$\sup_{I_d \preceq \Lambda \preceq \kappa I_d} d_{\mathrm{TV}}\big(\mathcal{L}(x_r^{\mathrm{alg},\Lambda}), \mathcal{N}(0,\Lambda^{-1})\big) \leq \delta.$$

For $\delta < 1/100$, equation 85 implies

$$\min\{1, \sqrt{d}\, E_{\kappa,r}\} \leq 100\delta < 1, \qquad \text{and hence} \qquad E_{\kappa,r} \leq \frac{100\delta}{\sqrt{d}}.$$

Combining this with the previously established elementary lower bound, obtained from equation 66 and equation 67, yields

$$\frac{1}{4}\exp\left(-\frac{\pi^2(2r+1)}{\ln\kappa}\right) \le E_{\kappa,r} \le \frac{100\delta}{\sqrt{d}}.$$

Taking logarithms gives the non-asymptotic necessary condition

$$2r+1 \ge \frac{\ln\kappa}{\pi^2}\ln\left(\frac{\sqrt{d}}{400\delta}\right). \tag{87}$$

Choose the universal constant $\delta_0 \triangleq 400^{-2}$. For $0 < \delta \le \delta_0$ and $d \ge 1$, we have $\sqrt{d}/\delta \ge 400^2$, so

$$\ln\left(\frac{\sqrt{d}}{400\delta}\right) = \ln\left(\frac{\sqrt{d}}{\delta}\right) - \ln 400 \ge \frac{1}{2}\ln\left(\frac{\sqrt{d}}{\delta}\right).$$

Since $r \ge 1$ implies $2r+1 \le 3r$, equation 87 therefore gives

$$r \ge \frac{1}{6\pi^2}\,\ln\kappa\,\ln\left(\frac{\sqrt{d}}{\delta}\right).$$

This holds for every algorithm and query budget satisfying the uniform accuracy guarantee. Taking the minimum over such budgets proves the lower bound of 3.7, for example with $c = 1/(6\pi^2)$ and $\delta_0 = 400^{-2}$.

Finally, Zoloflow belongs to $\mathcal{A}_r$, and Theorem 3.7 supplies a matching upper bound with a universal constant throughout this parameter regime. This proves the lower bound □

F Score-span Algorithms on Gaussians are Rational Functions

Proof of Lemma 3.6. Fix a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$ and an algorithm $\mathrm{alg} \in \mathcal{A}_r$. Recall that its iterates satisfy

$$x_j = \sum_{i=0}^{j-1} a_{ij}x_i + \sum_{i=0}^{j-1} b_{ij}s_{t_i}(x_i), \qquad j = 1,\dots,r. \tag{88}$$

For a precision eigenvalue $\lambda \in \mathrm{spec}(\Lambda)$, let $x_{\lambda,j} \triangleq u_\lambda u_\lambda^T x_j$ denote the corresponding eigenmode component of x_j. By the Gaussian score formula,

$$s_{t_i}(x_{\lambda,i}) = -\frac{\lambda}{\alpha_{t_i}^2\lambda + \beta_{t_i}^2}\,x_{\lambda,i}. \tag{89}$$

Therefore, projecting equation 88 onto this eigenmode gives

$$x_{\lambda,j} = \sum_{i=0}^{j-1}\left(a_{ij} - b_{ij}\frac{\lambda}{\alpha_{t_i}^2\lambda + \beta_{t_i}^2}\right)x_{\lambda,i}. \tag{90}$$

Indeed

$$x_j = \sum_{i=0}^{j-1}\left[a_{ij}I_d - b_{ij}\left(\alpha_{t_i}^2 I_d + \beta_{t_i}(\Lambda^{-1})\right)^{-1}\right]x_i \tag{91}$$

$$= \sum_{i=0}^{j-1} U\left[a_{ij}I_d - b_{ij}\left(\alpha_{t_i}^2 I_d + \beta_{t_i}^2(\mathrm{diag}(1/\lambda_1,\dots,1/\lambda_d)\right)^{-1}\right]U^T x_i \tag{92}$$

$$= \sum_{i=1}^{j-1} U\,\mathrm{diag}\left(a_{ij} - \frac{b_{ij}\lambda_1}{\alpha_{t_i}^2\lambda_1 + \beta_{t_i}^2},\dots,a_{ij} - \frac{b_{ij}\lambda_d}{\alpha_{t_i}^2\lambda_d + \beta_{t_i}^2}\right)U^T x_i \tag{93}$$

$$= \sum_{k=1}^{d}\sum_{i=1}^{j-1}\left(a_{ij} - \frac{b_{ij}\lambda_1}{\alpha_{t_i}^2\lambda_1 + \beta_{t_i}^2}\right)u_k u_k^T x_i \tag{94}$$

so along any λ-eigenmode $x_{\lambda,j}$

$$x_{\lambda,j} = \sum_{i=1}^{j-1} \left(a_{ij} - \frac{b_{ij}\lambda}{\alpha_{t_i}^2 \lambda + \beta_{t_i}^2} \right) x_{\lambda,i} \tag{95}$$

We prove inductively that for every $j = 0, \ldots, r$ there exists $q_j(\lambda) = \frac{P_j(\lambda)}{D_j(\lambda)} \in \mathcal{R}_{j,j}$ such that

$$x_{\lambda,j} = q_j(\lambda) x_{\lambda,0}. \tag{96}$$

For $j = 0$, take $q_0(\lambda) = 1$.

For the induction step, define

$$d_i(\lambda) \triangleq \alpha_{t_i}^2 \lambda + \beta_{t_i}^2, \qquad D_j(\lambda) \triangleq \prod_{\ell=0}^{j-1} d_\ell(\lambda), \qquad D_0(\lambda) \triangleq 1. \tag{97}$$

Suppose for each $i < j$ that

$$q_i(\lambda) = \frac{P_i(\lambda)}{D_i(\lambda)}, \qquad \deg P_i \le i. \tag{98}$$

Substituting into equation 90 yields

$$q_j(\lambda) = \sum_{i=0}^{j-1} \frac{(a_{ij} d_i(\lambda) - b_{ij}\lambda)\, P_i(\lambda)}{D_{i+1}(\lambda)} \tag{99}$$

$$= \frac{P_j(\lambda)}{D_j(\lambda)}, \tag{100}$$

where

$$P_j(\lambda) \triangleq \sum_{i=0}^{j-1} (a_{ij} d_i(\lambda) - b_{ij}\lambda)\, P_i(\lambda) \prod_{\ell=i+1}^{j-1} d_\ell(\lambda). \tag{101}$$

Each summand in equation 101 has degree at most

$$1 + i + (j - i - 1) = j,$$

and hence $\deg P_j \le j$. Since $\deg D_j = j$, it follows that $q_j \in \mathcal{R}_{j,j}$.

Taking $j = r$, there is therefore a rational function $q_r^{\mathrm{alg}} \in \mathcal{R}_{r,r}$ such that

$$x_{\lambda,r} = q_r^{\mathrm{alg}}(\lambda) x_{\lambda,0}$$

for every eigenmode λ. Since the same scalar function acts on every eigenmode,

$$x_r^{\mathrm{alg}} = q_r^{\mathrm{alg}}(\Lambda) z = U \operatorname{diag}\big(q_r^{\mathrm{alg}}(\lambda_1), \ldots, q_r^{\mathrm{alg}}(\lambda_d)\big)\, U^\top z. \tag{102}$$

Thus every algorithm in $\mathcal{A}_r$ acts on a centered Gaussian target through a type-(r, r) rational function of its precision matrix. □

G Approaches for handling $\mathcal{N}(0, \Lambda^{-1})$, where spec$\{\Lambda\} \in [\lambda_{\min}, \lambda_{\max}]$

Zoloflow is engineered to be optimal for Gaussian targets $x_1 \sim \mathcal{N}(0, \Sigma^{-1})$ which a normalized spectrum $\mathrm{spec}(\Lambda) \in [1, \kappa]$. In this section we ask how we can modify this algoirthm to achieve the same optimal relative error $\hat{E}_{\kappa,r}$ for when spec$(\Lambda^{-1}) \in [\lambda_{min}, \lambda_{max}]$ for $0 < \lambda_{\min} < \lambda_{\max}$ are arbitrary. We introduce two different techniques to accomodate these arbitrary endpoints.

1. **Target rescaling** works by 1) building a modified stochastic interpolant by rescaling the target x_1 so that the spectrum is normalialized to $[1, \kappa]$, 2) runing the new interpolant ODE where we know zoloflow is otpimal, then 3) rescaling the output back to the original spectrum
2. **Gauge rescaling** works by rescaling the interpolant schedules themselves by a *gauge* $g(t)$, a time dependent spacial rescaling which performs a rescaling but amortized over the interpolant time.

We discuss both of these options in the following subsections

Algorithm 2 Sampling with target rescaling

Require: Spectral endpoints $\lambda_{\min} > 0$ and $\lambda_{\max} \geq \lambda_{\min}$, condition number $\kappa = \lambda_{\max}/\lambda_{\min}$, step count r, discretization method $\mathcal{D}$, VP schedule (α_t, β_t)
Require: Score oracle $\widetilde{s}_t = \nabla \log p_{y_t}$ for $y_t = \alpha_t z + \sqrt{\lambda_{\min}}\,\beta_t x_1$
1: Sample $y_0 = z \sim \mathcal{N}(0, I_d)$
2: Define $b_t(y) \triangleq \dfrac{\dot\beta_t}{\beta_t}\big(y + \widetilde{s}_t(y)\big)$
3: Compute y_r using r steps of $\mathcal{D}$ for $\dot y_t = b_t(y_t)$ with schedule (α_t, β_t), initialized at y_0
4: $x_r \leftarrow \lambda_{\min}^{-1/2} y_r$
5: **return** x_r

G.1 Target rescaling

Definition G.1 (Target Rescaled Stochastic Interpolant)**.** For any VP interpolant pair (α, β) consider the rescaled target

$$x_1^{(\lambda_{\min})} \triangleq \sqrt{\lambda_{\min}}\, x_1,\ \Lambda^{(\lambda_{\min})} \triangleq \frac{\Lambda}{\lambda_{\min}},\ \implies x_1^{(\lambda_{\min})} \sim \mathcal{N}\Big(0, (\Lambda^{(\lambda_{\min})})^{-1}\Big),\ I_d \preceq \Lambda^{(\lambda_{\min})} \preceq \kappa I_d. \tag{103}$$

We define the *target rescaled Stochastic Interpolant* transporting z to $x_1^{(\lambda_{\min})}$ as

$$y_t = \alpha_t z + \beta_t x_1^{(\lambda_{\min})} = \alpha_t z + \beta_t \left[\lambda_{\min}^{1/2} x_1\right]. \tag{104}$$

Throughout this subsection, the scale $\lambda_{\min} > 0$ is fixed, and we suppress the dependence of y_t, $\widetilde{p}_t$, and $\widetilde{s}_t$ on $\lambda_{\min}$. Intuitively, this rescaling maps the left endpoint of the target precision spectrum to I_d, which again reduces the distance our initial Gaussian z must travel and recovers our original setup. The associated probability flow ODE is given by

$$\dot y_t = \frac{\dot\beta_t}{\beta_t}\left(y + \tilde s_t(y)\right), \quad y_0^{(l)} \sim \mathcal{N}(0, I), \quad \tilde s_t \triangleq \nabla \log p_{y_t}(y). \tag{105}$$

Equivalent schedule interpretation and reuse of the original score. Target rescaling can equivalently be viewed as retaining the original target x_1 and using coefficients

$$\alpha_t^* \triangleq \alpha_t, \qquad \beta_t^* \triangleq \sqrt{\lambda_{\min}}\,\beta_t, \qquad y_t = x_t^*.$$

Both interpretations define the same interpolant and hence the same probability-flow ODE. Although the starred coefficients are generally not VP or endpoint-normalized, the general case of Proposition I.1 applies. Under its assumptions, with $\sigma(t) = \alpha_t/\beta_t$, the time change and spatial scaling are

$$\tau(t) \triangleq \sigma^{-1}\left(\frac{\sigma(t)}{\sqrt{\lambda_{\min}}}\right), \qquad c_t = A_{\lambda_{\min}}(t) \triangleq \sqrt{\alpha_t^2 + \lambda_{\min}\beta_t^2}.$$

as $(\alpha_{\tau(t)}, \beta_{\tau(t)}) = \left(\frac{\alpha_t}{A_{\lambda_{\min}}(t)}, \frac{\beta_t \lambda_{\min}^{1/2}}{A_{\lambda_{\min}}(t)}\right)$ gives $c_t = \beta_t^*/\beta_{\tau(t)} = A_{\lambda_{\min}}(t)$. Consequently, the rescaled score and velocity satisfy

$$\widetilde{s}_t(y) = \frac{1}{A_{\lambda_{\min}}(t)} s_{\tau(t)}\left(\frac{y}{A_{\lambda_{\min}}(t)}\right), \tag{106}$$

$$\widetilde{b}_t(y) = \frac{\dot\beta_t}{\beta_t}\big(y + \widetilde{s}_t(y)\big) = \frac{\dot\beta_t}{\beta_t}\left[y + \frac{1}{A_{\lambda_{\min}}(t)} s_{\tau(t)}\left(\frac{y}{A_{\lambda_{\min}}(t)}\right)\right]. \tag{107}$$

Thus the target rescaled flow can be implemented using the original score oracle.

The following proposition states that Zoloflow on the rescaled interpolant also recovers the optimal Gaussian sampling rate. We articulate this method for any VP pair (α, β) through the following algorithm

Proposition G.2 (Zoloflow optimality on the target-rescaled interpolant)**.** *Let $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $\mathrm{spec}(\Lambda) \subseteq [\lambda_{\min}, \lambda_{\max}]$, set $\kappa \triangleq \lambda_{\max}/\lambda_{\min}$, and take the elliptic schedule $(\alpha_t, \beta_t) = (\mathrm{cn}(K(m)t \mid m), \mathrm{sn}(K(m)t \mid m))$ with $m = 1 - \kappa^{-1}$. Run Zoloflow, i.e. the finite-difference discretization of Definition 3.3 with r steps and exact scores, on the target-rescaled flow equation 104 from $y_0 = z \sim \mathcal{N}(0, I_d)$, and output $x_r \leftarrow \lambda_{\min}^{-1/2} y_r$. Then*

$$x_r = \lambda_{\min}^{-1/2}\, \hat{q}_{\kappa,r}\left(\frac{\Lambda}{\lambda_{\min}}\right) z = Q_{\lambda_{\min},\kappa,r}(\Lambda)\, z, \qquad Q_{\lambda_{\min},\kappa,r}(\lambda) \triangleq \lambda_{\min}^{-1/2}\, \hat{q}_{\kappa,r}(\lambda/\lambda_{\min}), \quad (108)$$

and

$$\max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} \left| \lambda^{1/2} Q_{\lambda_{\min},\kappa,r}(\lambda) - 1 \right| = \frac{2E_{\kappa,r}}{1 - E_{\kappa,r}} = \hat{E}_{\kappa,r}, \quad (109)$$

the same relative error that $\hat{q}_{\kappa,r}$ attains on $[1, \kappa]$. Moreover, in terms of the reference score s_t, the flow equation 104 reads

$$\dot{y}_t = \frac{\dot{\beta}_t}{\beta_t}\left(y_t + \frac{1}{A_{\lambda_{\min}}(t)}\, s_{\tau(t)}\left(\frac{y_t}{A_{\lambda_{\min}}(t)}\right)\right), \qquad A_{\lambda_{\min}}(t) = \sqrt{\alpha_t^2 + \lambda_{\min}\beta_t^2}, \quad (110)$$

where, writing $(\alpha_t, \beta_t) = (\cos\varphi(t), \sin\varphi(t))$, $\tau(t) = \varphi^{-1}\big(\arctan(\sqrt{\lambda_{\min}}\tan\varphi(t))\big)$. For the elliptic schedule,

$$\tau(t) = \frac{1}{K(m)}\, F\Big(\arctan\big[\sqrt{\lambda_{\min}}\ \mathrm{sc}(K(m)t \mid m)\big] \,\Big|\, m\Big). \quad (111)$$

Proof. Reduction to a single eigenmode. Write $\alpha_i \triangleq \alpha(i/(2r+1))$ and $\beta_i \triangleq \beta(i/(2r+1))$ for $i = 1, \ldots, 2r$, so that the query times and half-steps of Definition 3.3 are $t_j = \frac{2j-1}{2r+1}$ and $t_j + \frac{h}{2} = \frac{2j}{2r+1}$, and $\gamma_i = \beta_i^2/\alpha_i^2$ zeros and poles. The rescaled target $\sqrt{\lambda_{\min}}\, x_1$ is $\mathcal{N}\big(0, (\Lambda/\lambda_{\min})^{-1}\big)$. Fix an eigenvalue λ of Λ and set $\mu \triangleq \lambda/\lambda_{\min} \in [1, \kappa]$. By the Gaussian score formula, the rescaled score acts on this eigenmode as

$$\tilde{s}_t(y_\lambda) = -\frac{\mu}{\mu\alpha_t^2 + \beta_t^2}\, y_\lambda. \quad (112)$$

The same expression follows from equation 89: with $s_\tau(x_\lambda) = -\lambda x_\lambda/(\lambda\alpha_\tau^2 + \beta_\tau^2)$ and $(\alpha_{\tau(t)}, \beta_{\tau(t)}) = (\alpha_t, \sqrt{\lambda_{\min}}\beta_t)/A_{\lambda_{\min}}(t)$, one gets $A_{\lambda_{\min}}(t)^2(\lambda\alpha_{\tau(t)}^2 + \beta_{\tau(t)}^2) = \lambda\alpha_t^2 + \lambda_{\min}\beta_t^2 = \lambda_{\min}(\mu\alpha_t^2 + \beta_t^2)$.

One step. Let $D_i(\mu) \triangleq \mu\alpha_i^2 + \beta_i^2$. Since $\alpha_i^2 + \beta_i^2 = 1$, we have $D_i(\mu) - \mu = (1-\mu)\beta_i^2$, hence $y_\lambda + \tilde{s}_{t_j}(y_\lambda) = \frac{(1-\mu)\beta_{2j-1}^2}{d_{2j-1}(\mu)}\, y_\lambda$. The finite-difference update with coefficient $(\beta_{2j}^2 - \beta_{2j-1}^2)/\beta_{2j-1}^2$ therefore gives

$$y_{\lambda,j} = \left[1 + \frac{(1-\mu)(\beta_{2j}^2 - \beta_{2j-1}^2)}{d_{2j-1}(\mu)}\right] y_{\lambda,j-1} = \frac{\mu + (1-\mu)\beta_{2j}^2}{d_{2j-1}(\mu)}\, y_{\lambda,j-1} = \frac{\mu\,\alpha_{2j}^2 + \beta_{2j}^2}{\mu\,\alpha_{2j-1}^2 + \beta_{2j-1}^2}\, y_{\lambda,j-1}. \quad (113)$$

This identity holds for any VP schedule; the elliptic choice enters only in the next step.

All steps. Since $D_i(\mu) = \alpha_i^2(\mu + \gamma_i) = (\mu + \gamma_i)/(1 + \gamma_i)$,

$$y_{\lambda,r} = \prod_{j=1}^{r} \frac{1 + \gamma_{2j-1}}{1 + \gamma_{2j}} \cdot \frac{\mu + \gamma_{2j}}{\mu + \gamma_{2j-1}}\, y_{\lambda,0} = \hat{q}_{\kappa,r}(\mu)\, y_{\lambda,0} \quad (114)$$

by equation 55, exactly as in Lemma E.1 with Λ replaced by $\Lambda/\lambda_{\min}$. Reassembling the eigenmodes gives $y_r = \hat{q}_{\kappa,r}(\Lambda/\lambda_{\min})z$, and multiplying by $\lambda_{\min}^{-1/2}$ gives the claimed form of x_r.

Error. For $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, $\mu = \lambda/\lambda_{\min}$ ranges over $[1, \kappa]$, and $\lambda^{1/2} Q_{\lambda_{\min},\kappa,r}(\lambda) = \mu^{1/2}\hat{q}_{\kappa,r}(\mu)$. Hence the maximum relative error over $[\lambda_{\min}, \lambda_{\max}]$ equals that of $\hat{q}_{\kappa,r}$ over $[1, \kappa]$, which is $\hat{E}_{\kappa,r} = 2E_{\kappa,r}/(1 - E_{\kappa,r})$ by equation 56.

Reference-score form. Substituting the rescaled score formula I.1 into equation 104 gives the displayed ODE. The time change is defined by equality of noise-to-signal ratios, $\alpha_\tau/\beta_\tau =$

$\alpha_t/(\sqrt{\lambda_{\min}}\beta_t)$, i.e. $\cot\varphi(\tau) = \cot\varphi(t)/\sqrt{\lambda_{\min}}$, i.e. $\tan\varphi(\tau) = \sqrt{\lambda_{\min}}\tan\varphi(t)$. For the elliptic schedule $\varphi(t) = \mathrm{am}(K(m)t \mid m)$, so $\tan\varphi(t) = \mathrm{sc}(K(m)t \mid m)$, and inverting the amplitude with the incomplete elliptic integral $F(\cdot \mid m)$ gives $K(m)\tau(t) = F\big(\arctan[\sqrt{\lambda_{\min}}\,\mathrm{sc}(K(m)t \mid m)] \mid m\big)$. □

Corollary G.3 (Same complexity as Zoloflow)**.** *In the setting of Proposition G.2, let* $\hat{\pi} \triangleq \mathcal{L}(x_r)$. *Then*

$$d_{\mathrm{TV}}(\hat{\pi}, \pi) \le \sqrt{\tfrac{d}{2}}\,\hat{E}_{\kappa,r}, \tag{115}$$

and $d_{\mathrm{TV}}(\hat{\pi}, \pi) \le \varepsilon$ *whenever* $r \ge \frac{\ln(16\kappa)}{2\pi^2}\ln\big(\frac{8\sqrt{2d}}{\varepsilon}\big) - \frac{1}{2}$. *In particular* $r = O\big(\ln\kappa\,\ln(\sqrt{d}/\varepsilon)\big)$ *with* $\kappa = \lambda_{\max}/\lambda_{\min}$, *independently of the scale* $\lambda_{\min}$.

Proof. Total variation is invariant under the common bijection $x \mapsto \sqrt{\lambda_{\min}}\,x$, which maps $\hat{\pi}$ to $\mathcal{L}(y_r)$ and π to $\mathcal{N}\big(0, (\Lambda/\lambda_{\min})^{-1}\big)$. By Proposition G.2, $y_r = \hat{q}_{\kappa,r}(\Lambda/\lambda_{\min})z$ is exactly the output of Zoloflow with parameters (κ, r) on a centered Gaussian target whose precision satisfies $I_d \preceq \Lambda/\lambda_{\min} \preceq \kappa I_d$. Theorem 3.7 therefore applies verbatim and gives both claims. □

G.2 Gauge rescaling

Fix a VP interpolant schedule (α, β) and a scale parameter $\lambda_{\min} > 0$.

Definition G.4 (Gauge)**.** A gauge is a strictly positive function $g \in C^1([0,1])$ satisfying $g_0 = 1$ and $g_1 = \lambda_{\min}^{-1/2}$. The corresponding gauge-rescaled schedule is

$$\alpha_t^* \triangleq g_t\alpha_t, \qquad \beta_t^* \triangleq \sqrt{\lambda_{\min}}\,g_t\beta_t, \qquad t \in [0,1]. \tag{116}$$

These endpoint conditions ensure $(\alpha_0^*, \beta_0^*) = (1, 0)$ and $(\alpha_1^*, \beta_1^*) = (0, 1)$.

For independent $z \sim \mathcal{N}(0, I_d)$ and $x_1 \sim \pi$, the gauge-rescaled interpolant satisfies

$$x_t^* = \alpha_t^* z + \beta_t^* x_1 = g_t \underbrace{\Big(\alpha_t z + \sqrt{\lambda_{\min}}\beta_t x_1\Big)}_{y_t}. \tag{117}$$

Thus it is a time-dependent spatial rescaling of the target interpolant y_t, with $x_0^* = z$ and $x_1^* = x_1$. Every gauge preserves the shifted noise-to-signal ratio $\alpha_t^*/\beta_t^* = \alpha_t/(\sqrt{\lambda_{\min}}\beta_t)$ for $t \in (0,1)$.

Let p_t^* denote the density of x_t^* and define $s_t^*(x) \triangleq \nabla_x \log p_t^*(x)$. Using $\alpha_t^2 + \beta_t^2 = 1$, the associated probability-flow ODE becomes

$$\dot{X}_t = \frac{\dot{g}_t}{g_t}X_t + \frac{\dot{\beta}_t}{\beta_t}\left(X_t + g_t^2 s_t^*(X_t)\right), \qquad t \in (0,1), \qquad X_0 = z. \tag{118}$$

Equivalently, one can use I.1 to write the ODE in terms of the original score. We consider two choices of gauge.

G.2.1 VP gauge

The choice

$$g_t^{\mathrm{VP}} \triangleq \frac{1}{A_{\lambda_{\min}}(t)} = \frac{1}{\sqrt{\alpha_t^2 + \lambda_{\min}\beta_t^2}} \tag{119}$$

gives the variance-preserving schedule

$$\alpha_t^* = \frac{\alpha_t}{A_{\lambda_{\min}}(t)}, \qquad \beta_t^* = \frac{\sqrt{\lambda_{\min}}\beta_t}{A_{\lambda_{\min}}(t)}. \tag{120}$$

The associated probability-flow ODE simplifies to

$$\dot{X}_t = \frac{\dot{\beta}_t}{\beta_t\,(\alpha_t^2 + \lambda_{\min}\beta_t^2)}\left(X_t + s_t^*(X_t)\right), \qquad t \in (0,1), \qquad X_0 = z. \tag{121}$$

Proposition G.5 (Gaussian approximation under the VP gauge)**.** *Let $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $\lambda_{\min} I_d \preceq \Lambda \preceq \lambda_{\max} I_d$, set $\kappa \triangleq \lambda_{\max}/\lambda_{\min}$, and let (α_t^*, β_t^*) be the VP-gauge schedule equation 120 of the elliptic base schedule with parameter κ. Run the finite-difference discretization of Definition 3.3 with the schedule (α^*, β^*) itself with the exact scores s_t^* of the interpolant $x_t^* = \alpha_t^* z + \beta_t^* x_1$ for r steps. Then $x_r = Q_{\lambda_{\min},\kappa,r}(\Lambda) z$, where*

$$Q_{\lambda_{\min},\kappa,r}(\lambda) \triangleq \frac{q_{\kappa,r}(\lambda/\lambda_{\min})}{q_{\kappa,r}(1/\lambda_{\min})}. \tag{122}$$

Here $q_{\kappa,r}$ denotes the Zolotarev inverse-square-root approximant equation 53 on $[1, \kappa]$, with relative error $E_{\kappa,r}$. If $1 \in [\lambda_{\min}, \lambda_{\max}]$, equivalently $1/\lambda_{\min} \in [1, \kappa]$, then

$$\max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} \left| \sqrt{\lambda}\, Q_{\lambda_{\min},\kappa,r}(\lambda) - 1 \right| \le \frac{2E_{\kappa,r}}{1 - E_{\kappa,r}}. \tag{123}$$

Proof. Write $\alpha_i \triangleq \alpha(i/(2r+1))$, $\beta_i \triangleq \beta(i/(2r+1))$ and $D_i(\mu) \triangleq \mu\alpha_i^2 + \beta_i^2$ for $i = 1, \dots, 2r$, so that the query time t_j and the half-step $t_j + h/2$ of the j-th update correspond to $i = 2j - 1$ and $i = 2j$; define α_i^*, β_i^* and $D_i^*(\lambda) \triangleq \lambda\alpha_i^{*2} + \beta_i^{*2}$ in the same way for the VP-gauge schedule.

One step. Fix an eigenvalue λ of Λ and let $x_{\lambda,j}$ denote the corresponding component of x_j. The interpolant x_t^* has law $\mathcal{N}(0, \alpha_t^{*2} I_d + \beta_t^{*2} \Lambda^{-1})$, so its score acts on this eigenmode as $s_t^*(x_\lambda) = -\lambda x_\lambda / (\lambda\alpha_t^{*2} + \beta_t^{*2})$. Since $\alpha_t^{*2} + \beta_t^{*2} = 1$, the one-step identity in the proof of Proposition G.2, which uses only the VP property, gives

$$x_{\lambda,j} = \frac{D_{2j}^*(\lambda)}{D_{2j-1}^*(\lambda)}\, x_{\lambda,j-1}, \qquad j = 1, \dots, r. \tag{124}$$

The gauge. Put $\mu \triangleq \lambda/\lambda_{\min}$ and $\mu_0 \triangleq 1/\lambda_{\min}$. By equation 120,

$$D_i^*(\lambda) = \frac{\lambda\alpha_i^2 + \lambda_{\min}\beta_i^2}{\alpha_i^2 + \lambda_{\min}\beta_i^2} = \frac{\lambda_{\min}\,(\mu\alpha_i^2 + \beta_i^2)}{\lambda_{\min}\,(\mu_0\alpha_i^2 + \beta_i^2)} = \frac{D_i(\mu)}{D_i(\mu_0)}. \tag{125}$$

All steps. Multiplying the one-step factors,

$$x_{\lambda,r} = \prod_{j=1}^{r} \frac{D_{2j}(\mu)}{D_{2j-1}(\mu)} \Bigg/ \prod_{j=1}^{r} \frac{D_{2j}(\mu_0)}{D_{2j-1}(\mu_0)} \;\; x_{\lambda,0} = \frac{\widehat{q}_{\kappa,r}(\mu)}{\widehat{q}_{\kappa,r}(\mu_0)}\, x_{\lambda,0} = \frac{q_{\kappa,r}(\mu)}{q_{\kappa,r}(\mu_0)}\, x_{\lambda,0}. \tag{126}$$

The second equality is the identity $\prod_j D_{2j}(\mu)/D_{2j-1}(\mu) = \widehat{q}_{\kappa,r}(\mu)$ of Lemma E.1 for the elliptic base schedule, which holds for every $\mu > 0$; the third cancels the common factor $1/q_{\kappa,r}(1)$ in equation 55. Reassembling the eigenmodes gives $x_r = Q_{\lambda_{\min},\kappa,r}(\Lambda) z$.

Error. Let $e(\mu) \triangleq \sqrt{\mu}\, q_{\kappa,r}(\mu) - 1$, so that $|e(\mu)| \le E_{\kappa,r}$ for $\mu \in [1, \kappa]$ by equation 49. For $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ we have $\mu \in [1, \kappa]$, and $1 \in [\lambda_{\min}, \lambda_{\max}]$ is exactly $\mu_0 \in [1, \kappa]$. Since $\sqrt{\lambda} = \sqrt{\mu}/\sqrt{\mu_0}$,

$$\sqrt{\lambda}\, Q_{\lambda_{\min},\kappa,r}(\lambda) = \frac{\sqrt{\mu}\, q_{\kappa,r}(\mu)}{\sqrt{\mu_0}\, q_{\kappa,r}(\mu_0)} = \frac{1 + e(\mu)}{1 + e(\mu_0)} \in \left[\frac{1 - E_{\kappa,r}}{1 + E_{\kappa,r}},\, \frac{1 + E_{\kappa,r}}{1 - E_{\kappa,r}} \right], \tag{127}$$

where $E_{\kappa,r} < 1$ (see the proof of Corollary D.4). Both endpoints of this interval lie within $2E_{\kappa,r}/(1 - E_{\kappa,r})$ of 1, which proves the bound. □

G.2.2 Exponential gauge

The choice $g_t^{\exp} \triangleq \lambda_{\min}^{-t/2}$ gives the generally non-VP schedule

$$\alpha_t^* = \lambda_{\min}^{-t/2} \alpha_t, \qquad \beta_t^* = \lambda_{\min}^{(1-t)/2} \beta_t. \tag{128}$$

Its probability-flow ODE is

$$\dot{X}_t = -\frac{\log \lambda_{\min}}{2} X_t + \frac{\dot{\beta}_t}{\beta_t} \left(X_t + \lambda_{\min}^{-t} s_t^*(X_t) \right), \qquad t \in (0, 1), \qquad X_0 = z. \tag{129}$$

Generalized Zoloflow. We use an exponential integrator-like discretization of the exponential-gauge ODE: the score-dependent term is approximated by the elliptic finite-difference rule, while the linear term $-\frac{1}{2}\log(\lambda_{\min})X_t$ is integrated exactly. Distributing this linear evolution equally across r updates gives the factor

$$\exp\left(-\frac{\log\lambda_{\min}}{2r}\right) = \lambda_{\min}^{-1/(2r)} \tag{130}$$

per update.

Definition G.6 (Generalized Zoloflow)**.** Fix $\lambda_{\min} > 0$, $\kappa > 1$, and an integer $r \geq 1$. Let $(\alpha_t, \beta_t) = (\mathrm{cn}(K(m)t \mid m), \mathrm{sn}(K(m)t \mid m))$, where $m \triangleq 1 - \kappa^{-1}$, and let s_t^* denote the score under the exponential-gauge schedule. Define

$$h \triangleq \frac{2}{2r+1}, \qquad t_j \triangleq \frac{2j-1}{2r+1}, \qquad \omega_j \triangleq \frac{\beta_{t_j+h/2}^2 - \beta_{t_j}^2}{\beta_{t_j}^2}, \qquad j = 1, \ldots, r. \tag{131}$$

Starting from $x_0 = z \sim \mathcal{N}(0, I_d)$, Generalized Zoloflow performs the updates

$$x_j = \lambda_{\min}^{-1/(2r)}\left[x_{j-1} + \omega_j\left(x_{j-1} + \lambda_{\min}^{-t_j} s_{t_j}^*(x_{j-1})\right)\right], \qquad j = 1, \ldots, r, \tag{132}$$

and returns x_r. The finite-difference weights use the base schedule β, rather than the transformed schedule β^*.

For a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $\lambda_{\min} I_d \preceq \Lambda \preceq \lambda_{\min}\kappa I_d$, exact scores satisfy, along a precision eigenmode λ,

$$\lambda_{\min}^{-t} s_t^*(x_\lambda) = -\frac{\lambda/\lambda_{\min}}{(\lambda/\lambda_{\min})\alpha_t^2 + \beta_t^2}\, x_\lambda. \tag{133}$$

Consequently, the eigenmode update is

$$x_{\lambda,j} = \lambda_{\min}^{-1/(2r)} \frac{(\lambda/\lambda_{\min})\alpha_{t_j+h/2}^2 + \beta_{t_j+h/2}^2}{(\lambda/\lambda_{\min})\alpha_{t_j}^2 + \beta_{t_j}^2} x_{\lambda,j-1}. \tag{134}$$

Defining the positive zero and pole shifts

$$\xi_j \triangleq \frac{\beta_{t_j+h/2}^2}{\alpha_{t_j+h/2}^2}, \qquad \zeta_j \triangleq \frac{\beta_{t_j}^2}{\alpha_{t_j}^2}, \tag{135}$$

this becomes

$$x_{\lambda,j} = \lambda_{\min}^{-1/(2r)} \frac{\lambda/\lambda_{\min} + \xi_j}{\lambda/\lambda_{\min} + \zeta_j} \frac{1+\zeta_j}{1+\xi_j} x_{\lambda,j-1}. \tag{136}$$

Thus the output is

$$x_r = \lambda_{\min}^{-1/2} \widehat{q}_{\kappa,r}\left(\frac{\Lambda}{\lambda_{\min}}\right) z, \tag{137}$$

where $\widehat{q}_{\kappa,r}$ is the endpoint-normalized Zolotarev approximant on $[1, \kappa]$.

H Parameter-selection heuristic for general distributions

As described in Appendix G the elliptic schedule $(\alpha_t, \beta_t) = (\mathrm{cn}(tK \mid m), \mathrm{sn}(tK \mid m))$ may be arbitrarily rescaled with an extra parameter $\ell = \lambda_{\min}(\Lambda)$ in order to stay optimal for unnormalized Gaussian targets $\pi \sim \mathcal{N}(0, \Lambda^{-1})$ when $\lambda_{\min}(\Lambda) \neq 1$. To use the elliptic schedule correctly for arbitrary Gaussians, one genuinely must choose two parameters:

1. a condition number κ_{param} to handle the anisotropy of π (then set $m = 1 - \frac{1}{\kappa}$),
2. a scaling parameter $\ell > 0$ to handle unnormalized spectra.

Equivalently this choice can be described as the pairing (ℓ, L), where ℓ is again the scaling parameter, and $L \triangleq \ell\kappa_{\text{param}}$. For Gaussians, optimal parameter choices for (ℓ, L) depend on curvature endpoints $-\nabla^2 \log \pi$ or equivalently the smallest and largest eigenvalues of the precision matrix $(\lambda_{\min}(\Lambda), \lambda_{\max}(\Lambda))$, where

$$\lambda_{\min} I_d \preceq -\nabla^2 \log \pi(x) = \Lambda \preceq \lambda_{\max} I_d. \tag{138}$$

However, this equivalence decouples for general π, and so it's not obvious how to pick these parameters. For a general distribution π with precision matrix $\Lambda_\pi = \Sigma_\pi^{-1}$ the two immediate extensions of these choices are

1. **Curvature endpoints**:
$$(\ell, L) = \big(\lambda_{\min}(-\nabla^2 \log \pi), \lambda_{\max}(-\nabla^2 \log \pi)\big) \tag{139}$$
2. **Precision matrix endpoints**
$$(\ell, L) = (\lambda_{\min}(\Lambda), \lambda_{\max}(\Lambda)) \tag{140}$$

The immediate issue with the curvature endpoints is generally $\ell < 0$ can occur which is prohibited. Nevertheless even in the log-concave setting we provide evidence that this choice is too conservative in A.2. We propose the precision matrix endpoints as our heuristic:

The heuristic. Let $x_1 \sim \pi$ have mean μ and covariance $\Sigma = \mathbb{E}[(x_1 - \mu)(x_1 - \mu)^\top]$. We propose to run the Elliptic schedule, as described in (Section G.1, using the parameters $(\ell, L, \kappa_{\text{param}})$ instead of $(\lambda_{\min}, \lambda_{\max}, \kappa)$. We set $(\ell, L, \kappa_{\text{param}})$ as follows

$$\ell = \lambda_{\min}(\Sigma^{-1}) = \frac{1}{\lambda_{\max}(\Sigma)}, \qquad L = \lambda_{\max}(\Sigma^{-1}) = \frac{1}{\lambda_{\min}(\Sigma)}, \qquad \kappa_{\text{param}} = \frac{L}{\ell}. \tag{141}$$

The heuristic has three properties that motivate it:

1. *It is exact in the Gaussian case.* If $x_1 \sim \mathcal{N}(m, \Lambda^{-1})$ then $\Sigma^{-1} = \Lambda$, so equation 141 returns the endpoints of the precision spectrum, which is the optimal choice of Section 3.
2. *It reduces to scale matching when there is nothing else to exploit.* If $\Sigma = \sigma^2 I$, then $\ell = L = \sigma^{-2}$ and the schedule is the trig schedule rescaled to the data variance, the standard practice of Align Your Steps. In particular the heuristic never widens the schedule on a target whose scales agree across directions.
3. *It is bounded by the curvature.* For strongly log-concave π, we have $\lambda_{\min} \le \lambda_{\min}(\Sigma^{-1})$ and $\lambda_{\max}(\Sigma^{-1}) \le \lambda_{\max}$, so the interval equation 141 always lies inside $[\lambda_{\min}, \lambda_{\max}]$ and is never wider than the curvature recipe.

In practice, we do not have access to the exact covariance Σ, so we recommend computing the sample covariance $\hat{\Sigma}$ and taking the 0.5th and 99.5th quantiles of the spectrum of $\lambda_{\min}^{(0.5)}(\hat{\Sigma})$ and $\lambda_{\max}^{(99.5)}(\hat{\Sigma})$

$$(\ell, L) = \left(\frac{1}{\lambda_{\max}^{(99.5)}}, \frac{1}{\lambda_{\min}^{(0.5)}}\right) \tag{142}$$

and using this ℓ and L for target rescaled Elliptic schedule. Alternatively, it can serve to guide a more in-depth hyperparameter sweep.

I Interpolant schedules

I.1 Interpolant schedule catalog

The following is a list of interpolant schedules.

- **Linear Schedule:**
$$\alpha^{lin}(t) = 1 - t, \quad \beta^{lin}(t) = t. \tag{143}$$

- **Square Root Schedule:**
$$\alpha^{sqrt}(t) = \sqrt{1-t}, \quad \beta^{sqrt}(t) = \sqrt{t}. \tag{144}$$
- **Trigonometric Schedule:**
$$\alpha^{trig}(t) = \cos\left(\frac{1}{2}\pi t\right), \quad \beta^{trig}(t) = \sin\left(\frac{1}{2}\pi t\right). \tag{145}$$
- **Elliptic Schedule:** For a given $\kappa \geq 1$, with $m = 1 - \kappa^{-1}$ and $K = K(m)$,
$$\alpha^{ell}(t) = \mathrm{cn}(Kt \mid m), \quad \beta^{ell}(t) = \mathrm{sn}(Kt \mid m). \tag{146}$$
- **Lipschitz-Guided:** For a given $\kappa > 1$,
$$\alpha^{lip}(t) = \sqrt{\frac{\kappa^{1-t} - 1}{\kappa - 1}}, \quad \beta^{lip}(t) = \sqrt{\frac{\kappa - \kappa^{1-t}}{\kappa - 1}}. \tag{147}$$
This schedule was introduced in Chen et al. (2025).

I.2 Conversions between different Interpolant schedules

Proposition I.1 (Conversion between interpolant schedules)**.** *Let $z \sim \mathcal{N}(0, I_d)$ and $x_1 \sim \pi$ be independent, with $\mathbb{E}\|x_1\| < \infty$. For positive coefficient functions $\alpha, \beta, \alpha^*, \beta^* \in C^1((0,1))$, define*
$$x_t = \alpha_t z + \beta_t x_1, \qquad x_t^* = \alpha_t^* z + \beta_t^* x_1.$$
Denote their respective densities by p_t, p_t^*, their scores by $s_t = \nabla \log p_t$, $s_t^* = \nabla \log p_t^*$, and their velocities by $b_t(x) = \mathbb{E}[\dot{x}_t \mid x_t = x]$ and $b_t^*(x) = \mathbb{E}[\dot{x}_t^* \mid x_t^* = x]$.

Define $\sigma(t) \triangleq \alpha_t/\beta_t$ and $\sigma^*(t) \triangleq \alpha_t^*/\beta_t^*$. Assume
$$\dot{\sigma}(t) \neq 0 \quad \text{for all } t \in (0,1), \qquad \sigma^*((0,1)) \subseteq \sigma((0,1)).$$
Then the time change and spatial scaling
$$\tau(t) \triangleq \sigma^{-1}(\sigma^*(t)), \qquad c_t \triangleq \frac{\beta_t^*}{\beta_{\tau(t)}} = \frac{\alpha_t^*}{\alpha_{\tau(t)}}$$
are well-defined and C^1. In particular we pick $\tau(t)$ in order to make equal the SNRs of the two schedules.
$$\sigma_{\tau(t)} = \frac{\alpha_{\tau_t}}{\beta_{\tau_t}} = \frac{\alpha_t^*}{\beta_t^*} = \sigma_t^*. \tag{148}$$
Now for $t \in (0,1)$,
$$x_t^* = c_t x_{\tau(t)}, \tag{149}$$
$$s_t^*(x) = \frac{1}{c_t} s_{\tau(t)}(x/c_t), \tag{150}$$
$$b_t^*(x) = \frac{\dot{c}_t}{c_t} x + c_t \dot{\tau}(t) b_{\tau(t)}(x/c_t). \tag{151}$$
Consequently, any solution $\dot{X}_u = b_u(X_u)$ on $(0,1)$ yields a solution $X_t^* \triangleq c_t X_{\tau(t)}$ of $\dot{X}_t^* = b_t^*(X_t^*)$. If both coefficient pairs are variance-preserving, then $c_t = 1$, so
$$s_t^*(x) = s_{\tau(t)}(x), \qquad b_t^*(x) = \dot{\tau}(t) b_{\tau(t)}(x). \tag{152}$$
In short we have the following summary

	Object	*General schedules*	*Both VP* $(c_t = 1)$
Interpolant	x_t^*	$c_t x_{\tau(t)}$	$x_{\tau(t)}$
ODE solution	X_t^*	$c_t X_{\tau(t)}$	$X_{\tau(t)}$
Score	$s_t^*(x)$	$c_t^{-1} s_{\tau(t)}(x/c_t)$	$s_{\tau(t)}(x)$
Velocity field *(from velocity)*	$b_t^*(x)$	$\frac{\dot{c}_t}{c_t} x + c_t \dot{\tau}(t) b_{\tau(t)}(x/c_t)$	$\dot{\tau}(t) b_{\tau(t)}(x)$
Velocity field *(from score)*	$b_t^*(x)$	$\frac{\dot{\beta}_t^*}{\beta_t^*} x + \frac{(\alpha_t^*)^2}{c_t}\left(\frac{\dot{\beta}_t^*}{\beta_t^*} - \frac{\dot{\alpha}_t^*}{\alpha_t^*}\right) s_{\tau(t)}(x/c_t)$	$\frac{\dot{\beta}_t^*}{\beta_t^*}\left(x + s_{\tau(t)}(x)\right)$

where derivatives have closed forms: $\frac{\dot{c}_t}{c_t} = \frac{\dot{\alpha}_t^*}{\alpha_t^*} - \dot{\tau}_t \frac{\dot{\alpha}_{\tau_t}}{\alpha_{\tau_t}}$ and $\dot{\tau}_t = \frac{(\sigma^*)'(t)}{\sigma'(\tau(t))}$.

Proposition I.1 shows that any two linear stochastic interpolants with the same target $x_1 \sim \pi$ and Gaussian reference $z \sim \mathcal{N}(0, I_d)$ are related by a change of time and a rescaling of space. The time change τ is chosen so that the two schedules have the same noise-to-signal ratio α/β at corresponding times, and the scores and velocity fields then transform accordingly. When both schedules are variance-preserving, the spatial rescaling is trivial ($c_t = 1$) and only the time change remains.

I.3 Optimality properties of the elliptic schedule

Here we provide a proof of the optimality of the Jacobi amplitude, stated as Proposition 3.1. The basic idea is to construct a conformal map from the singularity-free region to the strip.

Proposition 3.1 (Jacobi amplitude as optimal phase)**.** *Consider the phase (angular clock) given by*

$$\varphi_\star(t) \triangleq \mathrm{am}(Kt \mid m), \tag{7}$$

where $m = 1 - \kappa^{-1}$, $K = K(m)$, $K' = K(1-m)$, and $\mathrm{am}()$ is the Jacobi amplitude function and $K()$ the complete elliptic integral of first kind (see Appendix C). Let $S_\gamma := \{t \in \mathbb{C} : |\operatorname{Im} t| < \gamma\}$ be a strip around the real axis of width 2γ. This choice of φ maximizes γ (analyticity strip width), subject to the constraints i) φ is holomorphic on S_γ, ii) $\varphi(0) = 0$, $\varphi(1) = \pi/2$, iii) $1 - (1 - \lambda^{-1}) \sin^2 \varphi(t) \neq 0$, $\forall t \in S_\gamma$, $\forall \lambda \in [1, \kappa]$. In particular, for this choice of $\varphi_\star$ we have $\gamma_\star = K'/K$.

Proof. We first rewrite the nonvanishing constraint as an avoidance condition in the complex φ-plane. For $\lambda \in [1, \kappa]$, $1 - \left(1 - \lambda^{-1}\right) \sin^2 \varphi = 0$ if and only if $\sin^2 \varphi = \lambda/(\lambda - 1)$. As λ ranges from 1 to κ, this ranges over the interval $[\kappa/(\kappa - 1), \infty) = [m^{-1}, \infty)$. Thus the third condition is equivalent to $\sin^2 \varphi(t) \notin [m^{-1}, \infty), \forall t \in S_\gamma$. Consider the domain Ω_m given by the connected component containing $[0, \pi/2]$ of $\left\{z \in \mathbb{C} : \sin^2 z \notin [m^{-1}, \infty)\right\}$. A standard conformal representation of Ω_m is given by the incomplete elliptic integral

$$F(z \mid m) = \int_0^z \frac{d\zeta}{\sqrt{1 - m \sin^2 \zeta}}.$$

Namely, $\Psi_m(z) \triangleq F(z \mid m)/K(m)$ is a biholomorphism $\Psi_m : \Omega_m \longrightarrow S_{\gamma_\star}$, with inverse $\Psi_m^{-1}(w) = \mathrm{am}(K(m)w \mid m)$. The boundary interpolation conditions are satisfied, since $\Psi_m(0) = 0$ and $\Psi_m(\pi/2) = 1$, since $F(0 \mid m) = 0$ and $F(\pi/2 \mid m) = K(m)$.

We now prove optimality, using Schwarz-Pick. Let φ be any other phase satisfying the three conditions of the proposition on S_γ. Since S_γ is connected, $\varphi(0) = 0$, its image must lie in the component Ω_m. Hence $h \triangleq \Psi_m \circ \varphi$ defines a holomorphic map $h : S_\gamma \longrightarrow S_{\gamma_\star}$. The endpoint constraints give $h(0) = 0$ and $h(1) = 1$. We now apply the Schwarz–Pick theorem to the two horizontal strips. Recall that $z \longmapsto \tanh\left(\frac{\pi z}{4a}\right)$ maps $S_a = \{z : |\operatorname{Im} z| < a\}$ biholomorphically onto the unit disk. Consequently, the hyperbolic distance between two points $x, y \in \mathbb{R}$ on the center line of S_a is $d_{S_a}(x, y) = \frac{\pi}{2a}|x - y|$. In particular, $d_{S_\gamma}(0, 1) = \pi/(2\gamma)$ and $d_{S_{\gamma_\star}}(0, 1) = \pi/(2\gamma_\star)$. Since holomorphic maps decrease hyperbolic distance, Schwarz–Pick gives $d_{S_{\gamma_\star}}\big(h(0), h(1)\big) \leq d_{S_\gamma}(0, 1)$. Using $h(0) = 0$ and $h(1) = 1$, we obtain

$$\frac{\pi}{2\gamma_\star} \leq \frac{\pi}{2\gamma},$$

and hence

$$\gamma \leq \gamma_\star = \frac{K(1-m)}{K(m)},$$

with equality if and only if $\varphi(t) = \mathrm{am}(K(m)t|m)$. Thus no admissible phase can be holomorphic on a wider horizontal strip.

Finally, we verify that the bound is attained. The inverse of the extremal conformal map is $\varphi_\star(t) = \mathrm{am}(Kt \mid m)$. It is holomorphic on $S_{\gamma_\star}$, satisfies the boundary conditions $\varphi_\star(0) = 0$, $\varphi_\star(1) = \mathrm{am}(K \mid m) = \frac{\pi}{2}$, and $1 - (1 - \lambda^{-1}) \sin^2 \varphi_\star(t)$ does not vanish on the strip $S_{\gamma_\star}$ for all $\lambda \in [1, \kappa]$. Therefore the Jacobi amplitude $\varphi_\star$ is admissible and attains the maximal strip width $\gamma_\star = K(1-m)/K(m)$. □

I.4 SPECTRAL PARAMETERIZATION

The following is restricted to variance-preserving schedules, and we count only *problem-dependent spectral parameters*; the path variable t and any numerical truncation endpoints are not included in this count. To accommodate schedules with square-root endpoint behavior, let $\widetilde{\mathcal{S}}$ denote the class of monotone VP schedules satisfying the usual endpoint conditions and

$$\alpha, \beta \in C([0,1]) \cap C^1((0,1)), \qquad \alpha^2, \beta^2 \in C^1([0,1]). \tag{153}$$

This class contains the $C^1([0,1])$ schedules used in the main text while also admitting the square-root and Lipschitz schedules below. The common linear interpolant $(1-t, t)$ is not VP and is therefore outside this catalog.

Parameter-free schedules. Two standard VP clocks require no target-dependent spectral information:

$$(\alpha_t^{\mathrm{trig}}, \beta_t^{\mathrm{trig}}) = \left(\cos\left(\frac{\pi t}{2}\right), \sin\left(\frac{\pi t}{2}\right)\right), \tag{154}$$

$$(\alpha_t^{\mathrm{sqrt}}, \beta_t^{\mathrm{sqrt}}) = \left(\sqrt{1-t}, \sqrt{t}\right). \tag{155}$$

The first is the trigonometric schedule of Albergo et al. (2025, Table 1). The second, which we call the *linear-variance* or *square-root* schedule because $(\alpha_t^2, \beta_t^2) = (1-t, t)$, is obtained from their VP/SBDM parameterization by the time change $u = \sqrt{t}$ (Albergo et al., 2025, Section 5.1).

One spectral shape parameter. Suppose the precision spectrum has been normalized to $[1, \kappa]$. Define $m \triangleq 1 - \kappa^{-1}$. The two κ-indexed schedules considered in this paper are

$$(\alpha_t^{\mathrm{ell}}, \beta_t^{\mathrm{ell}}) = (\mathrm{cn}(K(m)t \mid m), \mathrm{sn}(K(m)t \mid m)), \tag{156}$$

$$(\alpha_t^{b}, \beta_t^{b}) = \left(\sqrt{\frac{\kappa^{1-t}-1}{\kappa-1}}, \sqrt{\frac{\kappa-\kappa^{1-t}}{\kappa-1}}\right). \tag{157}$$

The first is our elliptic schedule. The second makes the worst-case Gaussian solution move at constant relative speed. It agrees algebraically with the Gaussian minimizer of averaged squared drift Lipschitzness in Chen et al. (2025, Proposition 3.9 and Equation (3.11)); Appendix I.5 additionally proves its minimax L^∞ drift-Lipschitz interpretation.

These families have two distinct parameter-free limits:

$$(\alpha^{\mathrm{ell}}, \beta^{\mathrm{ell}}) \xrightarrow[\kappa\downarrow 1]{} (\alpha^{\mathrm{trig}}, \beta^{\mathrm{trig}}), \qquad (\alpha^{b}, \beta^{b}) \xrightarrow[\kappa\downarrow 1]{} (\alpha^{\mathrm{sqrt}}, \beta^{\mathrm{sqrt}}). \tag{158}$$

At $\kappa = 1$, every VP schedule gives a constant Gaussian solution and zero drift. The square-root schedule is therefore the continuous extension of the two Lipschitz families, not a uniquely selected optimizer of the degenerate problem.

Two spectral endpoints. For a general precision interval $[\lambda_{\min}, \lambda_{\max}]$, write $\kappa = \lambda_{\max}/\lambda_{\min}$ and let $(\alpha^{\mathrm{ell}}, \beta^{\mathrm{ell}})$ be the corresponding elliptic schedule above. Its $\lambda_{\min}$-shift is

$$(\widetilde{\alpha}_t^{\mathrm{ell}}, \widetilde{\beta}_t^{\mathrm{ell}}) = \frac{(\alpha_t^{\mathrm{ell}}, \sqrt{\lambda_{\min}}\,\beta_t^{\mathrm{ell}})}{\sqrt{(\alpha_t^{\mathrm{ell}})^2 + \lambda_{\min}(\beta_t^{\mathrm{ell}})^2}}. \tag{159}$$

This family uses two independent pieces of spectral information: scale $\lambda_{\min}$ and shape κ, equivalently the two endpoints $(\lambda_{\min}, \lambda_{\max})$. It reduces to the one-parameter elliptic family when $\lambda_{\min} = 1$. We therefore refer to it as the *two-endpoint-adapted* or *normalization-agnostic* family, rather than as spectrum-agnostic. Its other one-parameter slice, $\kappa = 1$, varies only the target scale and contains the isotropic construction below. These categories therefore form an overlapping hierarchy of target adaptation, rather than disjoint schedule classes.

Remark (Relation to Align Your Steps). The term *sampling schedule* in Sabour et al. (2024) denotes a discrete VE noise-level grid rather than a continuous stochastic-interpolant schedule. For $x_1 \sim \mathcal{N}(0, c^2 I_d)$, their Gaussian Euler-optimal grid is

$$\sigma_i^\star = c\tan\left[\left(1 - \frac{i}{n}\right)\arctan\left(\frac{\sigma_{\min}}{c}\right) + \frac{i}{n}\arctan\left(\frac{\sigma_{\max}}{c}\right)\right], \qquad i = 0, \dots, n. \tag{160}$$

To compare this grid with our construction, take $\lambda_{\min} = \lambda_{\max} = c^{-2}$, so $\kappa = 1$, and write $\theta(u) = \pi u/2$. Since the elliptic schedule then reduces to $(\cos\theta, \sin\theta)$, its c^{-2}-shift is

$$(\widetilde{\alpha}_u, \widetilde{\beta}_u) = \frac{(c\cos\theta(u), \sin\theta(u))}{\sqrt{c^2\cos^2\theta(u) + \sin^2\theta(u)}}. \tag{161}$$

The induced VE noise coordinate therefore satisfies

$$\sigma(u) = \frac{\widetilde{\alpha}_u}{\widetilde{\beta}_u} = c\cot\theta(u), \qquad \arctan\left(\frac{\sigma(u)}{c}\right) = \frac{\pi}{2}(1-u). \tag{162}$$

Consequently, after reversing orientation and restricting to the finite noise interval $[\sigma_{\min}, \sigma_{\max}]$, a uniform interpolant-time grid gives exactly equation 160. This identifies the induced clocks and grids, not the numerical algorithms: Sabour et al. (2024) prove optimality for a forward-Euler/DDIM grid, whereas Zoloflow uses the elliptic finite-difference discretization.

For arbitrary $c > 0$, this correspondence is read most directly through the target rescaling of Appendix G.1; at the level of the induced VE clock, it is equivalently the c^{-2}-shift above. The error guarantee in Appendix G.2.1 additionally assumes $1 \in [\lambda_{\min}, \lambda_{\max}]$, which for the singleton $\{c^{-2}\}$ holds only when $c = 1$. Thus the clock identity above does not by itself invoke that normalization-agnostic error guarantee.

I.5 Lipschitz-optimal interpolant schedules

I.5.1 Solution regularity and drift regularity

Consider a centered Gaussian target $\pi = \mathcal{N}(0, \Lambda^{-1})$ with $I_d \preceq \Lambda \preceq \kappa I_d$, where $\kappa > 1$. Different interpolant schedules produce the same target law, but change how the solution and velocity field evolve along the interpolation. Two natural design criteria are to control the temporal variation of the solution or the spatial sensitivity of the drift.

We work over the endpoint-relaxed VP class $\widetilde{\mathcal{S}}$ defined in Appendix I.4.

For a precision eigenvalue $\lambda \in [1, \kappa]$, let $x_\lambda(t)$ denote the normalized probability-flow solution along the corresponding eigenmode:

$$x_\lambda(t) = \sqrt{\alpha_t^2 + \frac{\beta_t^2}{\lambda}}, \qquad x_\lambda(0) = 1, \qquad x_\lambda(1) = \lambda^{-1/2}. \tag{163}$$

The corresponding drift multiplier is

$$g_\lambda(t) = \frac{\lambda\alpha_t\dot{\alpha}_t + \beta_t\dot{\beta}_t}{\lambda\alpha_t^2 + \beta_t^2} = \frac{\dot{x}_\lambda(t)}{x_\lambda(t)} = \frac{d}{dt}\log x_\lambda(t), \tag{164}$$

so that $\dot{x}_\lambda(t) = g_\lambda(t)x_\lambda(t)$.

Following these two representations, define

$$L_\lambda^x[\alpha, \beta] \triangleq \sup_{t\in(0,1)} |\dot{x}_\lambda(t)|, \qquad L_\lambda^b[\alpha, \beta] \triangleq \sup_{t\in(0,1)} |g_\lambda(t)|. \tag{165}$$

The first measures the temporal Lipschitz constant of the normalized solution. The second measures the spatial Lipschitz constant of the eigenmode drift, uniformly over time: it controls the map $x \mapsto g_\lambda(t)x$, rather than the temporal derivative $\dot{g}_\lambda(t)$.

Under the VP constraint, $\alpha_t\dot{\alpha}_t = -\beta_t\dot{\beta}_t$, and hence

$$|\dot{x}_\lambda(t)| = \frac{(1-\lambda^{-1})\beta_t\dot{\beta}_t}{x_\lambda(t)}, \qquad |g_\lambda(t)| = \frac{(1-\lambda^{-1})\beta_t\dot{\beta}_t}{x_\lambda(t)^2}. \tag{166}$$

Both expressions are increasing in λ: the factor $1-\lambda^{-1}$ increases, while $x_\lambda(t)$ decreases. Consequently, both worst-case criteria are attained at the upper spectral endpoint:

$$\max_{\lambda\in[1,\kappa]} L_\lambda^x = L_\kappa^x, \qquad \max_{\lambda\in[1,\kappa]} L_\lambda^b = L_\kappa^b. \tag{167}$$

Moreover, specifying the worst-case solution x_κ uniquely determines a VP schedule, since

$$\alpha_t^2 = \frac{\kappa x_\kappa(t)^2 - 1}{\kappa - 1}, \qquad \beta_t^2 = \frac{\kappa(1 - x_\kappa(t)^2)}{\kappa - 1}. \tag{168}$$

Thus both schedule-design problems reduce to choosing how a scalar function moves from 1 to $\kappa^{-1/2}$.

I.5.2 Drift-Lipschitz optimality: constant relative speed

The drift-Lipschitz problem instead minimizes

$$\min_{(\alpha,\beta)\in\widetilde{\mathcal{S}}} \max_{\lambda\in[1,\kappa]} L_\lambda^b[\alpha,\beta]. \tag{169}$$

Writing $d_\kappa = \frac{1}{2}\log\kappa$, the endpoint lower bound now applies to the logarithm of the solution:

$$L_\kappa^b \geq \int_0^1 \left|\frac{d}{dt}\log x_\kappa(t)\right| dt = \log x_\kappa(0) - \log x_\kappa(1) = d_\kappa. \tag{170}$$

Equality is attained by making the logarithmic speed constant,

$$\log x_\kappa^{\mathrm{drift}}(t) = -d_\kappa t, \qquad x_\kappa^{\mathrm{drift}}(t) = \kappa^{-t/2}, \qquad g_\kappa^{\mathrm{drift}}(t) = -d_\kappa. \tag{171}$$

Reconstructing the schedule using equation 168 gives equation 157, and

$$\min_{(\alpha,\beta)\in\widetilde{\mathcal{S}}} \max_{\lambda\in[1,\kappa]} L_\lambda^b[\alpha,\beta] = \frac{1}{2}\log\kappa. \tag{172}$$

Again, equality forces the prescribed worst-case profile, so the VP minimizer is unique for $\kappa > 1$.

Unlike the solution-optimal schedule, this schedule equalizes proportional contraction:

$$\frac{x_\kappa^{\mathrm{drift}}(t+h)}{x_\kappa^{\mathrm{drift}}(t)} = \kappa^{-h/2}.$$

It therefore distributes the required logarithmic contraction uniformly over time.

I.5.3 Connection to averaged squared Lipschitzness

Chen et al. (2025) propose the averaged squared Lipschitzness criterion

$$A_2[\alpha,\beta;\pi] \triangleq \int_0^1 \mathbb{E}_{X\sim p_t}\left[\|\nabla_x b_t(X)\|_{\mathrm{op}}^2\right] dt, \tag{173}$$

where $\|\cdot\|_{\mathrm{op}}$ is the Euclidean operator norm. Their Proposition 3.9, with smallest covariance eigenvalue κ^{-1}, gives exactly equation 157.

To see the connection with our minimax criterion, observe that for Gaussian targets the spatial Jacobian is independent of X, and

$$A_2[\alpha,\beta;\mathcal{N}(0,\Lambda^{-1})] = \int_0^1 \max_{\lambda\in\mathrm{spec}(\Lambda)} |g_\lambda(t)|^2\, dt. \tag{174}$$

Taking the worst case over the Gaussian spectral class and applying Cauchy–Schwarz gives

$$\begin{aligned}\sup_{I_d\preceq\Lambda\preceq\kappa I_d} A_2[\alpha,\beta;\mathcal{N}(0,\Lambda^{-1})] &= \int_0^1 |g_\kappa(t)|^2\, dt \\ &\geq \left|\int_0^1 g_\kappa(t)\, dt\right|^2 = \frac{1}{4}(\log\kappa)^2.\end{aligned} \tag{175}$$

The drift-optimal schedule attains equality because $g_\kappa^{\mathrm{drift}}(t) = -\frac{1}{2}\log\kappa$ is constant. Thus the same schedule minimizes both the worst-case instantaneous drift Lipschitz constant and its time-averaged square. For a fixed Gaussian target, the same conclusion holds when κ is its largest precision eigenvalue.

Their path-space KL analysis, after optimal diffusion tuning, separates statistical equivalence from numerical efficiency; their transfer formula also enables inference-time schedule changes without retraining Chen et al. (2025).

The same paper extends the bound to targets $\pi \propto e^{-V}$ with $I_d \preceq \nabla^2 V \preceq \kappa I_d$: under the drift schedule,

$$-\frac{1}{2}\log\kappa\, I_d \preceq \nabla_x b_t(x) \preceq 0, \qquad \|\nabla_x b_t(x)\|_{\mathrm{op}} \le \frac{1}{2}\log\kappa. \tag{176}$$

This is a uniform guarantee, not a claim that the schedule minimizes A_2 for every non-Gaussian target (Chen et al., 2025, Example 3.11).

I.6 Comparison with the elliptic schedule

Set $m \triangleq 1-\kappa^{-1}$ and $c_\kappa \triangleq 1-\kappa^{-1/2}$. The three schedules have particularly simple worst-case solution profiles:

$$x_\kappa^{\mathrm{ell}}(t) = \mathrm{dn}(K(m)t \mid m), \qquad x_\kappa^{\mathrm{sol}}(t) = 1-c_\kappa t, \qquad x_\kappa^{\mathrm{drift}}(t) = \kappa^{-t/2}. \tag{177}$$

They interpolate the same endpoints through elliptic, linear, and exponential profiles, respectively.

Let L^b_{ell}, L^b_{sol}, and L^b_{drift} denote the worst-case drift Lipschitz constants of these three schedules. Their values are

$$L^b_{\mathrm{ell}} = c_\kappa K(m), \qquad L^b_{\mathrm{sol}} = \sqrt{\kappa}-1, \qquad L^b_{\mathrm{drift}} = \frac{1}{2}\log\kappa. \tag{178}$$

Indeed, the solution-optimal schedule has

$$g_\kappa^{\mathrm{sol}}(t) = -\frac{c_\kappa}{1-c_\kappa t},$$

whose largest magnitude is $c_\kappa\sqrt{\kappa} = \sqrt{\kappa}-1$. For the elliptic schedule, differentiating $\log \mathrm{dn}(K(m)t \mid m)$ gives

$$g_\kappa^{\mathrm{ell}}(t) = -K(m)m\frac{\mathrm{sn}(K(m)t \mid m)\,\mathrm{cn}(K(m)t \mid m)}{\mathrm{dn}(K(m)t \mid m)}, \tag{179}$$

whose largest magnitude is $K(m)(1-\sqrt{1-m}) = c_\kappa K(m)$.

Since $K(m) \sim \frac{1}{2}\log(16\kappa)$ as $\kappa \to \infty$, the elliptic schedule is asymptotically drift-Lipschitz optimal:

$$\frac{L^b_{\mathrm{ell}}}{L^b_{\mathrm{drift}}} = \frac{2(1-\kappa^{-1/2})K(m)}{\log\kappa} \longrightarrow 1. \tag{180}$$

In contrast, the solution-optimal schedule has an algebraically growing drift Lipschitz constant. The distinction comes from

$$g_\kappa(t) = \frac{\dot{x}_\kappa(t)}{x_\kappa(t)}:$$

keeping the absolute speed small does not keep the relative speed small as the solution contracts.

These criteria describe different aspects of regularity, rather than optimality for a particular numerical discretization. In particular, minimizing a first-order Lipschitz quantity alone does not determine which schedule–discretization pair gives the smallest sampling error.

Endpoint regularity and the limit $\kappa = 1$. For both Lipschitz schedules, β_t behaves as a constant times $\sqrt{t}$ near $t = 0$, and α_t behaves as a constant times $\sqrt{1-t}$ near $t = 1$. Thus the schedules are not $C^1([0,1])$, although their Gaussian solution and drift multipliers are smooth on the closed interval. This is why their minimizers are stated in the enlarged class $\widetilde{\mathcal{S}}$ defined in Appendix I.4.

As $\kappa \downarrow 1$, both Lipschitz schedules converge to $(\sqrt{1-t}, \sqrt{t})$. The elliptic schedule instead converges to $(\cos(\pi t/2), \sin(\pi t/2))$, since $m \to 0$ and $K(m) \to \pi/2$. At $\kappa = 1$, every VP schedule gives $x_1(t) \equiv 1$ and $g_1(t) \equiv 0$. The square-root schedule is therefore the continuous extension of the two optimizer families, not a uniquely selected optimizer of the degenerate $\kappa = 1$ problem.

I.7 Relation between interpolant schedules and noise schedules

Let $x_t = \alpha_t z + \beta_t x_1$ follow a VP interpolant. For every $t \in (0,1)$, division by the deterministic data coefficient gives the corresponding VE-corrupted variable

$$y_t \triangleq \frac{x_t}{\beta_t} = x_1 + \sigma_t z, \qquad \sigma_t \triangleq \frac{\alpha_t}{\beta_t}. \tag{181}$$

Conversely, a positive monotone noise clock σ_t determines the unique first-quadrant VP coefficients

$$\alpha_t = \frac{\sigma_t}{\sqrt{1+\sigma_t^2}}, \qquad \beta_t = \frac{1}{\sqrt{1+\sigma_t^2}}. \tag{182}$$

Thus the two parameterizations describe the same signal-to-noise clock, but their states differ by the time-dependent dilation $x_t = \beta_t y_t$. In particular, if X_t solves the VP probability-flow ODE $\dot{X}_t = b_t(X_t)$, then $Y_t = X_t/\beta_t$ satisfies

$$\dot{Y}_t = \frac{1}{\beta_t} b_t(\beta_t Y_t) - \frac{\dot{\beta}_t}{\beta_t} Y_t. \tag{183}$$

With our noise-to-data orientation, σ_t decreases as interpolant time increases. Hence a VE grid $\sigma_{\min} = \sigma_0 < \cdots < \sigma_n = \sigma_{\max}$ corresponds to the reverse-ordered interpolant times $t_i = \sigma^{-1}(\sigma_i)$. Finite VE endpoints restrict the continuous interpolant to the associated interior time interval. Finally, the continuous clock, the discrete query grid, and the numerical update rule are separate choices: transforming a VP clock into VE coordinates does not by itself turn one ODE discretization into another. The AYS correspondence in Appendix I.4 is an instance of precisely this clock-and-grid conversion.

J Probability-Flow Discretizations

J.1 Derivation of elliptic finite-difference discretization

Recall that if (α_t, β_t) satisfy $\alpha_t^2 + \beta_t^2 = 1$ (a *variance preserving* schedule), the drift terms in (32) can be simplified as

$$b_t(x) = \frac{\dot{\beta}_t}{\beta_t}\left(x + s_t(x)\right). \tag{184}$$

Since

$$\frac{\dot{\beta}_t}{\beta_t} = \frac{d}{dt}(\log \beta_t) = \frac{1}{2}\frac{d}{dt}(\log \beta_t^2) = \frac{1}{2\beta_t^2}\frac{d}{dt}(\beta_t^2),$$

we can rewrite the drift term to obtain

$$b_t(x) = \frac{\dot{\beta}_t}{\beta_t}\left(x + s_t(x)\right) = \frac{1}{2\beta_t^2}\frac{d}{dt}(\beta_t^2)\left(x + s_t(x)\right) \approx \frac{1}{2\beta_t^2}\frac{(\beta_{t+(h/2)}^2 - \beta_t^2)}{(h/2)}\left(x + s_t(x)\right) \tag{185}$$

where in the last step we used a half-step finite-difference scheme to approximate the derivative. This yields the discretization described in Definition 3.3.

J.2 Catalog of Discretizations

Notation. For a budget of r score queries,

$$h \triangleq \frac{2}{2r+1} \qquad \text{grid spacing,}$$

$$t_j \triangleq \frac{2j-1}{2r+1}, \quad j \in [r+1] \qquad \text{query times } t_1, \ldots, t_r \text{ and endpoint } t_{r+1} = 1, \tag{186}$$

$$x^0 \sim \mathcal{N}(0, I) \qquad \text{initial state at } t_1;\ x^j \text{ is the state after } j \text{ updates,}$$

$$\hat{x}_0(x,t) \triangleq -\alpha_t\, s_t(x), \quad \hat{x}_1(x,t) \triangleq \frac{x + \alpha_t^2\, s_t(x)}{\beta_t} \qquad \text{noise and data predictions (one score query each),} \tag{187}$$

$$\rho_t \triangleq \log \frac{\beta_t}{\alpha_t} \qquad \text{half-log-SNR.}$$

One query per step. These methods query once at each of $t_1, \dots, t_r$ and return x^r. Unless stated otherwise, $j = 1, \dots, r$.

- **Elliptic Finite-Difference**:
$$x^j = x^{j-1} + \frac{\beta^2_{t_j+h/2} - \beta^2_{t_j}}{\beta^2_{t_j}}\big(x^{j-1} + s_{t_j}(x^{j-1})\big). \tag{188}$$
- **Euler:**
$$x^j = x^{j-1} + h\, b_{t_j}(x^{j-1}). \tag{189}$$
- **DDIM** (Song et al., 2022):
$$x^j = \alpha_{t_{j+1}}\hat{x}_0(x^{j-1}, t_j) + \beta_{t_{j+1}}\hat{x}_1(x^{j-1}, t_j). \tag{190}$$
- **DPM-Solver++(2M)** (Lu et al., 2022b): let $d_j \triangleq \hat{x}_1(x^{j-1}, t_j)$ be the data prediction at step j and $\delta_j \triangleq \rho_{t_{j+1}} - \rho_{t_j}$ the step in half-log-SNR. For $j = 1, \dots, r-1$,
$$x^j = \frac{\alpha_{t_{j+1}}}{\alpha_{t_j}}x^{j-1} + \beta_{t_{j+1}}\big(1 - e^{-\delta_j}\big)D_j, \qquad D_j = \begin{cases} d_1, & j = 1, \\ \big(1 + \frac{\delta_j}{2\delta_{j-1}}\big)d_j - \frac{\delta_j}{2\delta_{j-1}}d_{j-1}, & j \geq 2, \end{cases} \tag{191}$$
and the last step is the DDIM readout $x^r = d_r$, since $\delta_r = +\infty$.

Multistage methods. An m-stage method (with $r > m$) takes $M \triangleq \lfloor (r-1)/m \rfloor$ macro-steps on the uniform macrogrid
$$\tau_n \triangleq t_1 + n\Delta, \qquad \Delta \triangleq \frac{t_r - t_1}{M}, \qquad n = 0, \dots, M, \tag{192}$$
so that $\tau_M = t_r$, with $\tau_n = t_{mn+1}$ when $m \mid r-1$. It then covers the last interval $[t_r, 1]$ with one Euler step,
$$x^{M+1} = x^M + h\, b_{t_r}(x^M), \tag{193}$$
and returns x^{M+1}. This uses $mM + 1 \leq r$ queries and never queries the model at $t = 1$, where the drift may blow up. Within a macro-step, write $x = x^n$, $x' = x^{n+1}$, $\tau = \tau_n$, and $\tau' = \tau_{n+1}$.

- **Heun** ($m = 2$):
$$k_1 = b_\tau(x), \qquad k_2 = b_{\tau'}(x + \Delta k_1), \qquad x' = x + \frac{\Delta}{2}(k_1 + k_2). \tag{194}$$
- **RK4** ($m = 4$):
$$\begin{aligned} &k_1 = b_\tau(x), && k_2 = b_{\tau+\Delta/2}\big(x + \tfrac{\Delta}{2}k_1\big), \\ &k_3 = b_{\tau+\Delta/2}\big(x + \tfrac{\Delta}{2}k_2\big), && k_4 = b_{\tau'}(x + \Delta k_3), \\ &x' = x + \tfrac{\Delta}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned} \tag{195}$$

K Log-Concave Setting

K.1 Holomorphic Extension of the Solution

Example K.1 (A pathological log-concave distribution with failed analyticity at $t = 1$)**.** Fix $\kappa > 1$ and set
$$\lambda = \frac{\kappa + 1}{2}, \quad \phi_{1/\lambda}(x) = \sqrt{\frac{\lambda}{2\pi}}\exp\left(-\lambda\frac{x^2}{2}\right), \quad \text{and} \quad \psi(x) = (1 - x^2)^4_+ = (1 - x^2)^4\, \mathbb{1}_{|x| \leq 1} \tag{196}$$
Consider the following perturbation of the Gaussian density $\phi_{1/\lambda}$:
$$p(x) = \phi_{1/\lambda}(x) + \eta\psi'(x) = \phi_{1/\lambda}(x) - 8\eta x(1 - x^2)^3\mathbb{1}_{|x|<1}, \tag{197}$$
where $\eta > 0$ is a sufficiently small parameter chosen such that $p(x) \geq \frac{1}{2}\phi_{1/\lambda}(x) > 0$ everywhere. Moreover, since the perturbation $\psi'(x)$ is an odd function, $\int_{-\infty}^{\infty} p(x)dx = 1$ and thus $p(x)$ is a probability density function supported on the entire real line.

Proof. Proof follows three steps:

i. Representing any SI flow map in terms of a canonical VP-SI flow map. The SI $\widetilde{X}_t$ between standard normal and the smoothed distribution $\widetilde{p}$ under a schedule (α_t, β_t) can be expressed as an SI between the standard normal and the unsmoothed distribution p under a different schedule $(\widetilde{\alpha}_t, \widetilde{\beta}_t)$ as follows

$$\widetilde{X}_t \overset{(d)}{=} \alpha_t \widetilde{Z} + \beta_t \widetilde{X}_t \overset{(d)}{=} \underbrace{\sqrt{\alpha_t^2 + \beta_t^2 \widetilde{\sigma}^2}}_{\triangleq \widetilde{\alpha}_t} Z + \underbrace{\beta_t \sqrt{1-\widetilde{\sigma}^2}}_{\triangleq \widetilde{\beta}_t} X_1. \tag{198}$$

Note that while the initialization of this new schedule satisfies $\widetilde{\alpha}_0 = 1$ and $\widetilde{\beta}_0 = 0$, the end points at time $t = 1$ are $\widetilde{\alpha}_1 = \widetilde{\sigma}$ and $\widetilde{\beta}_1 = \sqrt{1-\widetilde{\sigma}^2}$. . This interpolant can be expressed in terms of the canonical VP-SI $X_s^\star \overset{(d)}{=} \sqrt{1-s^2} Z + s X_1$ as follows:

$$\widetilde{X}_t \overset{(d)}{=} \sqrt{\widetilde{\alpha}_t^2 + \widetilde{\beta}_t^2} \left(\frac{\widetilde{\alpha}_t}{\sqrt{\widetilde{\alpha}_t^2 + \widetilde{\beta}_t^2}} Z + \frac{\widetilde{\beta}_t}{\sqrt{\widetilde{\alpha}_t^2 + \widetilde{\beta}_t^2}} X_1 \right), \tag{199}$$

$$\overset{(d)}{=} \sqrt{\alpha_t^2 + \beta_t^2}\, X^\star_{s(t)}, \quad \text{where} \quad s(t) \triangleq \frac{\beta_t \sqrt{1-\widetilde{\sigma}^2}}{\sqrt{\alpha_t^2 + \beta_t^2}}. \tag{200}$$

Note that the scaling term becomes simply one for VP schedules. Therefore, We will consider VP schedules for simplicity, without loss of generality. Similarly, the flow map $t \mapsto \widetilde{\Phi}_t(z)$ for any interpolant (α_t, β_t) can be expressed in terms of the flow map $s \mapsto \Phi^\star_s(z)$ for the canonical VP interpolant by a time reparameterization:

$$\widetilde{\Phi}_t(z) = \Phi^\star_{s(t)}(z), \qquad s(t) = \beta_t \sqrt{1-\widetilde{\sigma}^2}. \tag{201}$$

As the canonical VP-SI flow map $s \mapsto \Phi^\star_s(z)$ is independent of the choice of schedule (α_t, β_t), the holomorphic extension of $t \mapsto \widetilde{\Phi}_t(z)$ can be constructed by composing the holomorphic extensions of $t \mapsto s(t) = \beta_t \sqrt{1-\widetilde{\sigma}^2}$, which is determined only by the schedule, and $s \mapsto \Phi^\star_s(z)$, which is schedule independent and determined by the underlying distribution.

ii. Schedule-independent holomorphic extension of canonical VP-SI flow map. Lemma K.2 implies that $s \mapsto \Phi^\star_s(z)$ extends holomorphically to a ball around each real $s_\circ \in (-1, 1)$. Stitching together all these balls yields that $s \mapsto \Phi^\star_s(z)$ extends holomorphically to

$$\mathcal{D}_{\Phi^\star}(\kappa, d, B) \triangleq \bigcup_{s_\circ \in (-1,1)} \left\{ s \in \mathbb{C} \mid |s - s_\circ| < \frac{1 - s_\circ^2}{30(\sqrt{B^2 + d\log(\kappa)} + \sqrt{d})} \right\}. \tag{202}$$

iii. Elliptic schedule maps the domain $\mathcal{D}_{\Phi^\star}(\kappa, d, B)$ to an infinite strip. Consider the elliptic schedule: $\alpha_t = \mathrm{cn}(K(m)t \mid m)$ and $\beta_t = \mathrm{sn}(K(m)t \mid m)$ with $m = 1 - \kappa^{-1}$. It is known that $u \mapsto \mathrm{sn}(u \mid m)$ is meromorphic (holomorphic except at isolated poles) on the complex domain with order-1 simple poles at lattice points $u_{n_1,n_2} = 2n_1 K(m) + i(2n_2+1)K(1-m)$ for $n_1, n_2 \in \mathbb{Z}$. Therefore, $t \mapsto s(t) = \widetilde{\beta}_t = \mathrm{sn}(K(m)t \mid m)\sqrt{1-\widetilde{\sigma}^2}$ is analytic everywhere except at lattice points $t_{n_1,n_2} = 2n_1 + i(2n_2+1)\frac{K(\kappa^{-1})}{K(1-\kappa^{-1})}$ for $n_1, n_2 \in \mathbb{Z}$, hence, it is holomorphic on the strip $\mathcal{D}_s(\kappa) \triangleq \left\{ t \in \mathbb{C} \mid |\operatorname{Im}(t)| < \frac{K(\kappa^{-1})}{K(1-\kappa^{-1})} \right\}$.

We would like to derive a local bound of the form

$$\frac{|s(t) - s(t_\circ)|}{1 - s(t_\circ)^2} \le f(|t - t_\circ|), \tag{203}$$

for $t \in \mathcal{D}_s(\kappa)$ and real $t_\circ \in \mathbb{R}$ where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is some function and find a large $\gamma(\kappa, d, B) > 0$ such that $|\delta t| < \gamma(\kappa, d, B)$ implies $f(|\delta t|) \le (30(\sqrt{B^2 + d\log(\kappa)} + \sqrt{d}))^{-1}$, and hence $s(t) \in \mathcal{D}_{\Phi^\star}(\kappa, d, B)$, i.e., analyticity of $\widetilde{\Phi}_t = \Phi^\star_{s(t)}$ for all $t \in \mathbb{C}$ with $|\operatorname{Im}(t)| < \min(\gamma(\kappa, d, B), \frac{K(\kappa^{-1})}{K(1-\kappa^{-1})})$.

Note that the derivative $\dot{s}(t)$ of s satisfies

$$\dot{s}(t) = \sqrt{1-\widetilde{\sigma}^2} K(m)\, \mathrm{cn}(K(m)t \mid m)\, \mathrm{dn}(K(m)t \mid m). \tag{204}$$

Integrating along the linear segment $[0,1] \ni \theta \mapsto \tau(\theta) \triangleq t_\circ + \theta(t - t_\circ)$ from real $t_\circ \in \mathbb{R}$ to complex $t \in \mathcal{D}_s(\kappa)$ gives

$$|s(t) - s(t_\circ)| = \sqrt{1-\widetilde{\sigma}^2} K(m) \left| \int_0^1 \operatorname{cn}(K(m)\tau(\theta) \mid m) \operatorname{dn}(K(m)\tau(\theta) \mid m)(t - t_\circ) d\theta \right|, \quad (205)$$

$$\leq \sqrt{1-\widetilde{\sigma}^2} K(m)|t - t_\circ| \int_0^1 |\operatorname{cn}(K(m)\tau(\theta) \mid m)| \, |\operatorname{dn}(K(m)\tau(\theta) \mid m)| d\theta. \quad (206)$$

Defining $\widetilde{m} \triangleq 1 - \widetilde{\sigma}^2$, $r \triangleq \min\left(1, \sqrt{\frac{1-\widetilde{m}}{1-m}}\right) = \min(1, \widetilde{\sigma}\sqrt{\kappa})$ and defining $\Delta \triangleq r^{-1} K(m) \, |t - t_\circ|$, Lemma K.5 implies that

$$|\operatorname{cn}(K(m)t \mid m)| \leq \frac{\sqrt{1 - s(t_\circ)^2}}{1 - \Delta}, \quad |\operatorname{dn}(K(m)t \mid m)| \leq \frac{r^{-1}\sqrt{1 - s(t_\circ)^2}}{1 - \Delta}, \quad (207)$$

whenever $\Delta < 1$ where all elliptic functions are holomorphic since

$$|\operatorname{Im}(t)| = |t - \Re(t)| < \frac{r}{K(m)} < \frac{\pi/2}{K(m)} \leq \frac{K(1-m)}{K(m)}. \quad (208)$$

Thus, the following holds whenever $\Delta = r^{-1} K(m) \, |t - t_\circ| < 1$:

$$\frac{|s(t) - s(t_\circ)|}{1 - s(t_\circ)^2} \leq \sqrt{1-\widetilde{\sigma}^2} \int_0^1 \frac{\Delta d\theta}{(1 - \Delta\theta)^2} = \sqrt{1-\widetilde{\sigma}^2} \frac{\Delta}{1 - \Delta}. \quad (209)$$

In particular, the condition

$$|t - t_\circ| < \gamma(\kappa, d, B) \triangleq \frac{r}{K(m)(1 + \sqrt{1-\widetilde{\sigma}^2} D(\kappa, d, B))} = \frac{\min(1, \widetilde{\sigma}\sqrt{\kappa})}{K(1-\kappa^{-1})(1 + \sqrt{1-\widetilde{\sigma}^2} D(\kappa, d, B))}, \quad (210)$$

ensures that

$$\frac{|s(t) - s(t_\circ)|}{1 - s(t_\circ)^2} \leq \sqrt{1-\widetilde{\sigma}^2} \frac{\Delta}{1 - \Delta} \leq \frac{1}{D(\kappa, d, B)}. \quad (211)$$

Since this holds for all real $t_\circ \in \mathbb{R}$, the flow map $t \mapsto \widetilde{\Phi}_t(z)$ is holomorphic on the strip $\mathcal{D}_{\widetilde{\Phi}}(\kappa, d, B) \triangleq \{t \in \mathbb{C} \mid |\operatorname{Im}(t)| < \gamma(\kappa, d, B)\}$.

□

Remark.

Lemma K.2 (Holomorphic extension of the canonical VP solution around a real time point)**.** *Let $X \sim p(x) \propto \exp(-V(x))$ be log-concave with $V \in \mathcal{C}^2$ and $I_d \preceq \nabla^2 V \preceq \kappa I_d$ for $\kappa \geq 1$. Consider the following canonical interpolant*

$$X_s^\star \overset{(d)}{=} \sqrt{1 - s^2} Z + sX, \quad (212)$$

and denote by $[-1, 1] \ni s \mapsto \Phi_s^\star(z)$ the corresponding flow map which satisfies

$$\frac{d\Phi_s^\star(z)}{ds} = b^\star(\Phi_s^\star(z), s) = \frac{1}{1 - s^2} \left(\widehat{X}_1^\star(\Phi_s^\star(z), s) - s\Phi_s^\star(z) \right), \quad \Phi_0^\star(z) = z. \quad (213)$$

where $(x, s) \mapsto \widehat{X}_1^\star(x, s)$ is the denoiser defined as

$$\widehat{X}_1^\star(x, s) \triangleq \mathbb{E}\left[X_1 \mid \sqrt{1 - s^2} Z + sX_1 = x \right]. \quad (214)$$

Fix a real $s_\circ \in (-1, 1)$. Then, for all bounded initializations $\|z\| \leq B$, the flow map $s \mapsto \Phi_s^\star(z)$ extends holomorphically to a ball around $s_\circ$, namely,

$$B_{\rho_\circ}(s_\circ) \triangleq \{s \in \mathbb{C} \mid |s - s_\circ| < \rho_\circ\}, \quad \textit{where} \quad \rho_\circ = \frac{1 - s_\circ^2}{30(\sqrt{B^2 + d\log(\kappa)} + \sqrt{d})}. \quad (215)$$

Proof.

$$\widehat{X}_1^\star(x,s) \triangleq \mathbb{E}\left[X_1 \mid \sqrt{1-s^2}Z + sX_1 = x\right] = \int x_1 p_{1|s}(x_1 \mid x) dx_1, \tag{216}$$

where $p_{1|s}(x_1 \mid x)$ is the posterior density

$$p_{1|s}(x_1 \mid x) = \frac{\exp\left(-\frac{\|x-sx_1\|^2}{2(1-s^2)} - V(x_1)\right)}{\int \exp\left(-\frac{\|x-sx_1'\|^2}{2(1-s^2)} - V(x_1')\right) dx_1'}. \tag{217}$$

The flow map $\widetilde{\Phi}_t(z)$ for the smoothed variable $\widetilde{X}_1$ under a VP schedule (α_t, β_t) can be expressed as

$$\widetilde{\Phi}_t(z) = \Phi^\star_{s(t)}(z), \quad s(t) \triangleq \widetilde{\beta}_t = \beta_t\sqrt{1-\widetilde{\sigma}^2}. \tag{218}$$

Notice that the canonical VP-SI flow map $s \mapsto \Phi^\star_s(z)$ is independent of the choice of schedule (α_t, β_t), which acts solely as a time reparameterization through $t \mapsto s(t)$. Therefore, the holomorphic extension of $t \mapsto \widetilde{\Phi}_t(z)$ can be constructed by composing the holomorphic extensions of $t \mapsto s(t)$ and $s \mapsto \Phi^\star_s(z)$.

Note that the domain of the canonical flow map $s \mapsto \Phi^\star_s(z)$ can be extended from $[0,1)$ to $(-1,1)$ by reflection. The canonical drift $(x,s) \mapsto b^\star(x,s)$ can be extended to the complex domain as a meromorphic function with singularities as the posterior $(x,s) \mapsto p_{1|s}(x_1 \mid x)$ is a ratio of two holomorphic functions in $\mathbb{C}^d \times \mathbb{C}/\{-1,+1\}$.

Pick any $s_\circ \in (-1,1)$ and let $x_\circ = \Phi^\star_{s_\circ}(z)$. The posterior density $p_{1|s_\circ}(x_1 \mid x_\circ)$ at $(x_\circ, s_\circ)$ is a strongly log-concave density since

$$-\nabla^2_{x_1} \log p_{1|s_\circ}(x_1 \mid x_\circ) = \nabla^2_{x_1}\left(\frac{\|x_\circ - s_\circ x_1\|^2}{2(1-s_\circ^2)} + V(x_1)\right) \succeq \left(\frac{s_\circ^2}{1-s_\circ^2} + 1\right) I_d = \frac{1}{1-s_\circ^2} I_d. \tag{219}$$

Moreover, it admits a unique mode at

$$\overline{x}_{1|s_\circ} \triangleq \operatorname*{arg\,min}_{x_1 \in \mathbb{R}^d}\left\{\frac{\|x_\circ - s_\circ x_1\|^2}{2(1-s_\circ^2)} + V(x_1)\right\} = \operatorname{prox}_{\frac{1-s_\circ^2}{s_\circ^2}V}(x_\circ/s_\circ), \tag{220}$$

so that $\nabla_{x_1} \log p_{1|s_\circ}(\overline{x}_{1|s_\circ} \mid x_\circ) = 0$. For a conditional sample $X_{1|s_\circ} \sim p_{1|s_\circ}(\cdot \mid x_\circ)$, define its mode-centered and scaled version

$$\Xi_\circ \triangleq \frac{X_{1|s_\circ} - \overline{x}_{1|s_\circ}}{\sqrt{1-s_\circ^2}} \sim \pi_\circ \quad \text{where} \quad \pi_\circ(\xi_\circ) = (1-s_\circ^2)^{\frac{d}{2}}\, p_{1|s_\circ}(\sqrt{1-s_\circ^2}\,\xi_\circ + \overline{x}_{1|s_\circ}). \tag{221}$$

The distribution $\pi_\circ$ has its mode at 0 and is 1-strongly log-concave, i.e.,

$$\nabla_{\xi_\circ} \log \pi_\circ(0) = 0 \quad \text{and} \quad -\nabla^2_{\xi_\circ} \log \pi_\circ(\xi_\circ) \succeq I_d. \tag{222}$$

Using the transformation $\xi_\circ = \frac{x_1 - \overline{x}_{1|s_\circ}}{\sqrt{1-s_\circ^2}}$, the posterior density $p_{1|s}(x_1 \mid x)$ at complex valued $(x,s) \in \mathbb{C}^d \times \mathbb{C}/\{-1,+1\}$ can be expressed an exponential tilting of $\pi_\circ(\xi_\circ)$ as

$$p_{1|s}(x_1 \mid x) \propto \pi_\circ(\xi_\circ) \exp\left(\xi_\circ^\top \chi_\circ(x,s) + \eta_\circ(s)\|\xi_\circ\|^2\right), \tag{223}$$

$$\text{where} \quad \begin{aligned} \chi_\circ(x,s) &\triangleq \sqrt{1-s_\circ^2}\left(\frac{s(x - s\overline{x}_{1|\circ})}{1-s^2} - \frac{s_\circ(x_\circ - s_\circ \overline{x}_{1|\circ})}{1-s_\circ^2}\right), \\ \eta_\circ(s) &\triangleq -\frac{1-s_\circ^2}{2}\left(\frac{s^2}{1-s^2} - \frac{s_\circ^2}{1-s_\circ^2}\right) = \frac{s_\circ^2 - s^2}{2(1-s^2)}. \end{aligned} \tag{224}$$

Thus, the denoiser $(x,s) \mapsto \widehat{X}_1^\star(x,s)$ at complex values can be expressed as a shifted and scaled ratio of expectations under $\pi_\circ$:

$$\widehat{X}_1^\star(x,s) = \overline{x}_{1|\circ} + \sqrt{1-s_\circ^2}\frac{\mathbb{E}_{\Xi_\circ \sim \pi_\circ}\left[\Xi_\circ \exp\left(\Xi_\circ^\top \chi_\circ(x,s) + \eta_\circ(s)\|\Xi_\circ\|^2\right)\right]}{\mathbb{E}_{\Xi_\circ \sim \pi_\circ}\left[\exp\left(\Xi_\circ^\top \chi_\circ(x,s) + \eta_\circ(s)\|\Xi_\circ\|^2\right)\right]}, \tag{225}$$

whenever the denominator is non-zero.

Notice that Tweedie's identity implies

$$\frac{x_\circ - s_\circ \widehat{X}_1^\star(x_\circ, s_\circ)}{\sqrt{1-s_\circ^2}} = -\sqrt{\tau_\circ} s_{\tau_\circ}^{VE}(x_\circ/s_\circ) \quad \text{where} \quad \tau_\circ = \frac{1-s_\circ^2}{s_\circ^2}. \tag{226}$$

Since $x_\circ = \Phi_{s_\circ}^\star(z)$, $x_\circ/s_\circ = \Phi_{\tau_\circ}^{VE}(z)$ where $\tau \mapsto \Phi_\tau^{VE}(z)$ is the reverse flow map for the VE PF-ODE. Thus, Lemma K.3 implies that

$$\|x_\circ - s_\circ \widehat{X}_1^\star(x_\circ, s_\circ)\| = \sqrt{1-s_\circ^2}\sqrt{\tau_\circ}\|s_{\tau_\circ}^{VE}(\Phi_{\tau_\circ}^{VE}(z))\| \le \sqrt{1-s_\circ^2}\sqrt{\|z\|^2 + d\log(\kappa)}. \tag{227}$$

$$\widehat{X}_1^\star(x_\circ, s_\circ) - \overline{x}_{1|s_\circ} = \mathbb{E}[X_1 - \overline{x}_{1|s_\circ} \mid \sqrt{1-s_\circ^2}Z + s_\circ X_1 = x_\circ] = \sqrt{1-s_\circ^2}\,\mathbb{E}_{\pi_\circ}[\Xi_\circ] \tag{228}$$

$$\|\widehat{X}_1^\star(x_\circ, s_\circ) - \overline{x}_{1|s_\circ}\| \le \sqrt{1-s_\circ^2}\,\sqrt{\mathbb{E}_{\pi_\circ}[\|\Xi_\circ\|^2]} \tag{229}$$

□

Lemma K.3 (VE score is bounded for log-concave targets)**.** *For any $z \in \mathbb{R}^d$, let $[0,\infty) \ni \tau \mapsto \Phi_\tau^{VE}(z)$ denote the reverse flow map for the VE PF-ODE such that*

$$\frac{d\Phi_\tau^{VE}(z)}{d\tau} = -\frac{1}{2}s_\tau^{VE}(\Phi_\tau^{VE}(z)), \quad \textit{and} \quad \lim_{\tau\to\infty}\frac{\Phi_\tau^{VE}(z)}{\sqrt{\tau}} = z. \tag{230}$$

Then, for all $\tau > 0$, we have that

$$\|s_\tau^{VE}(\Phi_\tau^{VE}(z))\| \le \sqrt{\frac{\|z\|^2 + d\log(\kappa)}{\tau}} \tag{231}$$

Proof. Since $p \propto \exp(-V(x))$ is a log-concave distribution, the noisy distribution $p_\tau^{VE} = p * N(0, \tau I_d)$ is also log-concave. In fact, Tweedie's formula and posterior identities yield quantitative bounds

$$\frac{1}{\tau+1}I_d \preceq \nabla^2 V_\tau \preceq \frac{\kappa}{\kappa\tau+1} \tag{232}$$

for all $\tau \ge 0$, where $V_\tau(x) \triangleq -\log p_\tau^{VE}(x)$. Since V_τ is Lipschitz, we get the gradient estimate

$$\|\nabla V_\tau(x)\|^2 \le \frac{2}{\tau + \kappa^{-1}}\left(V_\tau(x) - \min V_\tau\right), \tag{233}$$

which can be used to get a bound on $s_\tau^{VE}(\Phi_\tau^{VE}(z)) = -V_\tau(\Phi_\tau^{VE}(z))$ as

$$\|s_\tau^{VE}(\Phi_\tau^{VE}(z))\|^2 \le \frac{2}{\tau+\kappa^{-1}}\left(-\log p_\tau^{VE}(\Phi_\tau^{VE}(z)) + \log\max_x p_\tau^{VE}(x)\right). \tag{234}$$

Therefore, we need a lower bound on $\log p_\tau^{VE}(\Phi_\tau^{VE}(z))$ and an upper bound on $\max_x p_\tau^{VE}(x)$.

i. Lower bound on $\log p_\tau^{VE}(\Phi_\tau^{VE}(z))$**.** We can utilize its evolution in τ, which is governed by the following ODE

$$\frac{d}{d\tau}\log p_\tau^{VE}(\Phi_\tau^{VE}(z)) = \frac{1}{p_\tau^{VE}(\Phi_\tau^{VE}(z))}\left(\nabla_x p_\tau^{VE}(\Phi_\tau^{VE}(z))^\top \frac{d\Phi_\tau^{VE}(z)}{d\tau} + \partial_\tau p_\tau^{VE}(\Phi_\tau^{VE}(z))\right), \tag{235}$$

$$= \frac{1}{2}\left(-\|s_\tau^{VE}(\Phi_\tau^{VE}(z))\|^2 + \nabla_x^\top s_\tau^{VE}(\Phi_\tau^{VE}(z)) + \|s_\tau^{VE}(\Phi_\tau^{VE}(z))\|^2\right), \tag{236}$$

$$= \frac{1}{2}\operatorname{tr}(\nabla s_\tau^{VE}(\Phi_\tau^{VE}(z))), \tag{237}$$

where equation 235 follows from chain rule and 236 follows from the continuity equation

$$\partial_\tau p_\tau^{VE}(x) + \nabla\cdot\left(-\frac{1}{2}s_\tau^{VE}(x)p_\tau^{VE}(x)\right) = \partial_\tau p_\tau^{VE}(x) - \frac{1}{2}\Delta p_\tau^{VE}(x) = 0. \tag{238}$$

Since $\nabla s_\tau^{VE}(\Phi_\tau^{VE}(z)) = -\nabla^2 V_\tau(\Phi_\tau^{VE}(z))$, the lower curvature bound in 232 implies

$$\frac{d}{d\tau}\log p_\tau^{VE}(\Phi_\tau^{VE}(z)) \leq -\frac{d}{2(\tau+1)}, \tag{239}$$

By Gronwall's inequality, integrating both sides yields

$$\log p_\tau^{VE}(\Phi_\tau^{VE}(z)) \geq \log p_T^{VE}(\Phi_T^{VE}(z)) + \frac{d}{2}\log\frac{T+1}{\tau+1}, \tag{240}$$

for any $0 \leq \tau \leq T$. In particular, this inequality also holds as $T \longrightarrow \infty$. To get the limiting term, note that

$$(2\pi T)^{d/2} p_T^{VE}(\Phi_T^{VE}(z)) = \mathbb{E}_{X\sim p}\left[\exp\left(-\frac{1}{2}\left\|\frac{\Phi_\tau^{VE}(z) - X}{\sqrt{T}}\right\|^2\right)\right] \longrightarrow \exp(-\|z\|^2/2) \tag{241}$$

as $T \longrightarrow \infty$ by the entrance condition $\lim_{T\to\infty} T^{-1/2}\Phi_T^{VE}(z) = z$. Therefore,

$$\log p_T^{VE}(\Phi_T^{VE}(z)) + \frac{d}{2}\log(T) \longrightarrow -\frac{\|z\|^2}{2} - \frac{d}{2}\log(2\pi) \quad \text{as} \quad T \longrightarrow \infty. \tag{242}$$

Together with equation 240, this gives the lower bound

$$\log p_\tau^{VE}(\Phi_\tau^{VE}(z)) \geq -\frac{\|z\|^2 + d\log(2\pi(\tau+1))}{2}. \tag{243}$$

ii. Upper bound on $\max_x p_\tau^{VE}(x)$. Note that the upper Hessian bound in 232 implies that

$$V_\tau(x) \leq V_\tau(x_\tau^\star) + \frac{1}{2(\tau+\kappa^{-1})}\|x - x_\tau^\star\|^2, \tag{244}$$

where $x_\tau^\star = \arg\min_{x'} V_\tau(x')$. Negating and exponentiating both sides yield

$$p_\tau^{VE}(x) \geq \max_{x'} p_\tau^{VE}(x')\ \exp\left(-\frac{\|x - x_\tau^\star\|^2}{2(\tau+\kappa^{-1})}\right). \tag{245}$$

Integrating both sides gives the upper bound

$$\max_{x'} p_\tau^{VE}(x') \leq \left(2\pi(\tau+\kappa^{-1})\right)^{-d/2}. \tag{246}$$

Substituting the lower bound 243 and upper bound 246 into 234 yields

$$\|s_\tau^{VE}(\Phi_\tau^{VE}(z))\|^2 \leq \frac{1}{\tau+\kappa^{-1}}\left(\|z\|^2 + d\log\left(\frac{\tau+1}{\tau+\kappa^{-1}}\right)\right). \tag{247}$$

Applying the inequalities $\frac{1}{\tau+\kappa^{-1}} \leq \frac{1}{\tau}$ and $\frac{\tau+1}{\tau+\kappa^{-1}} \leq \kappa$ gives the inequality 231. □

Lemma K.4 (Complex-valued tilting). *Let $X \sim p(x) \propto \exp(-V(x))$ be such that $\nabla^2 V \succeq I_d$ and $\nabla V(0) = 0$, i.e., the mode is at 0. Then, for complex valued $\chi \in \mathbb{C}^d$ and $\eta \in \mathbb{C}$ satisfying*

$$\|\chi\| < \frac{1}{2} \quad \textit{and} \quad |\eta| < \frac{1}{16\sqrt{d}}, \tag{248}$$

we have that

$$\mathbb{E}\left[e^{X^\top\chi + \eta\|X\|^2}\right] \neq 0 \quad \textit{and} \quad \left\|\frac{\mathbb{E}\left[Xe^{X^\top\chi+\eta\|X\|^2}\right]}{\mathbb{E}\left[e^{X^\top\chi+\eta\|X\|^2}\right]}\right\| \leq 2\sqrt{d}. \tag{249}$$

Proof. Let us split the exponent into its real and imaginary parts $x^\top\chi + \eta\|x\|^2 = R(x \mid \chi, \eta) + i\Theta(x \mid \chi, \eta)$ where

$$R(x \mid \chi, \eta) = x^\top \Re(\chi) + \Re(\eta)\|x\|^2 \quad \text{and} \quad \Theta(x \mid \chi, \eta) = x^\top \operatorname{Im}(\chi) + \operatorname{Im}(\eta)\|x\|^2. \tag{250}$$

After exponentiation, the real part can be absorbed into the distribution $p(x)$ yielding a tilted distribution

$$\widetilde{p}(x \mid \chi, \eta) \triangleq \frac{1}{Z(\chi,\eta)} e^{R(x|\chi,\eta)} p(x), \quad Z(\chi,\eta) \triangleq \mathbb{E}_p[e^{R(X|\chi,\eta)}] > 0. \tag{251}$$

Note that the curvature of this tilted distribution is bounded below by

$$-\nabla_x^2 \log \widetilde{p}(x \mid \chi, \eta) \succeq (1 - 2\Re(\eta)) I_d \succeq (1 - 2|\eta|) I_d \succeq \frac{7}{8} I_d. \tag{252}$$

Thus, the tilted distribution remains strongly log-concave.

The expressions in 249 can be written as expectations under the tilted distribution $\widetilde{X} \sim \widetilde{p}(\cdot \mid \chi, \eta)$ as

$$\mathbb{E}_p\left[e^{X^\top \chi + \eta \|X\|^2}\right] = Z(\chi,\eta) \mathbb{E}_{\widetilde{p}(\cdot|\chi,\eta)}\left[e^{i\Theta(\widetilde{X}|\chi,\eta)}\right], \tag{253}$$

$$\frac{\mathbb{E}_p\left[X e^{X^\top \chi + \eta \|X\|^2}\right]}{\mathbb{E}_p\left[e^{X^\top \chi + \eta \|X\|^2}\right]} = \frac{\mathbb{E}_{\widetilde{p}(\cdot|\chi,\eta)}\left[\widetilde{X} e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]}{\mathbb{E}_{\widetilde{p}(\cdot|\chi,\eta)}\left[e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]}. \tag{254}$$

Since $Z(\chi,\eta) > 0$, we only need to show that the magnitude of the purely complex-phase integral $\mathbb{E}_{\widetilde{p}(\cdot|\chi,\eta)}\left[e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]$ is bounded below by a non-zero constant and control the magnitude of the numerator $\mathbb{E}_{\widetilde{p}(\cdot|\chi,\eta)}\left[\widetilde{X} e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]$ from above.

To show that $\mathbb{E}_{\widetilde{p}}[e^{i\Theta(\widetilde{X}|\chi,\eta)}]$ is non-vanishing, notice that its magnitude remains the same under constant phase shifts, thus

$$\left|\mathbb{E}_{\widetilde{p}}\left[e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]\right| = \left|\mathbb{E}_{\widetilde{p}}\left[e^{i(\Theta(\widetilde{X}|\chi,\eta) - \mathbb{E}_{\widetilde{p}}[\Theta(\widetilde{X}|\chi,\eta)])}\right]\right|, \tag{255}$$

$$\geq \mathbb{E}_{\widetilde{p}}\left[\cos(\Theta(\widetilde{X} \mid \chi,\eta) - \mathbb{E}_{\widetilde{p}}[\Theta(\widetilde{X} \mid \chi,\eta)])\right], \tag{256}$$

$$\geq 1 - \frac{1}{2}\operatorname{Var}_{\widetilde{p}}(\Theta(\widetilde{X} \mid \chi,\eta)), \tag{257}$$

where the last inequality follows from $\cos(\theta) \geq 1 - \theta^2/2$ for $\theta \in \mathbb{R}$.

On the other hand, the magnitude of the numerator can be bounded above by the second moment $\widetilde{\mathsf{M}}_2(\chi,\eta) \triangleq \mathbb{E}_{\widetilde{p}}[\|\widetilde{X}^2\|]^{1/2}$ of the tilted distribution $\widetilde{p}$ as

$$\left\|\mathbb{E}_{\widetilde{p}}\left[\widetilde{X} e^{i\Theta(\widetilde{X}|\chi,\eta)}\right]\right\| \leq \mathbb{E}_{\widetilde{p}}[\|\widetilde{X}\|] \leq \mathbb{E}_{\widetilde{p}}[\|\widetilde{X}^2\|]^{1/2} = \widetilde{\mathsf{M}}_2(\chi,\eta). \tag{258}$$

Hence, $\operatorname{Var}_{\widetilde{p}}(\Theta(\widetilde{X} \mid \chi,\eta)) < 2$ ensures that the first integral in 249 is non-vanishing and the bound on the ratio-of-expectations in 249 is

$$\left\|\frac{\mathbb{E}_p\left[X e^{X^\top \chi + \eta \|X\|^2}\right]}{\mathbb{E}_p\left[e^{X^\top \chi + \eta \|X\|^2}\right]}\right\| \leq \frac{\mathbb{E}_{\widetilde{p}}[\|\widetilde{X}^2\|]^{1/2}}{1 - \frac{1}{2}\operatorname{Var}_{\widetilde{p}}(\Theta(\widetilde{X} \mid \chi,\eta))}. \tag{259}$$

Since $\widetilde{p}$ remains $7/8$-strongly log-concave, Brascamp-Lieb variance inequality in Lemma K.11 implies

$$\operatorname{Var}_{\widetilde{p}}(\Theta(\widetilde{X} \mid \chi,\eta)) \leq \frac{8}{7}\mathbb{E}_{\widetilde{p}}[\|\nabla\Theta(\widetilde{X} \mid \chi,\eta)\|^2], \tag{260}$$

$$= \frac{8}{7}\mathbb{E}_{\widetilde{p}}[\|\operatorname{Im}(\chi) + 2\operatorname{Im}(\eta)\widetilde{X}\|^2], \tag{261}$$

$$\leq \frac{8}{7}\left(\|\operatorname{Im}(\chi)\| + 2|\operatorname{Im}(\eta)|\mathbb{E}_{\widetilde{p}}[\|\widetilde{X}^2\|]^{1/2}\right)^2 \leq \frac{2}{7}\left(1 + \frac{\widetilde{\mathsf{M}}_2(\chi,\eta)}{4\sqrt{d}}\right)^2. \tag{262}$$

Thus, both the numerator and the denominator are controlled by the second moment:

$$\left\| \frac{\mathbb{E}_p\left[X e^{X^\top \chi + \eta \|X\|^2}\right]}{\mathbb{E}_p\left[e^{X^\top \chi + \eta \|X\|^2}\right]} \right\| \leq \frac{\widetilde{\mathsf{M}}_2(\chi, \eta)}{1 - \frac{1}{7}\left(1 + \frac{\widetilde{\mathsf{M}}_2(\chi,\eta)}{4\sqrt{d}}\right)^2}. \tag{263}$$

To get an upper bound on the second moment $\widetilde{\mathsf{M}}_2(\chi, \eta)$, note that the tilted distribution $\widetilde{p}(\widetilde{x} \mid \chi, \eta)$ is $\frac{7}{8}$-strongly log-concave and $\|\nabla_x \log \widetilde{p}(0 \mid \chi, \eta)\| = \|\Re(\chi))\| < 1/2$. Thus, Lemma K.14 implies

$$\widetilde{\mathsf{M}}_2(\chi, \eta) \leq \frac{2}{7}(1 + \sqrt{1 + 14d}) \leq \sqrt{2d}. \tag{264}$$

Substituting this bound on $\widetilde{\mathsf{M}}_2$ in 262 and 263 yields $\mathrm{Var}_{\widetilde{p}}(\Theta(\widetilde{X} \mid \chi, \eta)) < \frac{4}{7} < 2$, ensuring the first expectation in 249 is non-vanishing and the upper bound on the ratio of expectations holds. □

Lemma K.5 (Local Bounds on Jacobi Elliptic Functions)**.** *Let $m, \widetilde{m} \in (0, 1)$ and define $r \triangleq \min\left(1, \sqrt{\frac{1-\widetilde{m}}{1-m}}\right)$. For any real $u_\circ \in \mathbb{R}$ and $u \in \mathbb{C}$ such that $|u - u_\circ| < r$, the following holds:*

$$\max\left(|\operatorname{sn}(u \mid m)|, \frac{|\operatorname{cn}(u \mid m)|}{\sqrt{1 - \widetilde{m}\operatorname{sn}^2(u_\circ \mid m)}}, \frac{r|\operatorname{dn}(u \mid m)|}{\sqrt{1 - \widetilde{m}\operatorname{sn}^2(u_\circ \mid m)}}\right) \leq \frac{1}{1 - r^{-1}|u - u_\circ|}. \tag{265}$$

Proof. For notational simplicity, let $s_\circ \triangleq \sqrt{\widetilde{m}}\operatorname{sn}(u_\circ \mid m)$ and define the following functions under the coordinate transformation $v \triangleq \frac{u - u_\circ}{r}$:

$$f_1(v) \triangleq \operatorname{sn}(u_\circ + rv \mid m), \quad f_2(v) \triangleq \frac{\operatorname{cn}(u_\circ + rv \mid m)}{\sqrt{1 - s_\circ^2}}, \quad f_3(v) \triangleq \frac{r \operatorname{dn}(u_\circ + rv \mid m)}{\sqrt{1 - s_\circ^2}}, \tag{266}$$

$$\text{and} \quad h(\rho) \triangleq \max_{i=1,2,3} \max_{|v| \leq \rho} |f_i(v)| \quad \text{for} \quad \rho \geq 0 \tag{267}$$

We aim to show that $h(\rho) \leq \frac{1}{1-\rho}$ whenever $\rho < 1$.

First, all functions in the triplet (f_1, f_2, f_3) are holomorphic on the open unit ball $\mathbb{D} \triangleq \{v \in \mathbb{C} \mid |v| < 1\}$, since $|u - u_\circ| < r \leq 1$ implies $|\operatorname{Im}(u)| < 1 < \pi/2 \leq K(1 - m)$ where all three Jacobi elliptic functions $\operatorname{sn}(u \mid m)$, $\operatorname{cn}(u \mid m)$, and $\operatorname{dn}(u \mid m)$ are holomorphic.

Second, we show that $h(0) \leq 1$. Clearly $|f_1(0)| = |\operatorname{sn}(u_\circ \mid m)| \leq 1$ for any real $u_\circ \in \mathbb{R}$. Similarly, $|f_2(0)| \leq 1$ since $\operatorname{cn}^2(u_\circ \mid m) \leq \operatorname{cn}^2(u_\circ \mid m) + (1 - \widetilde{m})\operatorname{sn}^2(Kt_\circ \mid m) = 1 - s_\circ^2$. To show $|f_3(0)| \leq 1$, one can use the standard Jacobi elliptic function identities:

$$\operatorname{dn}^2(u_\circ \mid m) = 1 - m\operatorname{sn}^2(u_\circ \mid m), \tag{268}$$

$$= \operatorname{cn}^2(u_\circ \mid m) + (1 - m)\operatorname{sn}^2(u_\circ \mid m), \tag{269}$$

$$\leq \max\left(1, \frac{1-m}{1-\widetilde{m}}\right)\left[\operatorname{cn}^2(u_\circ \mid m) + (1 - \widetilde{m})\operatorname{sn}^2(Kt_\circ \mid m)\right], \tag{270}$$

$$= \frac{1 - s_\circ^2}{r^2}. \tag{271}$$

Third, we aim to derive a bound on each $|f_i(v)|$ in terms of $|v|$ for $v \in \mathbb{D}$. Standard derivative identities of Jacobi elliptic functions imply that the triplet (f_1, f_2, f_3) satisfies the following system of differential equations:

$$\frac{d}{dv} f_1(v) = c_1 f_2(v) f_3(v), \tag{272}$$

$$\frac{d}{dv} f_2(v) = c_2 f_1(v) f_3(v), \tag{273}$$

$$\frac{d}{dv} f_3(v) = c_3 f_1(v) f_2(v), \tag{274}$$

where the coefficients are $c_1 \triangleq 1 - s_\circ^2$, $c_2 \triangleq -1$, and $c_3 \triangleq -r^2 m$, with all three of them satisfying $|c_i| \leq 1$. Integrating these ODEs along the straight line from the origin to any $v \in \mathbb{D}$ and using the bounds $|f_i(0)| \leq 1$ and $|c_i| \leq 1$ give

$$|f_i(v)| = \left| f_i(0) + c_i \int_0^1 f_j(\theta v)\, f_k(\theta v)\, v d\theta \right|, \tag{275}$$

$$\leq |f_i(0)| + |c_i| \int_0^1 |f_j(\theta v)|\, |f_k(\theta v)||v| d\theta, \tag{276}$$

$$\leq 1 + \int_0^1 h^2(\theta|v|)|v| d\theta. \tag{277}$$

Taking the maximum of of both sides over $i = 1, 2, 3$ and $|v| \leq \rho < 1$ gives the following integral inequality for the monotonic function $[0, 1) \ni \rho \mapsto h(\rho)$:

$$h(\rho) \leq H(\rho) \triangleq 1 + \int_0^\rho h^2(\rho') d\rho'. \tag{278}$$

Taking the derivative of $H(\rho)$ gives

$$\frac{d}{d\rho} H(\rho) = h^2(\rho) \leq H^2(\rho) \implies \frac{d}{d\rho}\left(-\frac{1}{H(\rho)}\right) \leq 1 \tag{279}$$

Noting $H(0) = 1$ and applying the fundamental theorem of calculus, we get the following bound

$$-\frac{1}{H(\rho)} = -1 + \int_0^\rho \frac{d}{d\rho'}\left(-\frac{1}{H(\rho')}\right) d\rho' \leq -1 + \rho. \tag{280}$$

Thus, $h(\rho) \leq H(\rho) \leq \frac{1}{1-\rho}$. □

Lemma K.6 (Wasserstein-2 distance and KL divergence between smoothed and unsmoothed log–concave distributions)**.** *Let $X \sim p$ be zero-mean where p is log-concave whose curvature is bounded in $[1, \kappa]$. Define the Gaussian smoothed distribution $\widetilde{X} \overset{(d)}{=} \widetilde{\sigma} Z + \sqrt{1 - \widetilde{\sigma}^2} X$ and denote by $\widetilde{p}_{\widetilde{\sigma}} = \mathcal{L}(\widetilde{X})$ it probability density, which remains log-concave with condition number bounded by $\widetilde{\kappa} \triangleq \frac{\kappa}{1 + \widetilde{\sigma}^2(\kappa - 1)}$. Then,*

$$W_2(\widetilde{p}_{\widetilde{\sigma}}, p) \leq \sqrt{\frac{d}{\kappa}}(\sqrt{\kappa/\widetilde{\kappa}} - 1) \tag{281}$$

$$D_{\mathrm{KL}}(\widetilde{p}_{\widetilde{\sigma}} \| p) \leq \frac{d}{2}\left[\kappa/\widetilde{\kappa} - 1 - \log\left(\kappa/\widetilde{\kappa}\right)\right] \leq \frac{d}{4}\left(\kappa/\widetilde{\kappa} - 1\right)^2. \tag{282}$$

Moreover, both first bounds are attained by $p = N(0, \kappa^{-1} I_d)$.

K.2 Oracle Complexity

Lemma K.7 (Lipschitz Constant of Drift under Elliptic Schedule)**.** *Fix $\widetilde{\sigma} \in (0, 1)$, set $\widehat{m} = 1 - \widetilde{\sigma}^2$ and consider the elliptic schedule $\alpha_t = \mathrm{cn}(K(\widehat{m})t \mid \widehat{m})$, $\beta = \mathrm{sn}(K(\widehat{m})t \mid \widehat{m})$ interpolating between $N(0, I_d)$ and $\widetilde{X} \overset{(d)}{=} \widetilde{\sigma} Z + \sqrt{1 - \widetilde{\sigma}^2} X$. Its PF-ODE drift $\widetilde{b}_t(x)$ satisfies:*

$$-K(\widehat{m}) I_d \preceq \nabla \widetilde{b}_t(x) \preceq 0. \tag{283}$$

Thus, $\mathrm{Lip}(\widetilde{b}_t(\bullet)) \leq K(\widehat{m}) = K(1 - \widetilde{\sigma}^2)$ for all $t \in [0, 1]$ and the resulting flow map $z \mapsto \Phi_t(z)$ is non-expansive, i.e.,

$$\|\Phi_t(z_1) - \Phi_t(z_2)\| \leq \|\Phi_s(z_1) - \Phi_s(z_2)\| \leq \|z_1 - z_2\|, \tag{284}$$

for all $0 \leq s \leq t \leq 1$ and $z_1, z_2 \in \mathbb{R}^d$.

Theorem K.8 (ODEs with Analytic and Bounded Solutions)**.** *Consider the following ODE*

$$\frac{dx_t}{dt} = b(x_t, t), \quad x_0 = z \in \mathcal{Z}, \quad t \in [0, 1], \tag{285}$$

where the initialization belongs to a set $\mathcal{Z} \subset \mathbb{R}^d$. Suppose the following assumptions hold:

i. There exists $\rho > 0$ such that the solution map $t \mapsto \Phi_t(z)$ extends holomorphically to $\Omega_\rho = \{t \in \mathbb{C} \mid \mathrm{dist}(t, [0,1]) < \rho\}$ for all initialization $z \in \mathcal{Z}$.

ii. There exists $M > 0$ such that the solution map $t \mapsto \Phi_t(z)$ is bounded on Ω_ρ by M for all initialization $z \in \mathcal{Z}$, i.e., $\sup_{\substack{t\in\Omega_\rho,\\ z\in\mathcal{Z}}} \|\Phi_t(z)\| \le M$.

iii. There exists $L > 0$ such that $\|b(x,t) - b(y,t)\| \le L\|x - y\|$ for all $x, y \in \mathbb{R}^d$ and $t \in [0,1]$.

Suppose $\max(L, \rho^{-1}) \ge 1$. *Then, for any $\zeta > 0$, there exists an algorithm generating a terminal estimate $\widehat{x}_1(z)$ with an error satisfying*

$$\sup_{z\in\mathcal{Z}} \|\widehat{x}_1(z) - \Phi_1(z)\| \le \zeta,$$

using at most

$$N(\rho, M, L, \zeta) \le C \max(L, \rho^{-1}) \left(1 + L + \log\left(\frac{M \max(L, \rho^{-1})}{\zeta}\right)\right)^2, \tag{286}$$

function evaluations $(x,t) \mapsto b(x,t)$.

Proof. Proof directly follows from (Lee et al., 2018, Thm2.3) together with Bernstein's approximation theorem using Chebyshev polynomials Trefethen (2019). □

K.3 Technical Lemmas

Lemma K.9 (Cauchy's Estimate)**.** *Let $f : B_\gamma(\tau_0) \mapsto \mathbb{C}^d$ be holomorphic and bounded on the ball $B_\gamma(\tau_0) = \{\tau \in \mathbb{C} \mid |\tau - \tau_0| < \gamma\}$. Then,*

$$\|f^{(n)}(\tau_0)\| \le \frac{n!}{\gamma^n} \sup_{\tau\in B_\gamma(\tau_0)} \|f(\tau)\|. \tag{287}$$

Lemma K.10 (Existence of Holomorphic Solutions of ODEs)**.** *(Teschl, 2012, Thm 4.1)*

Given $(x_0, \tau_0) \in \mathbb{C}^d \times \mathbb{C}$, let $\Omega \triangleq \{(x,\tau) \in \mathbb{C}^d \times \mathbb{C} \mid \|x - x_0\| < \delta, |\tau - \tau_0| < \varepsilon\}$ be an open rectangle and suppose $F : \Omega \to \mathbb{C}$ is analytic and bounded with

$$M \triangleq \sup_{(x,\tau)\in\Omega} \|F(x,\tau)\| < \infty. \tag{288}$$

Then the initial value problem

$$\frac{dx_\tau}{d\tau} = F(x_\tau, \tau), \quad x_{\tau_0} = x_0, \tag{289}$$

has a unique analytic solution defined in the disc $\{\tau \in \mathbb{C} \mid |\tau - \tau_0| < \varepsilon_0\}$, where

$$\varepsilon_0 \triangleq \min\left(\varepsilon, \frac{\delta}{M}\right). \tag{290}$$

Lemma K.11 (Brascamp-Lieb variance inequality)**.** *For $p(x) \propto \exp(-V(x))$ with $V \in \mathcal{C}^2$ and $\nabla^2 V(x) \succ 0$ and any $f \in \mathcal{C}^1$, we have that*

$$\mathrm{Var}_p(f) \le \mathbb{E}_{X\sim p}\left[\nabla f(X)^\top (\nabla^2 V(X))^{-1} \nabla f(X)\right]. \tag{291}$$

Lemma K.12 (Posterior Identities)**.** *For $X \sim p(x) \propto \exp(-V(x))$ with $V \in \mathcal{C}^2$ and $\nabla^2 V(x) \succ 0$ and independent standard normal $Z \sim N(0, I_d)$, we have that*

$$\frac{(\mathbb{E}[\nabla^2 V(X)])^{-1}}{\tau} \preceq \nabla_y \mathbb{E}\left[X \mid X + \sqrt{\tau} Z = y\right] \preceq \frac{\mathbb{E}[(\nabla^2 V(X))^{-1}]}{\tau}. \tag{292}$$

In particular, if $\lambda_{\min} I_d \preceq \nabla^2 V(x) \preceq \lambda_{\max} I_d$ for constants $0 < \lambda_{\min} \le \lambda_{\max}$ one gets

$$\frac{1}{\tau\lambda_{\max} + 1} I_d \preceq \nabla_y \mathbb{E}\left[X \mid X + \sqrt{\tau} Z = y\right] \preceq \frac{1}{\tau\lambda_{\min} + 1} I_d. \tag{293}$$

Lemma K.13 (Stein's Identity). *Let $X \sim p(x)$ be a random variable with $\mathcal{C}^1(\mathcal{X})$ probability density over a domain $\mathcal{X} \subseteq \mathbb{R}^d$. For a smooth function $f \in \mathcal{C}^1(\mathcal{X})$, it holds that*

$$\mathbb{E}_p[\nabla f(X)] = -\mathbb{E}_p[f(X)\,\nabla \log p(X)] \quad \textit{whenever} \quad \int_{\mathcal{X}} \nabla \cdot (f(x)p(x))dx = 0. \tag{294}$$

In particular, for $\mathcal{X} = \mathbb{R}^d$, this holds as long as $f(x)p(x) \longrightarrow 0$ as $\|x\| \longrightarrow \infty$.

Lemma K.14 (Second-moment bound under strong log-concavity). *Let $X \sim p(x) \propto \exp(-V(x))$ with $V \in \mathcal{C}^2$ be such that $\nabla^2 V \succeq \lambda_{\min} I_d$ and $\|\nabla V(0)\| \leq M$ for $\lambda_{\min} > 0$ and $M \geq 0$. Then,*

$$\mathbb{E}_p[\|X\|^2]^{1/2} \leq \frac{M + \sqrt{M^2 + 4d\lambda_{\min}}}{2\lambda_{\min}} \leq \frac{M + \sqrt{M^2 + 4\lambda_{\min}}}{2\lambda_{\min}}\sqrt{d}. \tag{295}$$

Proof. Using the fundamental theorem of calculus, one gets

$$\nabla V(x) = \nabla V(0) + \int_0^1 \nabla^2 V(\theta x) x d\theta. \tag{296}$$

Taking the inner product of both sides with x yields

$$\begin{aligned} x^\top \nabla V(x) &= x^\top \nabla V(0) + \int_0^1 x^\top \nabla^2 V(\theta x) x d\theta, && (297)\\ &\geq -\|x\|\|\nabla V(0)\| + \lambda_{\min}\|x\|^2, && (298)\\ &\geq \lambda_{\min}\|x\|^2 - M\|x\|. && (299) \end{aligned}$$

where the second inequality follows from Cauchy-Schwarz and the strong log-concavity. Note that, for any $\rho \in (0, 1)$, it holds that

$$x^\top \nabla V(x) \geq \lambda_{\min}\|x\|^2 - M\|x\| \geq \rho\lambda_{\min}\|x\|^2 - \frac{M^2}{4\lambda_{\min}(1-\rho)}. \tag{300}$$

Taking the expectation of both sides under $p(x)$ and noting $\mathbb{E}_p[X^\top V(X)] = d$ by Stein's identity in Lemma K.13 since $\lim_{\|x\|\to\infty} xp(x) = 0$ due to strong log-concavity, we get

$$\mathbb{E}_p[\|X\|^2] \leq \frac{1}{\rho}\left(\frac{d}{\lambda_{\min}} + \frac{M^2}{4\lambda_{\min}^2(1-\rho)}\right), \forall \rho \in (0, 1). \tag{301}$$

Taking the infimum over $\rho \in (0, 1)$ yields

$$\mathbb{E}_p[\|X\|^2] \leq \left(\frac{M + \sqrt{M + 4d\lambda_{\min}}}{2\lambda_{\min}}\right)^2. \tag{302}$$

□

Lemma K.15 (Gaussian initialization tails). *For any $B > 0$, let $\Pi_B : z \mapsto z\mathbb{1}_{\|z\|\leq B} + B\frac{z}{\|z\|}\mathbb{1}_{\|z\|>B}$ denote the Euclidean projection onto the ball $\{x \in \mathbb{R}^d \mid \|z\| \leq B\}$. Suppose $T : \mathbb{R}^d, \to \mathbb{R}^d$ is globally L-Lipschitz and $\widehat{T} : \mathbb{R}^d \to \mathbb{R}^d$ is measurable such that*

$$\sup_{\|z\|\leq B_\delta} \|\widehat{T}(z) - T(z)\| \leq \zeta, \quad \textit{where} \quad B_\delta \triangleq \sqrt{d} + \sqrt{4\log\left(\frac{\sqrt{2}}{\delta}\right)},$$

for some $\zeta > 0$ and $\delta \in (0, \sqrt{2})$. Then,

$$\mathbb{E}_{Z\sim N(0,I_d)}\left[\|\widehat{T}(\Pi_{B_\delta} Z) - T(Z)\|^2\right]^{1/2} \leq \zeta + L\delta. \tag{303}$$

L Extended Related Works

Interpolant and noise schedules. Stochastic interpolants (Albergo et al., 2025) separate the path, set by the schedule (α_t, β_t), from its numerical integration. Schedules are often chosen by hand, as with the trig schedule (Nichol & Dhariwal, 2021) and the noise levels of EDM (Karras et al., 2022); the continuous-time variational bound depends on the schedule only through its endpoint signal-to-noise ratios (Kingma et al., 2023), and Dieleman (2024) argues that the schedule entangles several design choices. Principled designs optimize the time grid for a given solver (Sabour et al., 2024), minimize an averaged Lipschitz constant of the drift (Chen et al., 2025), or adapt to high-dimensional mixtures (Aranguri et al., 2025), and Gao et al. (2024) study the regularity of the resulting flows. We instead choose the schedule by the analyticity of the flow in complex time. The elliptic family reduces to the trig schedule at $\kappa = 1$, contains the closed-form Gaussian schedule of Align Your Steps after rescaling (Section I.4), and is asymptotically optimal for the worst-case drift-Lipschitz criterion whose Gaussian minimizer is the schedule of Chen et al. (2025) (Section I.5).

Gaussian sampling and rational approximation. Sampling from $\mathcal{N}(0, \Lambda^{-1})$ amounts to computing $\Lambda^{-1/2}z$, for which rational approximations, and Zolotarev's optimal ones in particular, converge far faster than polynomial ones when the spectrum is wide (Higham, 2008; Hale et al., 2008; Nakatsukasa & Freund, 2016; Gawlik, 2019; Beckermann & Townsend, 2017; Trefethen & Wilber, 2025). The closest work to ours is Liu (2026), which samples Gaussians from *smoothed-score* queries, our interpolant scores in variance-exploding coordinates (Section I.7). Combining these queries through a sinc-quadrature rational approximation of $x^{-1/2}$, Liu obtains $O\big((\log\kappa + \log(\sqrt{d}/\varepsilon))\log(\sqrt{d}/\varepsilon)\big)$ queries, improving the $\sqrt{\kappa}$ dependence of unsmoothed scores to $\log\kappa$; the accompanying lower bounds count bits received from a finite-precision oracle rather than exact queries. We improve on this in two ways. Zoloflow realizes the optimal (Zolotarev) approximant, which removes the additive $\log^2(\sqrt{d}/\varepsilon)$ term Theorem 3.7, and we prove a matching lower bound on the number of exact queries for oblivious score-span algorithms, a class that contains Liu's sampler and the standard diffusion solvers (Theorem 3.7). Moreover, unlike Liu's sampler, Zoloflow is a schedule plus a discretization of the probability-flow ODE, so the schedule carries over to non-Gaussian targets.

Fast samplers and discretizations. Many works accelerate sampling for a fixed schedule by improving the discretization: DDIM (Song et al., 2022), the exponential integrators DEIS (Zhang & Chen, 2023) and DPM-Solver(++) (Lu et al., 2022a;b), and the Heun sampler of EDM (Karras et al., 2022); Align Your Steps (Sabour et al., 2024) instead tunes the time grid for a fixed solver. We design the schedule and the discretization together. On Gaussian targets all of these solvers are oblivious score-span algorithms (Lemma 3.6 and Section J), so our lower bound covers them, and on non-Gaussian targets the elliptic schedule also improves standard solvers such as DPM-Solver++ (Section A.2).

Complexity of diffusion and log-concave sampling. Diffusion samplers converge in polynomially many steps given accurate scores, for both the reverse SDE (Chen et al., 2023b) and the probability-flow ODE (Chen et al., 2023a), while classical samplers for strongly log-concave targets pay polynomially in the condition number (Chewi, 2026). When an ODE solution is analytic, collocation reaches accuracy ε in $\mathrm{polylog}(1/\varepsilon)$ steps (Lee et al., 2018), which Gatmiry et al. (2026) use to give high-accuracy guarantees for diffusion samplers. The step count depends on how far the solution extends into complex time; we choose the schedule to maximize this region, which gives a query complexity polylogarithmic in both κ and $1/\varepsilon$ for log-concave targets (Section 4).

Schedules as approximation problems. Choosing a schedule so that a fixed budget of steps is as accurate as possible has close analogues. For gradient descent on a quadratic, stepsizes define a polynomial in the Hessian, and Chebyshev stepsizes (Agarwal et al., 2021) attain the minimax $O(\sqrt{\kappa}\log(1/\varepsilon))$ rate; the silver stepsize schedule (Altschuler & Parrilo, 2025; 2026) accelerates gradient descent beyond quadratics through the stepsizes alone. Score queries instead produce rational functions of the precision (Lemma 3.6), so Zolotarev's approximant replaces the Chebyshev polynomial and $\sqrt{\kappa}$ becomes $\log\kappa$. Similarly, double-exponential and sinc quadrature (Takahasi & Mori, 1973; Mori, 2005; Stenger, 2000), like the sampler of Liu (2026), change variables so that

the integrand is analytic in a wide strip and then use a uniform grid; the elliptic schedule is the time change that maximizes this strip, and Zoloflow queries on a uniform grid in the new time.