NYU | COURANT

Beyond Gaussian tilting: Importance sampling for rare events in stochastic systems

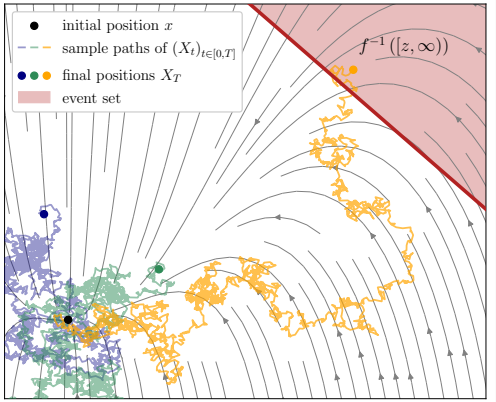
Georg Stadler*

collaborators: Mingxin Li*, Hans Reimann†, Timo Schorlepp*

*Courant Institute School of Mathematics, Computing and Data Science, New York University
†Institut für Mathematik, Universität Heidelberg

Sep 27, 2026

Rare events in stochastic dynamics



Small-noise SDE (or SPDE) on $[0,T]$:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sqrt{\varepsilon}\,\sigma\,\mathrm{d}B_t$$

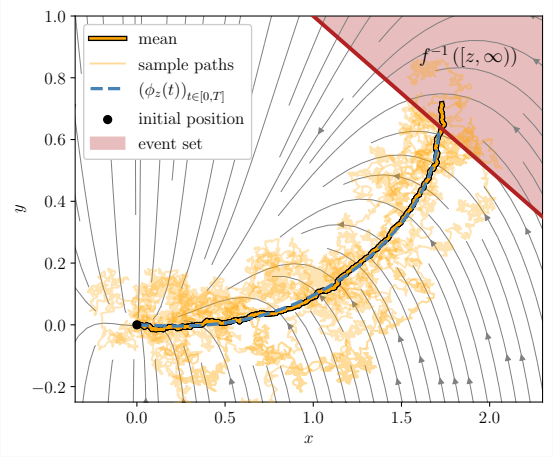
Observable f, threshold z: rare event $\{f(X_T) \geq z\}$.

Question

$P_\varepsilon(z) = \mathbb{P}\big[f(X_T) \geq z\big]$ as $\varepsilon \downarrow 0$ – how large, and how to estimate it?

Gray: flow of b. The noise must work *against* the drift.

Direct sampling: 100 hits in 12 million runs



$\varepsilon = 0.5$, event $x + 2y \geq 3$:

$$\begin{cases} \mathrm{d}x = (-x - xy)\,\mathrm{d}t + \sqrt{\varepsilon}\,\mathrm{d}B^x \\ \mathrm{d}y = (-4y + x^2)\,\mathrm{d}t + \frac{1}{2}\sqrt{\varepsilon}\,\mathrm{d}B^y \end{cases}$$

- $1.2 \cdot 10^7$ runs for 100 hits: $P \approx 8 \cdot 10^{-6}$
- Halve ε: hopeless
- The rare paths cluster around one path – the instanton

Large deviation theory: the instanton

Noise $\sqrt{\varepsilon}\,\eta$, η standard Gaussian ($\mathbb{R}^N$ or L^2); solution map $F(\eta) = f(X_T[\eta])$, so $P_\varepsilon(z) = \mathbb{P}\big[F(\sqrt{\varepsilon}\,\eta) \geq z\big]$.

Rate function and instanton

$$I(z) = \min_{F(\eta)\geq z} \tfrac{1}{2}\|\eta\|^2, \quad \eta_z = \arg\min \qquad \Rightarrow \qquad P_\varepsilon(z) = e^{-I(z)/\varepsilon + o(1/\varepsilon)}$$

- η_z: the *most likely way* for the event to happen ("design point").
- Computable by adjoint-based optimization, even for SPDEs.
- $\eta_z = \lambda_z \nabla F(\eta_z)$: the instanton is normal to the boundary.

Sharp asymptotics: the prefactor

Laplace's method around η_z gives more than the exponent:

Sharp estimate [Breitung '84; Schorlepp, Tong, Grafke, S. '23]

$$P_\varepsilon(z) \overset{\varepsilon\downarrow 0}{\sim} \sqrt{\tfrac{\varepsilon}{2\pi}}\, C_{\mathrm{SORM}}(z,\lambda_z)\, e^{-I(z)/\varepsilon}, \quad C_{\mathrm{SORM}}(z,\mu) := \left[2I(z)\det(\mathrm{Id}-\mu H_z)\right]^{-\frac{1}{2}}$$

- $H_z = \mathrm{pr}_{\eta_z^\perp} \nabla^2 F(\eta_z)\, \mathrm{pr}_{\eta_z^\perp}$: curvature of the event boundary at η_z.
- Needs $\mathrm{Id} - \lambda_z H_z \succ 0$.
- Infinite dimensions: Fredholm determinant, computed from the *dominant eigenvalues* of H_z – typically $O(10)$–$O(100)$, independent of the discretization.

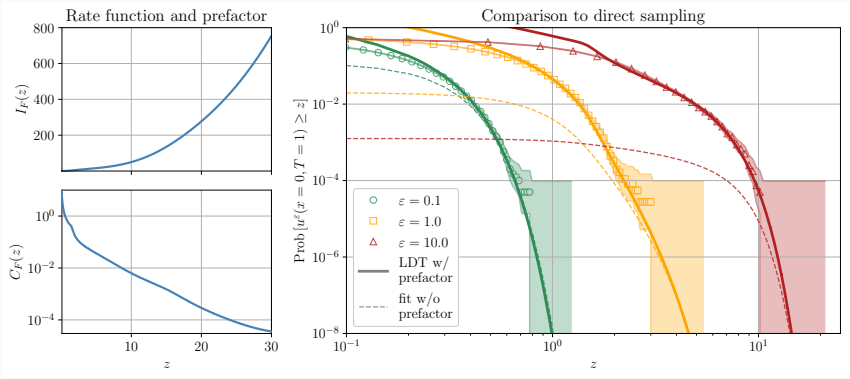
What this needs: the assumptions behind sharp LDT

Standing assumptions (finite-dimensional setting, $\eta \in \mathbb{R}^N$)

1. F is smooth (C^3).
2. $\min_{F(\eta)\geq z} \frac{1}{2}\|\eta\|^2$ has a unique global minimizer η_z.
3. $F(\eta_z) = z$ and $\nabla F(\eta_z) \neq 0$.
4. Second-order condition: $\mathrm{Id} - \lambda_z H_z \succ 0$ on $\eta_z^{\perp}$.

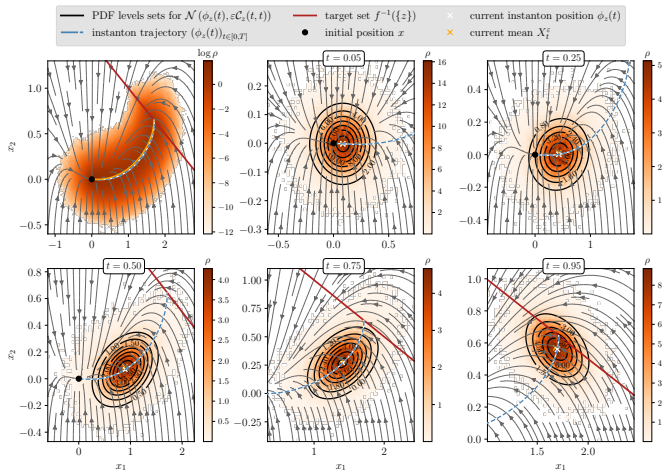
- 1, 3, 4 are *local*: checkable from the optimization output.
- 2 is *global*: hard to verify in high dimensions. Fails for metastable transitions (families of instantons); we stay with "large excursions".
- Infinite dimensions: same, with H_z trace class; rigorous for SDEs, empirical for SPDEs.
- For *sampling* we will need a little more – stated when it appears.

The prefactor matters: stochastic Korteweg–de Vries



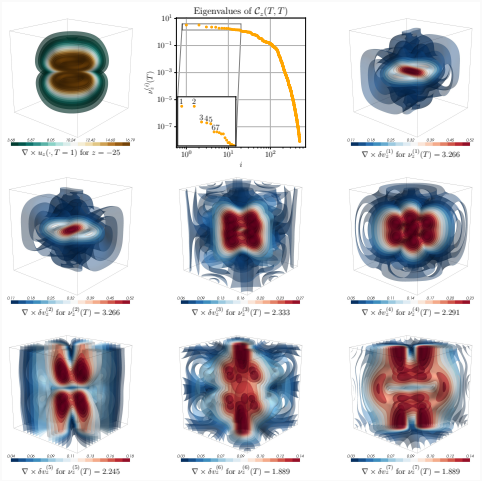
- $\partial_t u + u\,\partial_x u - \nu\,\partial_{xx} u + \kappa\,\partial_{xxx} u = \sqrt{\varepsilon}\,\eta$, observable $u(0,T)$.
- Sharp estimate (solid) vs. direct sampling (markers): no fitted parameters. Dashed: rate function with a constant prefactor.

What the prefactor knows



Histograms: 10^5 conditioned paths. Black: Gaussian with covariance from the eigenpairs of H_z. *We know where the event happens and what it looks like there.*

. . . and this scales: 3D Navier–Stokes



Strain event $\partial_3 u_3(0,T) = -25$: instanton (vortex-ring pair) and dominant fluctuation directions, from 600 eigenpairs of H_z.

This talk

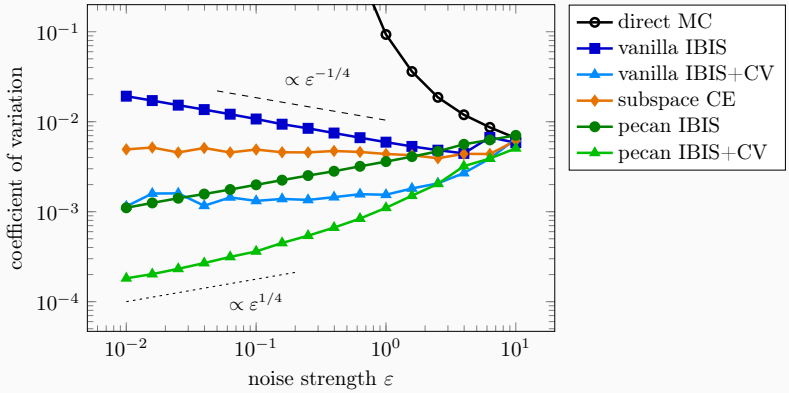
Question

We can compute the instanton *and* the local geometry at scale. Can we turn this into an importance sampler with guarantees?

1. Shifting a Gaussian to the instanton ("vanilla IBIS") is log-efficient, yet its relative error diverges like $\varepsilon^{-1/4}$.
2. Why: curvature and anisotropic fluctuations are ignored.
3. Two remedies: a control variate (bounded error) and a non-Gaussian proposal (vanishing error).
4. Two ways to measure success: coefficient of variation and KL-based sample size.
5. Numerics: 2D SDE and stochastic KdV, down to $P \sim 10^{-270}$.

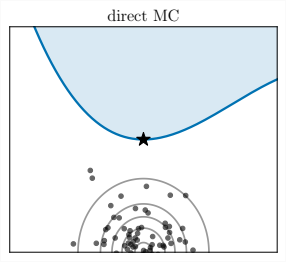
Six samplers on the toy SDE: coefficient of variation

For unbiased estimator $\hat{P}$ of $P_\varepsilon(z)$: $\mathrm{CoV}(\hat{P}) = \sqrt{\mathrm{Var}(\hat{P})}\,/\,\mathbb{E}[\hat{P}]$: relative error.



- 10^5 samples per point; $P_\varepsilon(z)$ from 10^{-1} ($\varepsilon = 10$) to 10^{-203} ($\varepsilon = 10^{-2}$).
- Vanilla IBIS helps, but its error grows as the event gets rarer ($\varepsilon^{-1/4}$, dashed).
- The two new samplers (green) get *better* as $\varepsilon \downarrow 0$.

The classical answer: vanilla IBIS



Shift the Gaussian to the instanton (Cramér tilt):

$$\pi_\varepsilon = \mathcal{N}(0, \varepsilon\,\mathrm{Id}) \;\longrightarrow\; \mu_\varepsilon = \mathcal{N}(\eta_z,\ \varepsilon\,\mathrm{Id})$$

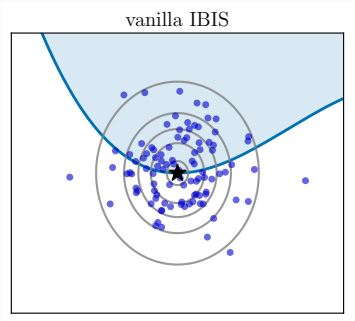
Coordinates at η_z: normal s, tangential u,

$$\eta = \eta_z + \varepsilon\, s\, e_z + \sqrt{\varepsilon}\, u$$

Classical result

Log-efficient: the exponential factor in the variance is gone.

. . . but log-efficiency is not enough



Prefactor estimator $X_\varepsilon = \sqrt{2\pi/\varepsilon}\, \mathbf{1}_{\{F(\eta)\geq z\}}\, e^{-rs}$, $r = \|\eta_z\|$.

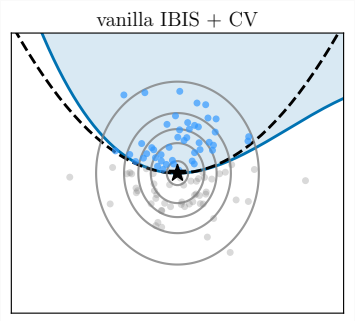
Theorem (vanilla IBIS)

$$\mathrm{CoV}(X_\varepsilon) \overset{\varepsilon\downarrow 0}{\sim} \text{const} \times \varepsilon^{-1/4}, \qquad M \sim \beta\, \delta^{-1} \varepsilon^{-1/2}$$

for relative error δ; constant explicit in $C_{\mathrm{SORM}}(z, \lambda_z)$, $C_{\mathrm{SORM}}(z, 2\lambda_z)$.

Why: the proposal spreads s over a width $\varepsilon^{-1/2}$, but the weight e^{-rs} only lets an $O(1)$ window above the boundary contribute. Effective sample size $M\sqrt{\varepsilon}$.

Remedy 1: a control variate from the parabola



Boundary near η_z is a parabola:
$s \geq \tau(u) = -\frac{\lambda_z}{2r}\langle u, H_z u\rangle$. Use it instead of $F \geq z$, and cancel the Gaussian density of s:

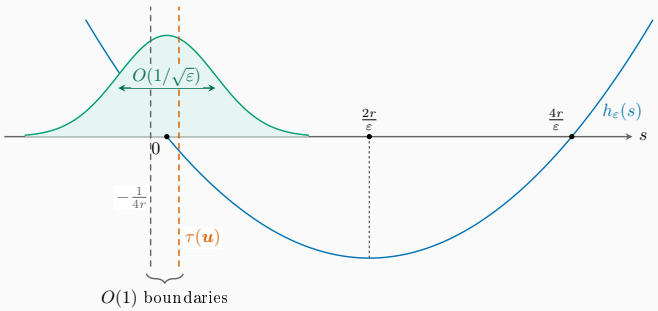
$$Y_\varepsilon = \sqrt{2\pi/\varepsilon}\; \mathbf{1}_{\{s \geq \tau(u)\}}\, e^{-rs+\varepsilon s^2/2}$$

$$\mathbb{E}[Y_\varepsilon] = C_{\mathrm{SORM}}(z, \lambda_z) \quad \text{exactly, for every } \varepsilon$$

Theorem (IBIS + CV)

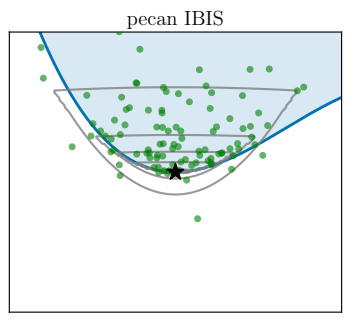
$X_\varepsilon + \hat{c}\,(Y_\varepsilon - C_{\mathrm{SORM}})$ with empirical $\hat{c}$: $\hat{c} \to -1$, and $\mathrm{CoV} \to \mathrm{const}$ as $\varepsilon \downarrow 0$.

Fine print: the control variable has infinite variance



- $\mathbb{E}[Y_\varepsilon^2] \propto \int e^{-2rs+\varepsilon s^2/2}\,\mathrm{d}s = \infty$: the exponent turns around at $s = 2r/\varepsilon$.
- That is $2r/\sqrt{\varepsilon}$ standard deviations out – never seen for $M \ll e^{c/\varepsilon}$.
- Analysis on truncated variables; statements hold with probability $\to 1$.

Remedy 2: change the proposal – pecan IBIS



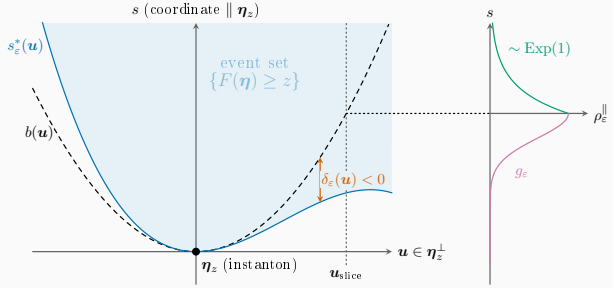
Piecewise-**E**xponential, **C**urvature-corrected **AN**isotropic:

- *Tangential:* $\sqrt{\varepsilon}\, G_z^{-1/2} u$, $G_z := \mathrm{Id} - \lambda_z H_z$ – the conditioned Gaussian.
- *Above the parabola* $b(u)$ ($= \tau(u)$ in these coordinates): $s = b(u) + \mathrm{Exp}(1)$ – the exact limit law.
- *Below,* with probability w_ε: $s = b(u) - V$, $V \sim g_\varepsilon$.

$$P_\varepsilon(z) = \sqrt{\tfrac{\varepsilon}{2\pi}}\, C_{\mathrm{SORM}}(z, \lambda_z)\, e^{-I(z)/\varepsilon}\, \mathbb{E}[J_\varepsilon]$$

The sharp asymptotics are factored out; we only sample what is left: $\mathbb{E}[J_\varepsilon] \to 1$.

Pecan IBIS: what happens below the parabola



- True boundary: $s^*_\varepsilon(u) = b(u) + \sqrt{\varepsilon}\, c_3(u) + O(\varepsilon)$, c_3 odd.
- Where $c_3 < 0$ the event pokes below the parabola – the bottom branch covers it.
- Take any density h on the $\sqrt{\varepsilon}$ scale: $g_\varepsilon(v) = \varepsilon^{-1/2} h(\varepsilon^{-1/2} v)$.

Pecan IBIS: theory

Theorem (pecan IBIS, weight $w_\varepsilon = \alpha\sqrt{\varepsilon}$)

$$\mathrm{Var}(J_\varepsilon) = \Big(\alpha + \tfrac{D_h}{\alpha} - m_1\Big)\sqrt{\varepsilon} + o(\sqrt{\varepsilon}), \qquad \mathrm{CoV}(J_\varepsilon) \propto \varepsilon^{1/4} \to 0$$

$D_h = \mathbb{E}\big[\int_0^{(c_3)-} h^{-1}\big]$, $m_1 = \mathbb{E}[(c_3)_+]$; optimal $w_\varepsilon^* \sim \sqrt{D_h\,\varepsilon}$.

- The relative error vanishes as the event gets rarer.
- Unscaled g still works, but only gives $\mathrm{Var} \propto \varepsilon^{1/4}$.
- In practice: $w_\varepsilon = \min\{\frac{1}{2}, \sqrt{\hat{D}_h\,\varepsilon}\}$, $\hat{D}_h$ from a few boundary evaluations.
- A control variate $\mathbf{1}_{\{s \geq b(u)\}}/(1 - w_\varepsilon)$ (mean 1) comes for free.

The whole family is one algorithm

	vanilla IBIS	pecan IBIS
tangential part of η	$\sqrt{\varepsilon}\,u$	$\sqrt{\varepsilon}\,G_z^{-1/2}u$
normal part of η	$\varepsilon\, s\, e_z$	$(\varepsilon s/r)\, e_z$
law of s given u	$\mathcal{N}(0,\varepsilon^{-1})$	$\mathrm{Exp}(1)$ above $b(u)$, g_ε below
prefactor W_ε	X_ε	J_ε
control V_ε, mean	Y_ε, C_{SORM}	$\mathbf{1}_{\{s\ge b(u)\}}/(1-w_\varepsilon)$, 1

1: **for** $m = 1, \dots, M$ **do**
2: draw $u^{(m)} \sim \mathcal{N}(0, \mathrm{Id})$ and $s^{(m)}$; form $\eta^{(m)}$; evaluate $F(\eta^{(m)})$
3: $W^{(m)} \leftarrow W_\varepsilon(s^{(m)}, u^{(m)})$, $V^{(m)} \leftarrow V_\varepsilon(s^{(m)}, u^{(m)})$
4: **end for**
5: $\hat{c} \leftarrow -\widehat{\mathrm{Cov}}(W, V)/\widehat{\mathrm{Var}}(V)$
6: **return** $c_\varepsilon \frac{1}{M}\sum_m \left[W^{(m)} + \hat{c}\,(V^{(m)} - \mathbb{E}[V_\varepsilon])\right]$

Ingredients: instanton η_z, C_{SORM}, and $r \ll N$ eigenpairs of H_z – exactly what Part I computes. Blue: control-variate steps.

From CoV to sample size

Relative error δ with M samples:
$\mathrm{CoV}(\hat{P}_{M,\varepsilon})^2 = \frac{1}{M}\,\mathrm{CoV}(Z_\varepsilon)^2 = \delta \;\Leftrightarrow\; M = \delta^{-1}\,\mathrm{CoV}(Z_\varepsilon)^2.$

sampler	$\mathrm{CoV}(Z_\varepsilon)$	M for fixed δ	assumptions
direct MC	$P_\varepsilon^{-1/2}$	$\delta^{-1}\,\varepsilon^{-1/2} e^{I(z)/\varepsilon}$	–
vanilla IBIS	$\propto \varepsilon^{-1/4}$	$\beta_{\mathrm{CoV}}\,\delta^{-1}\varepsilon^{-1/2}$	$\mathrm{Id} - 2\lambda_z H_z \succ 0$
vanilla + CV	$\to$ const	$O(\delta^{-1})$	+ dominating η_z for $\hat{c}$
pecan IBIS	$\propto \varepsilon^{1/4}$	$\to 0$	+ cubic boundary term

- The tilt turns exponential cost into polynomial cost; the geometry removes even that.
- All constants are explicit in $I(z)$, λ_z, and the spectrum of H_z.
- Next: is the CoV the right measure at all?

A second criterion: KL divergence and sample size

Optimal proposal $\pi_{\varepsilon,z}$ = law conditioned on the event; ours μ_ε. The CoV *is* a divergence:

$$\mathrm{CoV}(\hat{P}_{M,\varepsilon})^2 = \frac{1}{M}\,\chi^2(\pi_{\varepsilon,z} \,\|\, \mu_\varepsilon), \qquad \chi^2 = \mathbb{E}_{\pi_{\varepsilon,z}}\left[\frac{\mathrm{d}\pi_{\varepsilon,z}}{\mathrm{d}\mu_\varepsilon}\right] - 1 \quad \geq \quad e^{\mathrm{KL}(\pi_{\varepsilon,z}\|\mu_\varepsilon)} - 1$$

It can lie: $\pi = \mathcal{N}(0,\sigma^2)$, $\mu = \mathcal{N}(0,\sigma^2/2)$, event $\{x \geq z\}$: $\chi^2 = \infty$ (huge, unseen weights) but $\mathrm{KL} < \infty$.

Chatterjee & Diaconis (2018), informal

$M \approx \exp\{\mathrm{KL}(\pi_{\varepsilon,z} \,\|\, \mu_\varepsilon)\}$ is *critical*: $M = e^{\mathrm{KL}+2t}$ suffices, $M = e^{\mathrm{KL}-2t}$ fails, w.h.p.

- Separates *necessary* from *sufficient*; needs only $\log$ of the weights, no second moments.

KL results: same rates, different constants

Vanilla IBIS

$$\mathrm{KL} = -\tfrac{1}{2}\log\varepsilon + \log\beta_{\mathrm{KL}} + o(\sqrt{\varepsilon}) \quad \Rightarrow \quad M \sim \beta_{\mathrm{KL}}\,\varepsilon^{-1/2}$$

Pecan IBIS ($w_\varepsilon = \alpha\sqrt{\varepsilon}$)

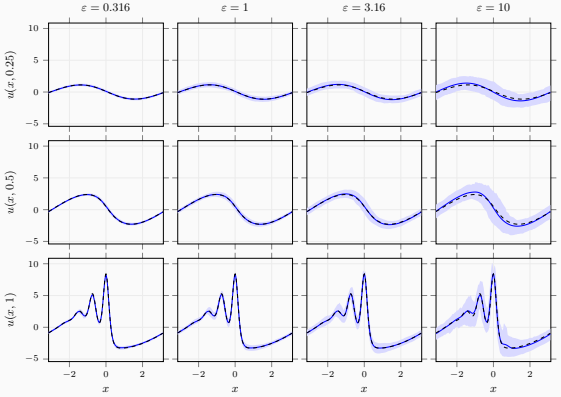
$$\mathrm{KL} = \sqrt{\varepsilon}\left(\alpha + \widetilde{D}_h - \log\alpha\right) + o(\sqrt{\varepsilon}) \quad \Rightarrow \quad M \to 1, \ \text{optimal } \alpha = 1$$

- Same ε-rates as the CoV analysis; constants differ: $\beta_{\mathrm{KL}}/\beta_{\mathrm{CoV}} \le 1/(2e)$.
- KL needs only first-order geometry; variance also needs $\mathrm{Id} - 2\lambda_z H_z \succ 0$.

Stochastic KdV: setup

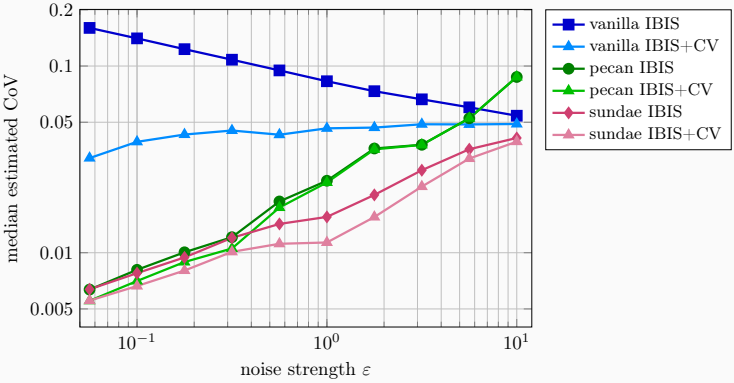
$$\partial_t u + u\,\partial_x u - \nu\,\partial_{xx} u + \kappa\,\partial_{xxx} u = \sqrt{\varepsilon/\pi}\,(\dot{B}_1 \sin x + \dot{B}_2 \cos x)$$

- Event: wave height $u(0,1) \geq 8.39$.
- Noise dimension $N = 2n_t \in \{1000, 2000, 4000\}$.
- 100 dominant curvature modes of H_z.
- 10^4 samples per estimate, 20 repetitions.



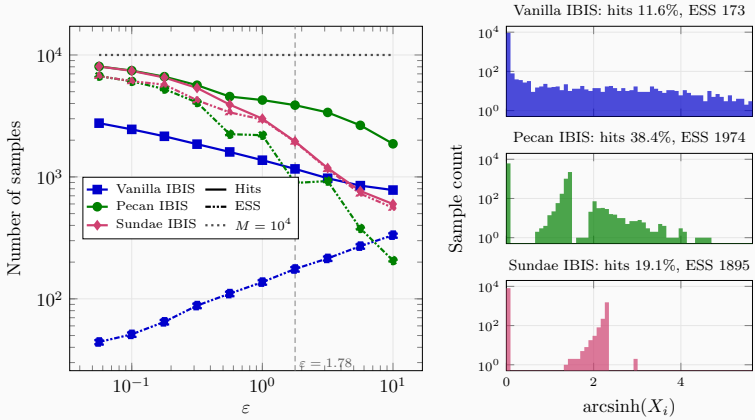
Samples near the event vs. instanton (dashed); ε across, time down.

Stochastic KdV: the six samplers



- $P_\varepsilon(z)$: $3.5 \cdot 10^{-4}$ $(\varepsilon = 10)$, $3 \cdot 10^{-154}$ $(\varepsilon = 0.1)$, $\sim 10^{-270}$ $(\varepsilon = 10^{-1.25})$.
- As in 2D: vanilla grows, +CV flattens, pecan decreases.
- "Sundae": curvature damped by $1/(1+\varepsilon^2)$; one eigenvalue ≈ -57 makes pecan too narrow at large ε.

KdV: hits are not the point, effective sample size is



- Solid: hits; dashed: ESS $= (\sum X_i)^2 / \sum X_i^2$. Vanilla: hits up, ESS *down*.
- $\varepsilon \approx 1.8$: ESS 173 (vanilla), 1974 (pecan), 1895 (sundae, half the hits).

Dimension robustness

$10^2\times$ median CoV, 20 repetitions of $M = 10^4$; $N = 1000$ / 2000 / 4000

	$\varepsilon = 0.1$			$\varepsilon = 1$			$\varepsilon = 10^{0.5}$		
Method	Low	Med	High	Low	Med	High	Low	Med	High
Vanilla IBIS	14.43	14.07	14.12	8.23	8.29	8.41	6.53	6.63	6.63
Vanilla IBIS + CV	4.34	3.93	3.46	4.66	4.64	4.34	4.61	4.88	4.81
Pecan IBIS	0.93	0.81	0.80	1.83	2.43	1.88	4.21	3.79	3.53
Pecan IBIS + CV	0.88	0.71	0.71	1.71	2.38	1.78	4.19	3.78	3.51
Sundae IBIS	0.93	0.78	0.77	1.54	1.56	1.54	2.76	2.77	2.77
Sundae IBIS + CV	0.88	0.66	0.67	1.12	1.13	1.12	2.25	2.26	2.26

Refining the discretization $4\times$ changes nothing: constants depend on the instanton and a few curvature modes, not on N.

Summary and outlook

What we showed

- The Gaussian tilt is log-efficient, yet its error diverges like $\varepsilon^{-1/4}$.
- Sharp asymptotics fix this: control variate (bounded), pecan proposal (vanishing).
- CoV and KL analyses agree on the rates; KL needs weaker assumptions.
- SPDEs with $N \sim 10^3$–10^4, P down to 10^{-270}, N-independent constants.

Open

Pecan + CV jointly; non-unique instantons; infinite-dimensional limit; $z \to \infty$ at fixed ε.

Li, Reimann, Schorlepp, S., *Beyond Gaussian tilting: importance sampling for rare events in stochastic systems*, in preparation.