Elliptic Sampling: Optimal Diffusion Schedules for Gaussians and Beyond





Pablo A. Parrilo
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Joint work with **Cole Becker**, **Kimia Hassibi**
Taylan Kargin, and **Asu Ozdaglar**

arXiv:2611.18542

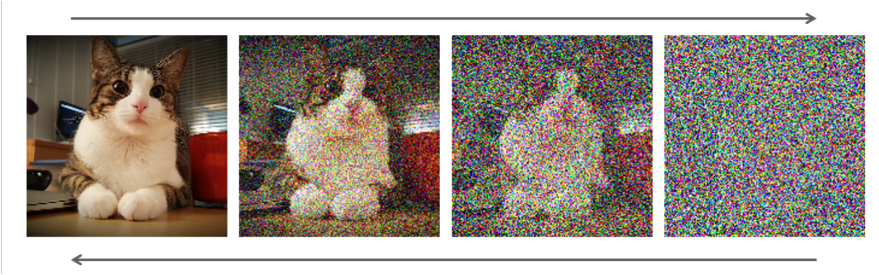
Princeton ORFE, November 2026





Context: Generative Modeling via Diffusion and Flows

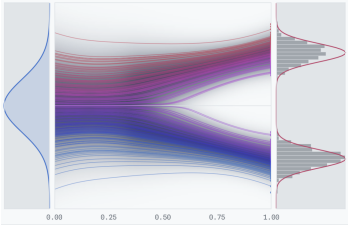
- **Goal:** produce new samples $X \sim \pi$ from a complicated distribution π on $\mathbb{R}^d$ (images, audio, molecules, ...), learned from data.
- Modern approach (diffusion models, flow matching): learn an ODE/SDE that *transports* a simple Gaussian $\mathcal{N}(0, I_d)$ to π.
- The vector field is a large neural network (the score or velocity), learned once.
- Sampling = numerically integrating this ODE. Every step is a network evaluation.



Stochastic Interpolants and Probability Flow ODEs

Consider the stochastic process (ABV25)

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



where $z \sim \mathcal{N}(0, I_d)$ and $x_1 \sim \pi$.

- (α_t, β_t) is the interpolant schedule:
  Boundary conditions: $(\alpha_0, \beta_0) = (1, 0)$, $(\alpha_1, \beta_1) = (0, 1)$.
- Key idea: reverse a diffusion (Anderson 1979, Haussmann-Pardoux 1986)
- The *probability-flow ODE*

$$\dot{X}_t = b_t(X_t), \qquad X_0 \sim \mathcal{N}(0, I_d), \qquad b_t(x) := \mathbb{E}[\dot{x}_t \mid x_t = x]$$

  has the same marginals as x_t. In particular, $X_1 \sim \pi$.
- Via Tweedie's formula, the drift is a function of the score $s_t(x) = \nabla \log p_t(x)$:

$$b_t(x) = (\dot{\beta}_t/\beta_t)\, x + \alpha_t^2 \left(\dot{\beta}_t/\beta_t - \dot{\alpha}_t/\alpha_t\right) s_t(x).$$

Two Key Design Decisions

To fully specify a diffusion sampling algorithm, one must choose:

- An interpolant schedule (α_t, β_t): linear, trigonometric, EDM, Lipschitz-guided, ...
- A numerical method for integrating the ODE: Euler, Runge–Kutta, DDIM, DPM-Solver++, ...

Are these choices independent? Should they be chosen jointly? If so, how?

Furthermore: how do the properties of the target distribution affect these choices?

The Central Question

Q: Given a fixed budget of r score evaluations, how should one *jointly* choose the interpolant schedule and the numerical method for the ODE?

- In practice, schedules and solvers are designed *separately*, mostly by heuristics and extensive tuning.
- Schedules can significantly change how *hard* the ODE is to integrate.
- An algorithm design problem, with a minimax (worst-case) flavor.
- Natural parameter: the condition number κ of the target.

Our Results

A new construction: the Elliptic Schedule.
Based on Jacobi elliptic functions, and parametrized by condition number κ.

Can be motivated either from:

- Analyticity of (complexified) ODE trajectories
- Optimal rational approximation of the inverse square root (Zolotarev)

Notably, the *same* construction simultaneously provides:

- Optimal answers for Gaussians (minimax up to constants, within oblivious score-span)
- Great bounds for log-concave targets (polylog in κ and $1/\varepsilon$)
- An empirically strong schedule for general targets

Problem Formulation

$$\inf_{\text{alg} \in \mathcal{A}_r} \sup_{\pi \in \mathsf{P}} \operatorname{dist}\left(\hat{\pi}_{\text{alg}}(\pi), \pi\right)$$

Problem Formulation

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

Oblivious score-span algorithms $\mathcal{A}_r$

A schedule (α, β), together with query times $t_0, \dots, t_{r-1}$ and coefficients a_{ij}, b_{ij} fixed in advance:

$$x_0 = z, \qquad x_j = \sum_{i<j} a_{ij}\, x_i + \sum_{i<j} b_{ij}\, s_{t_i}(x_i), \qquad j = 1, \dots, r.$$

- Access to π only through an exact *score oracle* $(t, x) \mapsto s_t(x)$.
- Contains standard solvers: Euler, Heun, DDIM, DPM-Solver++, ...
- Oracles for different schedules are interchangeable (time + space rescaling), so the schedule is a natural design choice.

Target Classes

Two nested classes, both parametrized by a condition number $\kappa \geq 1$:

$$\underbrace{\mathsf{G}_\kappa = \left\{\mathcal{N}(0, \Lambda^{-1}) : \ I_d \preceq \Lambda \preceq \kappa I_d\right\}}_{\text{Gaussians (total variation)}} \ \subset \ \underbrace{\mathsf{P}_\kappa = \left\{\pi \propto e^{-V} : \ I_d \preceq \nabla^2 V \preceq \kappa I_d\right\}}_{\text{Strongly log-concave (Wasserstein-2)}}$$

- (Zero mean is w.l.o.g.; the mean can be recovered from one score query.)

Target Classes

Two nested classes, both parametrized by a condition number $\kappa \geq 1$:

$$\underbrace{\mathsf{G}_\kappa = \left\{\mathcal{N}(0, \Lambda^{-1}) : \ I_d \preceq \Lambda \preceq \kappa I_d\right\}}_{\text{Gaussians (total variation)}} \ \subset \ \underbrace{\mathsf{P}_\kappa = \left\{\pi \propto e^{-V} : \ I_d \preceq \nabla^2 V \preceq \kappa I_d\right\}}_{\text{Strongly log-concave (Wasserstein-2)}}$$

- (Zero mean is w.l.o.g.; the mean can be recovered from one score query.)

Why care about Gaussians?

- Simplest nontrivial case, and a lower bound for any larger class.
- Even here, the question is surprisingly rich: it connects to classical approximation theory and elliptic functions.
- Empirically, the resulting schedule works well far beyond Gaussians.

Gaussian Targets: Explicit Trajectories

For $\pi = \mathcal{N}(0, \Lambda^{-1})$, the score is linear, and thus the ODE can be solved in closed form:

$$X_t = \left(\alpha_t^2 I_d + \beta_t^2 \Lambda^{-1}\right)^{1/2} X_0.$$

On each eigenmode of Λ (eigenvalue $\lambda \in [1, \kappa]$):

$$x_\lambda(t) = \sqrt{\alpha_t^2 + \lambda^{-1}\beta_t^2}\, x_\lambda(0)$$

Gaussian Targets: Explicit Trajectories

For $\pi = \mathcal{N}(0, \Lambda^{-1})$, the score is linear, and thus the ODE can be solved in closed form:

$$X_t = \left(\alpha_t^2 I_d + \beta_t^2 \Lambda^{-1}\right)^{1/2} X_0.$$

On each eigenmode of Λ (eigenvalue $\lambda \in [1, \kappa]$):

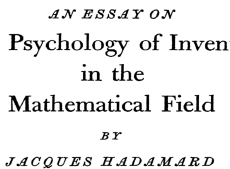
$$x_\lambda(t) = \sqrt{\alpha_t^2 + \lambda^{-1}\beta_t^2}\, x_\lambda(0)$$

- Smooth for real t. But in complex time, there are branch points wherever

$$\alpha_t^2 + \lambda^{-1}\beta_t^2 = 0.$$

- The singularity locations depend on the choice of schedule!
- Why do we care about complex singularities of a real ODE?

Real and Complex Truths



AN ESSAY ON

The Psychology of Invention
in the
Mathematical Field

BY

JACQUES HADAMARD

in the nineteenth century, established on solid and rigorous bases. It has been written that the shortest and best way between two truths of the real domain often passes through the imaginary one.

We have mentioned Cardan's case with Socrates' and

Often attributed to Hadamard, who was quoting Painlevé: "Il apparut que, entre deux vérités du domaine réel, le chemin le plus facile et le plus court passe bien souvent par le domaine complexe." (*Analyse des travaux scientifiques*, 1900)

Why Analyticity?

- Classical principle in numerical analysis: convergence rates are often governed by the domain of (complex) analyticity.
  - Chebyshev interpolation: error $\sim \rho^{-N}$, with ρ given by the largest Bernstein ellipse.
  - Trapezoidal rule/sinc methods: error $\sim e^{-c\,\omega N}$ for functions analytic on a strip $|\mathrm{Im}(t)| < \omega$.
  - Collocation for analytic ODEs (Lee–Song–Vempala): $\mathrm{polylog}(1/\varepsilon)$ steps, depending on the strip width.

Why Analyticity?

- Classical principle in numerical analysis: convergence rates are often governed by the domain of (complex) analyticity.
  - Chebyshev interpolation: error $\sim \rho^{-N}$, with ρ given by the largest Bernstein ellipse.
  - Trapezoidal rule/sinc methods: error $\sim e^{-c\,\omega N}$ for functions analytic on a strip $|\mathrm{Im}(t)| < \omega$.
  - Collocation for analytic ODEs (Lee–Song–Vempala): $\mathrm{polylog}(1/\varepsilon)$ steps, depending on the strip width.
- Large region of analyticity $\Rightarrow$ fewer steps. Natural design criterion:

Q: Find schedule α_t, β_t such that the trajectories $x_\lambda(t)$ are holomorphic on the widest possible strip $S_\gamma = \{t \in \mathbb{C} : |\mathrm{Im}(t)| < \gamma\}$, simultaneously for all $\lambda \in [1, \kappa]$.

Why Analyticity?

- Classical principle in numerical analysis: convergence rates are often governed by the domain of (complex) analyticity.
  - Chebyshev interpolation: error $\sim \rho^{-N}$, with ρ given by the largest Bernstein ellipse.
  - Trapezoidal rule/sinc methods: error $\sim e^{-c\,\omega N}$ for functions analytic on a strip $|\mathrm{Im}(t)| < \omega$.
  - Collocation for analytic ODEs (Lee–Song–Vempala): $\mathrm{polylog}(1/\varepsilon)$ steps, depending on the strip width.
- Large region of analyticity $\Rightarrow$ fewer steps. Natural design criterion:

> **Q:** Find schedule α_t, β_t such that the trajectories $x_\lambda(t)$ are holomorphic on the widest possible strip $S_\gamma = \{t \in \mathbb{C} : |\mathrm{Im}(t)| < \gamma\}$, simultaneously for all $\lambda \in [1, \kappa]$.

The answer involves beautiful and classical 19th century mathematics: elliptic functions!

Interlude: Jacobi Elliptic Functions

For a parameter $m \in [0, 1)$, the elliptic integral of the first kind

$$u = F(\varphi \mid m) = \int_0^{\varphi} \frac{d\theta}{\sqrt{1 - m \sin^2 \theta}}, \qquad K(m) := F(\tfrac{\pi}{2} \mid m).$$

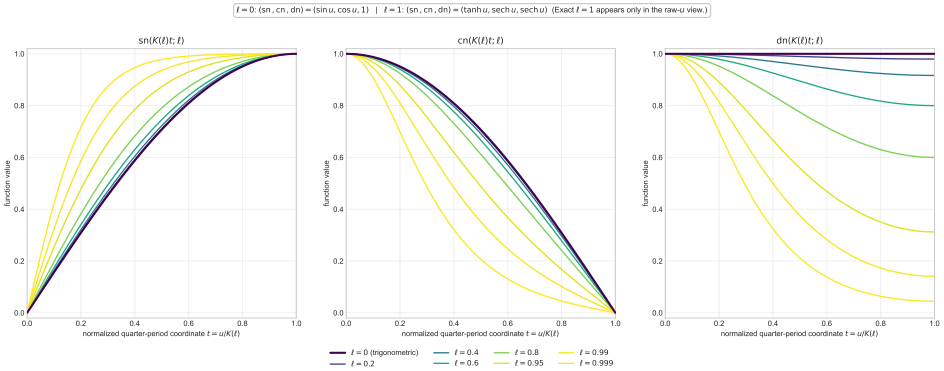
Its inverse is the Jacobi amplitude $\varphi = \operatorname{am}(u \mid m)$. Then define

$$\operatorname{sn}(u \mid m) = \sin \varphi, \qquad \operatorname{cn}(u \mid m) = \cos \varphi, \qquad \operatorname{dn}(u \mid m) = \sqrt{1 - m \sin^2 \varphi}$$

- "Sine and cosine with a nonlinear clock" $u \mapsto \varphi$, slowed near $\varphi = \pi/2$.
- $\operatorname{sn}^2 + \operatorname{cn}^2 = 1$, $\operatorname{dn}^2 + m \operatorname{sn}^2 = 1$.
- In SciPy: `scipy.special.ellipj`.

	$m = 0$	$m \to 1$
$\operatorname{am} u$	u	$\arcsin(\tanh u)$
$\operatorname{sn} u$	$\sin u$	$\tanh u$
$\operatorname{cn} u$	$\cos u$	$\operatorname{sech} u$
$\operatorname{dn} u$	1	$\operatorname{sech} u$
$K(m)$	$\pi/2$	∞

Jacobi Elliptic Functions: The Real Picture



- In the normalized clock $t = u/K(m)$, all functions run over a quarter period on $[0, 1]$.
- As $m \to 1$, the phase moves fast early and slows down near $t = 1$.

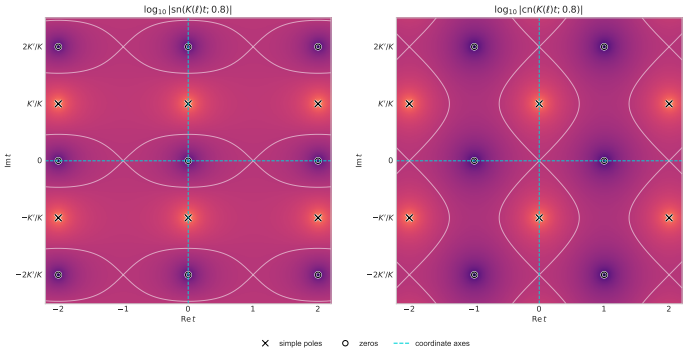
Jacobi Elliptic Functions: The Complex Picture

- Meromorphic and doubly periodic: periods $4K$ and $2iK'$ (for sn), where $K' := K(1-m)$.
- Natural domain is a torus.
- Common simple poles at

$$u = 2jK + (2n+1)iK'.$$

- Hence $\mathrm{sn}(Kt \mid m)$, $\mathrm{cn}(Kt \mid m)$ are holomorphic on the strip

$$|\mathrm{Im}(t)| < K'/K.$$



$|\,\mathrm{sn}\,|$ and $|\,\mathrm{cn}\,|$ in normalized time t, $m = 0.64$. Zeros ○, poles ×.

The Optimal Phase

Consider a VP schedule $(\alpha_t, \beta_t) = (\cos\varphi(t), \sin\varphi(t))$ with phase $\varphi : [0,1] \to [0, \pi/2]$.

Proposition (Jacobi amplitude is the optimal phase)

Let $m = 1 - \kappa^{-1}$, $K = K(m)$, $K' = K(1-m)$. The phase $\varphi_\star(t) = \mathrm{am}(Kt \mid m)$ maximizes the strip width γ, subject to

i. φ holomorphic on S_γ,

ii. $\varphi(0) = 0$, $\varphi(1) = \pi/2$,

iii. $\alpha_t^2 + \lambda^{-1}\beta_t^2 = 1 - (1 - \lambda^{-1})\sin^2\varphi(t) \neq 0$ for all $t \in S_\gamma$, $\lambda \in [1, \kappa]$.

The optimal strip width is $\gamma_\star = K'/K$.

The Optimal Phase

Consider a VP schedule $(\alpha_t, \beta_t) = (\cos\varphi(t), \sin\varphi(t))$ with phase $\varphi : [0,1] \to [0, \pi/2]$.

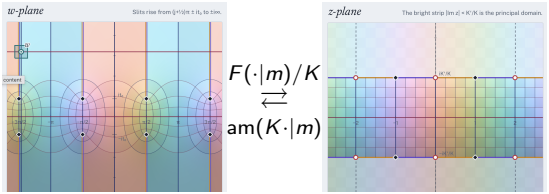
Proposition (Jacobi amplitude is the optimal phase)

Let $m = 1 - \kappa^{-1}$, $K = K(m)$, $K' = K(1-m)$. The phase $\varphi_\star(t) = \mathrm{am}(Kt \mid m)$ maximizes the strip width γ, subject to

i. φ holomorphic on S_γ,
ii. $\varphi(0) = 0$, $\varphi(1) = \pi/2$,
iii. $\alpha_t^2 + \lambda^{-1}\beta_t^2 = 1 - (1-\lambda^{-1})\sin^2\varphi(t) \neq 0$ for all $t \in S_\gamma$, $\lambda \in [1, \kappa]$.

The optimal strip width is $\gamma_\star = K'/K$.

- $\gamma_\star \approx \dfrac{\pi}{\log(16\kappa)}$: degrades only logarithmically in κ.
- $\kappa = 1$: $\varphi_\star(t) = \frac{\pi}{2}t$ gives the trig schedule.

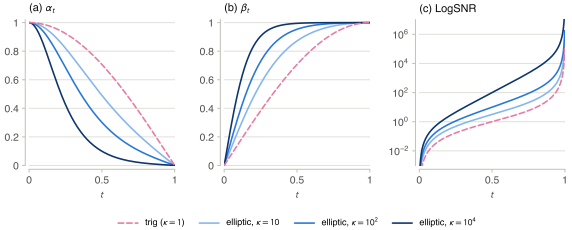


The Elliptic Schedule

Definition

For $\kappa \geq 1$, with $m = 1 - \kappa^{-1}$ and $K = K(m)$:

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



- VP, holomorphic on $S_{\gamma_\star}$, and explicitly adapted to κ. Reduces to trig for $\kappa = 1$.
- Signal-to-noise ratio $\beta_t^2/\alpha_t^2 = \mathrm{sc}^2(Kt \mid m)$.

Second Lens: Gaussian Sampling as Rational Function Approximation

Lemma

Let $\pi = \mathcal{N}(0, \Lambda^{-1})$. For every $\mathrm{alg} \in \mathcal{A}_r$, there is a rational function $q \in \mathcal{R}_{r,r}$ (numerator and denominator degrees $\leq r$) with

$$x_r^{\mathrm{alg}} = q(\Lambda)\, z.$$

- Gaussian scores are linear: $s_t(x) = -(\alpha_t^2 I + \beta_t^2 \Lambda^{-1})^{-1} x$, a *resolvent* of Λ.
- Each score query raises the numerator and denominator degrees by at most one.

Second Lens: Gaussian Sampling as Rational Function Approximation

Lemma

Let $\pi = \mathcal{N}(0, \Lambda^{-1})$. For every $\mathrm{alg} \in \mathcal{A}_r$, there is a rational function $q \in \mathcal{R}_{r,r}$ (numerator and denominator degrees $\leq r$) with

$$x_r^{\mathrm{alg}} = q(\Lambda)\, z.$$

- Gaussian scores are linear: $s_t(x) = -(\alpha_t^2 I + \beta_t^2 \Lambda^{-1})^{-1} x$, a *resolvent* of Λ.
- Each score query raises the numerator and denominator degrees by at most one.
- But exact sampling is $\Lambda^{-1/2} z \sim \mathcal{N}(0, \Lambda^{-1})$. Hence:

> Good Gaussian sampling with r queries $\iff$ good rational approximation of $\lambda^{-1/2}$ on $[1, \kappa]$ by $q \in \mathcal{R}_{r,r}$.

(cf. matrix sign function, lattice QCD in physics, Liu 2026, etc.)

An Optimization Analogy

Compare with the classical story for quadratic minimization:

	Gradient descent on $\frac{1}{2}x^T\Lambda x$	Sampling with score queries
Oracle	$\nabla f(x) = \Lambda x$	$s_t(x) = -(\alpha_t^2 I + \beta_t^2 \Lambda^{-1})^{-1} x$
After r steps	polynomial $p(\Lambda)$	rational $q(\Lambda)$
Approximate	0 on $[1, \kappa]$, $p(0) = 1$	$\lambda^{-1/2}$ on $[1, \kappa]$
Optimal	**Chebyshev**	**Zolotarev**
Complexity	$\sqrt{\kappa} \log(1/\varepsilon)$	$\log \kappa \, \log(1/\varepsilon)$

- Chebyshev stepsizes (and fractal/silver stepsizes) are the GD analogue of a "schedule".
- Rational approximation is exponentially better in κ: from $\sqrt{\kappa}$ to $\log \kappa$.

Zolotarev's Problem (1877)

Optimal relative-error approximation of the inverse square root:

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

- Equivalent to Zolotarev's best rational approximation of $\mathrm{sign}(x)$ on $[-\sqrt{\kappa}, -1] \cup [1, \sqrt{\kappa}]$: the optimal approximant is odd, $Z(x) = x\, q_{\kappa,r}(x^2)$, and

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

- Widely used in numerical linear algebra (matrix sign, polar decomposition).

Zolotarev's Problem (1877)

Optimal relative-error approximation of the inverse square root:

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

- Equivalent to Zolotarev's best rational approximation of $\mathrm{sign}(x)$ on $[-\sqrt{\kappa}, -1] \cup [1, \sqrt{\kappa}]$: the optimal approximant is odd, $Z(x) = x\, q_{\kappa,r}(x^2)$, and

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

- Widely used in numerical linear algebra (matrix sign, polar decomposition).
- **Rates:** with $\rho_\kappa = \exp(\pi K'/K)$,

$$E_{\kappa,r} \sim 4\rho_\kappa^{-(2r+1)}, \qquad \log \rho_\kappa \sim \frac{\pi^2}{\log(16\kappa)}.$$

Note: the *same* ratio K'/K as in the strip width!

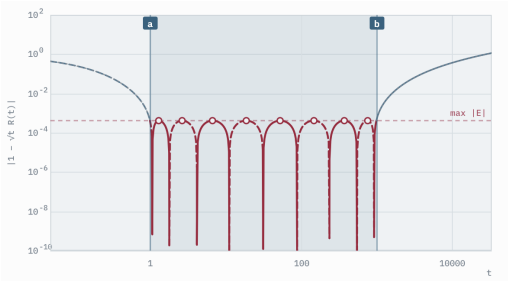
Zolotarev's Explicit Solution

The optimal approximant is given in terms of elliptic functions. For $i = 1, \ldots, 2r$, let

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

Then

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



- Zeros and poles interlace on the negative axis, at $-\gamma_{2j}$ and $-\gamma_{2j-1}$.
- Uniform grid in elliptic time u, evaluated through β_t^2/α_t^2, the SNR of the elliptic schedule!
- The relative error equioscillates at $2r+2$ points.

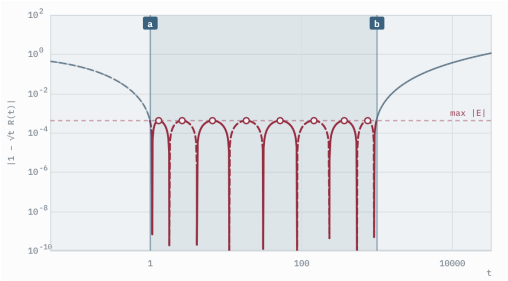
Zolotarev's Explicit Solution

The optimal approximant is given in terms of elliptic functions. For $i = 1, \ldots, 2r$, let

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

Then

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



- Zeros and poles interlace on the negative axis, at $-\gamma_{2j}$ and $-\gamma_{2j-1}$.
- Uniform grid in elliptic time u, evaluated through β_t^2/α_t^2, the SNR of the elliptic schedule!
- The relative error equioscillates at $2r+2$ points.
- **Q:** Is there a *dynamical* realization, i.e., a schedule + ODE solver, with output $q_{\kappa,r}(\Lambda)z$?

Zoloflow

Algorithm (Zoloflow: a custom ODE solver for the elliptic schedule)

Input: number of steps r, parameter κ.

1. $x_0 \sim \mathcal{N}(0, I_d)$; $h = \frac{2}{2r+1}$, $t_j = \frac{2j-1}{2r+1}$.
2. Elliptic schedule: $m = 1 - \kappa^{-1}$, $\alpha_t = \mathrm{cn}(Kt \mid m)$, $\beta_t = \mathrm{sn}(Kt \mid m)$.
3. For $j = 1, \ldots, r$:
$$x_j \leftarrow x_{j-1} + \frac{\beta^2_{t_j + h/2} - \beta^2_{t_j}}{\beta^2_{t_j}} \big(x_{j-1} + s_{t_j}(x_{j-1})\big)$$
4. Return x_r.

- A natural "log-finite-difference" step: $\frac{\dot\beta_t}{\beta_t} = \frac{1}{2\beta_t^2}\frac{d}{dt}\beta_t^2$, approximated with a half step.
- Can be applied to *any* target distribution; only needs one score query per step.
- On Gaussians, Zoloflow gives $x_r = \widehat{q}_{\kappa,r}(\Lambda) z$ exactly.
- The two lenses (analyticity and rational approximation) lead to the same elliptic schedule.

Main Result: Gaussian Sampling

Theorem (minimax query complexity)

There are universal constants $c, C > 0$ such that, for all $d \geq 1$, $\kappa \geq 2$, and ε small enough, the minimal number of exact score queries for an algorithm in $\mathcal{A}_r$ to reach TV error ε uniformly over G_κ satisfies

$$c \, \log \kappa \, \log\left(\frac{\sqrt{d}}{\varepsilon}\right) \;\leq\; r^*(d, \kappa, \varepsilon) \;\leq\; C \, \log \kappa \, \log\left(\frac{\sqrt{d}}{\varepsilon}\right).$$

Zoloflow attains the upper bound.

Main Result: Gaussian Sampling

Theorem (minimax query complexity)

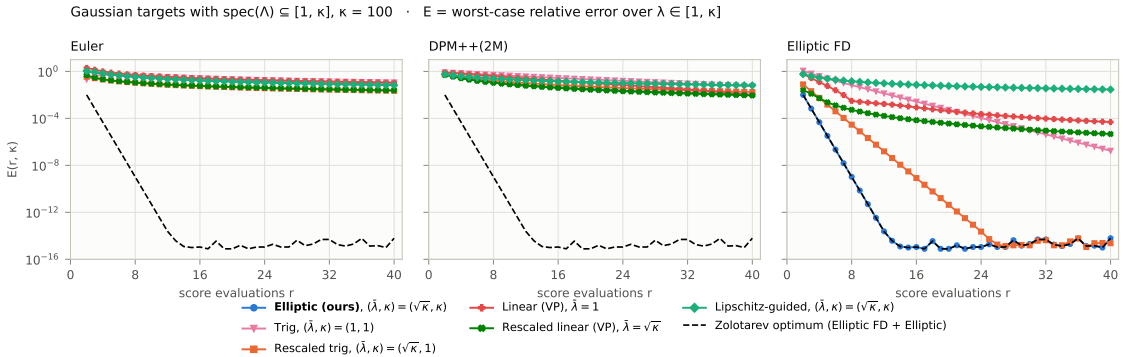
There are universal constants $c, C > 0$ such that, for all $d \geq 1$, $\kappa \geq 2$, and ε small enough, the minimal number of exact score queries for an algorithm in $\mathcal{A}_r$ to reach TV error ε uniformly over G_κ satisfies

$$c \, \log \kappa \, \log\left(\frac{\sqrt{d}}{\varepsilon}\right) \;\leq\; r^*(d, \kappa, \varepsilon) \;\leq\; C \, \log \kappa \, \log\left(\frac{\sqrt{d}}{\varepsilon}\right).$$

Zoloflow attains the upper bound.

- Upper bound: Zolotarev rates + Gaussian TV/KL bounds, $D_{\mathrm{KL}} \lesssim d \, E_{\kappa,r}^2$.
- Lower bound: *any* algorithm in $\mathcal{A}_r$ is a rational function, hence no better than Zolotarev.
- Compare Liu (2026), smoothed scores + sinc quadrature: $O\big((\log \kappa + \log \frac{\sqrt{d}}{\varepsilon}) \log \frac{\sqrt{d}}{\varepsilon}\big)$.
- Classical (unsmoothed) log-concave samplers: polynomial in κ.

Gaussian Experiments



Worst-case relative error E vs. number of score evaluations r, $\mathrm{spec}(\Lambda) \subset [1, 100]$.

Panels: Euler, DPM-Solver++(2M), elliptic finite-difference. Dashed: Zoloflow.

- With standard solvers (Euler, DPM-Solver++), all schedules stall far above the optimum.
- Elliptic schedule + elliptic finite differences (Zoloflow) tracks the Zolotarev optimum: exponential convergence down to machine precision.

Log-Concave Targets

- Scores are now nonlinear, so the rational approximation argument no longer applies. Use the analyticity lens instead.
- *Holomorphic width* $\Omega_B(b)$: largest ω such that all trajectories with $\|z\| \le B$ extend holomorphically to S_ω. Collocation needs $\approx \max(L, \omega^{-1})\,\mathrm{polylog}$ steps.
- Analyticity can fail at $t = 1$; fix by mild Gaussian smoothing ($=$ early stopping).

Log-Concave Targets

- Scores are now nonlinear, so the rational approximation argument no longer applies. Use the analyticity lens instead.
- *Holomorphic width* $\Omega_B(b)$: largest ω such that all trajectories with $\|z\| \le B$ extend holomorphically to S_ω. Collocation needs $\approx \max(L, \omega^{-1})\,\mathrm{polylog}$ steps.
- Analyticity can fail at $t = 1$; fix by mild Gaussian smoothing ($=$ early stopping).

Theorem (holomorphic width)

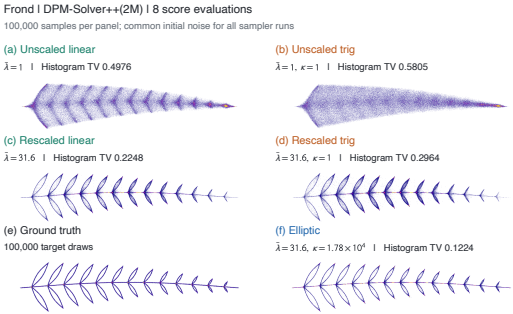
$\omega_\star$: best worst-case width over P_κ (smoothing $\widetilde{\sigma}$, $\|z\| \le B$). Lower bound: elliptic schedule, for *every* π. Upper bound: Gaussians.

$$\frac{1}{D\log(4/\widetilde{\sigma})} \lesssim \omega_\star \lesssim \frac{\pi}{\log(16\widetilde{\kappa})}, \qquad \widetilde{\kappa} \le \kappa \text{ (smoothed)}, \quad D \asymp \sqrt{d\log\kappa} + B.$$

Theorem (complexity)

Elliptic schedule + collocation reaches W_2 error ε with $O\left(\sqrt{d}\,\log^{3.5}\frac{\kappa d}{\varepsilon}\right)$ score evaluations.

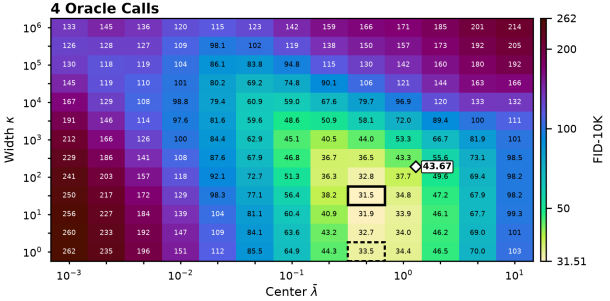
Non-Log-Concave Examples





- Elliptic family has two parameters: center $\bar{\lambda}$ (target rescaling, possible for any schedule) and width κ (new).
- **2D frond:** DPM-Solver++(2M), exact scores, 8 evaluations. Rescaled trig (d) vs. elliptic (f) at the same center $\bar{\lambda}$ isolates the effect of the *width* κ: histogram TV 0.30 $\rightarrow$ 0.12.
- **Images:** Stable Diffusion 3.5M, 4 evaluations. Left to right: linear, trig, elliptic.

Image Generation



SiT-XL/2, FID-10K over the (center, width) grid $(\bar{\lambda}, \kappa)$, 4 calls.

Diamond: covariance heuristic.

SiT-XL/2 (ImageNet), FID-50K, rescaled schedules

Calls	Schedule	FID↓	IS↑
4	Linear	32.36	92.5
	Trig	27.34	101.8
	Elliptic	**25.46**	**108.5**
8	Linear	6.07	198.4
	Trig	4.52	201.9
	Elliptic	**3.85**	**212.6**
32	Linear	2.31	249.6
	Trig	2.07	250.3
	Elliptic	**2.06**	**250.4**

- Largest gains at *low budgets*; as r grows, all schedules converge to the same exact flow.
- Similar results for Stable Diffusion 3.5M (text-to-image).

Summary and Open Questions

- Joint design of schedule + discretization, as a minimax problem.
- Two complementary lenses, same answer: the elliptic schedule $(\mathrm{cn}, \mathrm{sn})$.
  - Analyticity: conformal map, optimal strip width K'/K.
  - Gaussians: rational approximation, Zolotarev, minimax-optimal Zoloflow, $\Theta(\log \kappa \log(\sqrt{d}/\varepsilon))$.
- Log-concave: $O(\sqrt{d}\,\mathrm{polylog}(\kappa, d, 1/\varepsilon))$. Strong empirical performance beyond.

Summary and Open Questions

- Joint design of schedule + discretization, as a minimax problem.
- Two complementary lenses, same answer: the elliptic schedule $(\mathrm{cn}, \mathrm{sn})$.
  - Analyticity: conformal map, optimal strip width K'/K.
  - Gaussians: rational approximation, Zolotarev, minimax-optimal Zoloflow, $\Theta(\log \kappa \log(\sqrt{d}/\varepsilon))$.
- Log-concave: $O(\sqrt{d}\,\mathrm{polylog}(\kappa, d, 1/\varepsilon))$. Strong empirical performance beyond.

Open questions:

- Approximate (learned) scores?
- Optimal discretization for log-concave targets? Matching lower bounds? Is $\sqrt{d}$ necessary?
- Principled choice of parameters (e.g., κ) for general distributions.
- Other classical approximation-theory gems hiding in generative modeling?

A Few References


C. Becker, T. Kargin, K. Hassibi, A. Ozdaglar, and P. A. Parrilo,
"Elliptic Sampling: Optimal interpolant schedules for Gaussians and beyond," arXiv:2611.18542, 2026.


M. Albergo, N. M. Boffi, and E. Vanden-Eijnden,
"Stochastic interpolants: A unifying framework for flows and diffusions," *JMLR* 26(209):1–80, 2025.


E. I. Zolotarev,
"Application of elliptic functions to questions of functions deviating least and most from zero," 1877.


Y. Nakatsukasa and R. 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.


J. Liu,
"Smoothed score queries and the complexity of sampling," arXiv:2605.27769, 2026.


Y. Chen, E. Vanden-Eijnden, and J. Xu,
"Lipschitz-guided design of interpolation schedules in generative models," arXiv:2509.01629, 2025.


Y. T. Lee, Z. Song, and S. Vempala,
"Algorithmic theory of ODEs and sampling from well-conditioned logconcave densities," arXiv:1812.06243, 2018.


A. Sabour, S. Fidler, and K. Kreis,
"Align your steps: Optimizing sampling schedules in diffusion models," *ICML* 2024.

Thank You for your Attention!

Questions?