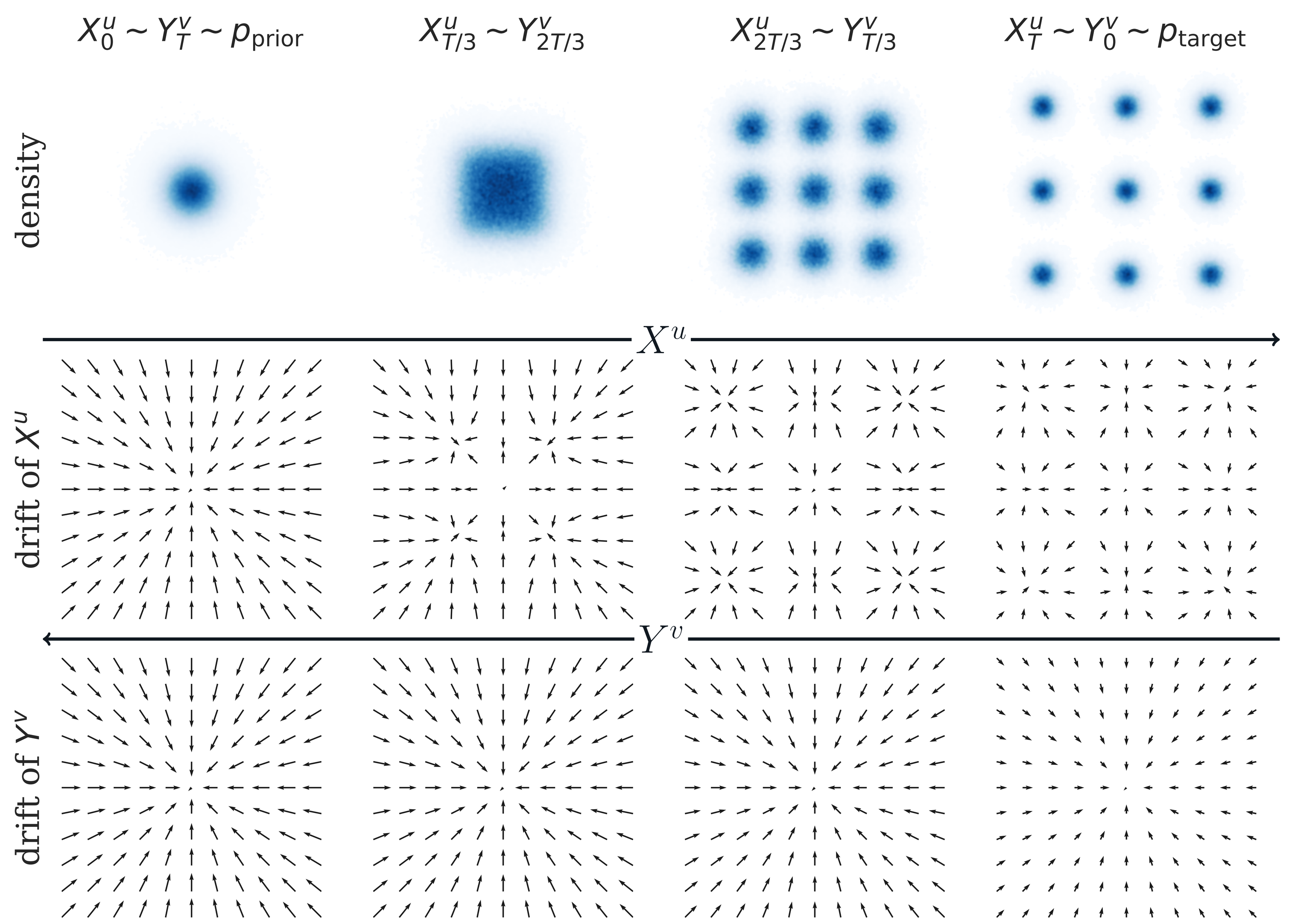
Improved Sampling via Learned Diffusions

Lorenz Richter[⋆,1,2], **Julius Berner**[⋆,3]

[⋆] Equal contribution, [1] dida Datenschmiede GmbH, [2] Zuse Institute Berlin, [3] Caltech

Sampling as time-reversal problem

Recent methods in **generative modeling and sampling** can be viewed as **time-reversals of controlled diffusion processes**.



Setting: Generative and inference SDEs

$$\mathrm{d}X_s^u = (f + \sigma u)(X_s^u, s)\,\mathrm{d}s + \sigma(s)\,\mathrm{d}W_s, \quad X_0^u \sim p_{\text{prior}},$$

$$\mathrm{d}Y_s^v = (-\bar{f} + \bar{\sigma}\bar{v})(Y_s^v, s)\,\mathrm{d}s + \bar{\sigma}(s)\,\mathrm{d}W_s, \quad Y_0^v \sim p_{\text{target}}.$$

Goal: Identify controls u^*, v^* such that (notation: $\bar{\sigma}(t) = \sigma(T-t)$)

$$p_{\text{prior}} \underset{Y^{v^*}}{\overset{X^{u^*}}{\rightleftarrows}} p_{\text{target}}$$

in order to achieve $X_T^{u^*} \sim p_{\text{target}}$ and $Y_T^{v^*} \sim p_{\text{prior}}$.

Path space measure perspective

The **path space measure** perspective provides a **unifying framework**.

Setting: Let D be a divergence, $\mathbb{P}_{X^u}$ be the path space measure of X^u and $\mathbb{P}_{\bar{Y}^v}$ be the path measure of the time reversal of Y^v.

Goal: Identify controls u^*, v^* such that

$$u^*, v^* \in \operatorname*{arg\,min}_{u,v} D\big(\mathbb{P}_{X^u} | \mathbb{P}_{\bar{Y}^v}\big).$$

Connections and equivalences

We show that **SB, DIS, DDS, PIS are special cases** of our framework when using the (reverse) **KL divergence**.

Our proposed tool: Likelihood for path measures is given by

$$\log \frac{\mathrm{d}\mathbb{P}_{X^u}}{\mathrm{d}\mathbb{P}_{\bar{Y}^v}}(X^w) = \int_0^T \Big((u+v)\cdot\Big(w + \frac{v-u}{2}\Big) + \nabla\cdot(\sigma v - f)\Big)(X_s^w, s)\,\mathrm{d}s$$

$$+ \int_0^T (u+v)(X_s^w, s)\cdot \mathrm{d}W_s + \log \frac{p_{\text{prior}}(X_0^w)}{p_{\text{target}}(X_T^w)}.$$

Generalization of recent methods: With $D = D_{\text{KL}}$ we recover:

- Likelihood training of general Schrödinger bridges (Chen et al. 2021).
- Diffusion models and corresponding Time-Reversed Diffusion Sampler (Berner et al. 2024) and Denoising Diffusion Sampler (Vargas, Grathwohl, et al. 2023) with $p_{\text{prior}} = \mathcal{N}(0, \mathrm{I})$ and $v = 0$.
- Path Integral Samplers (Richter 2021; Zhang and Chen 2022; Vargas, Ovsianas, et al. 2023) based on Schrödinger half-bridges with $p_{\text{prior}} = \delta_{x_0}$ and $v = \sigma^\top \nabla \log p_{X^0}$.

The log-variance divergence

The **log-variance divergence** has **provably better properties** than the (reverse) KL divergence.

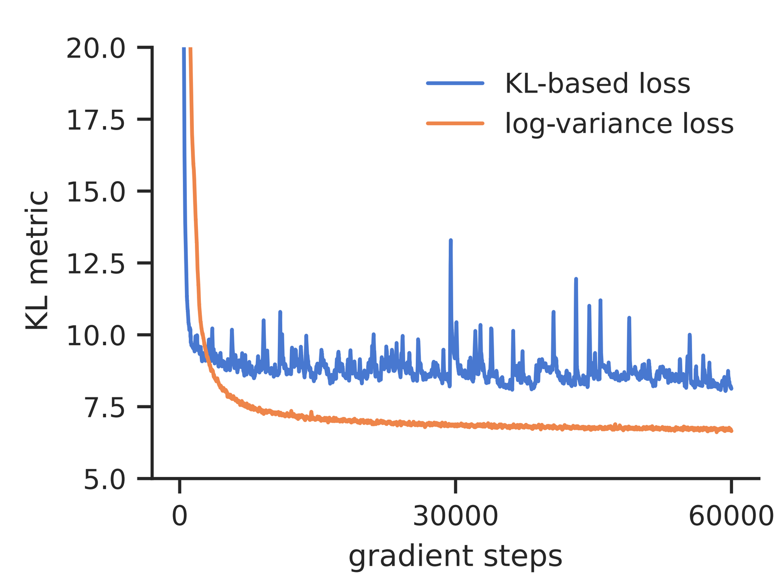
Problem: Reverse KL divergence is prone to mode collapse.

Idea: Our perspective allows for different divergences.

Our proposed divergence: log-variance divergence $D = D_{\text{LV}}$:

$$D_{\text{LV}}^{\mathbb{P}_{X^w}}(\mathbb{P}_{X^u}, \mathbb{P}_{\bar{Y}^v}) := \operatorname{Var}\left[\log \frac{\mathrm{d}\mathbb{P}_{X^u}}{\mathrm{d}\mathbb{P}_{\bar{Y}^v}}(X^w)\right].$$

- Balance exploitation and exploration by the choice of w.
- No differentiation through the SDE solver (detach X^w).
- Gradients have zero variance at the optimum (sticking-the-landing).

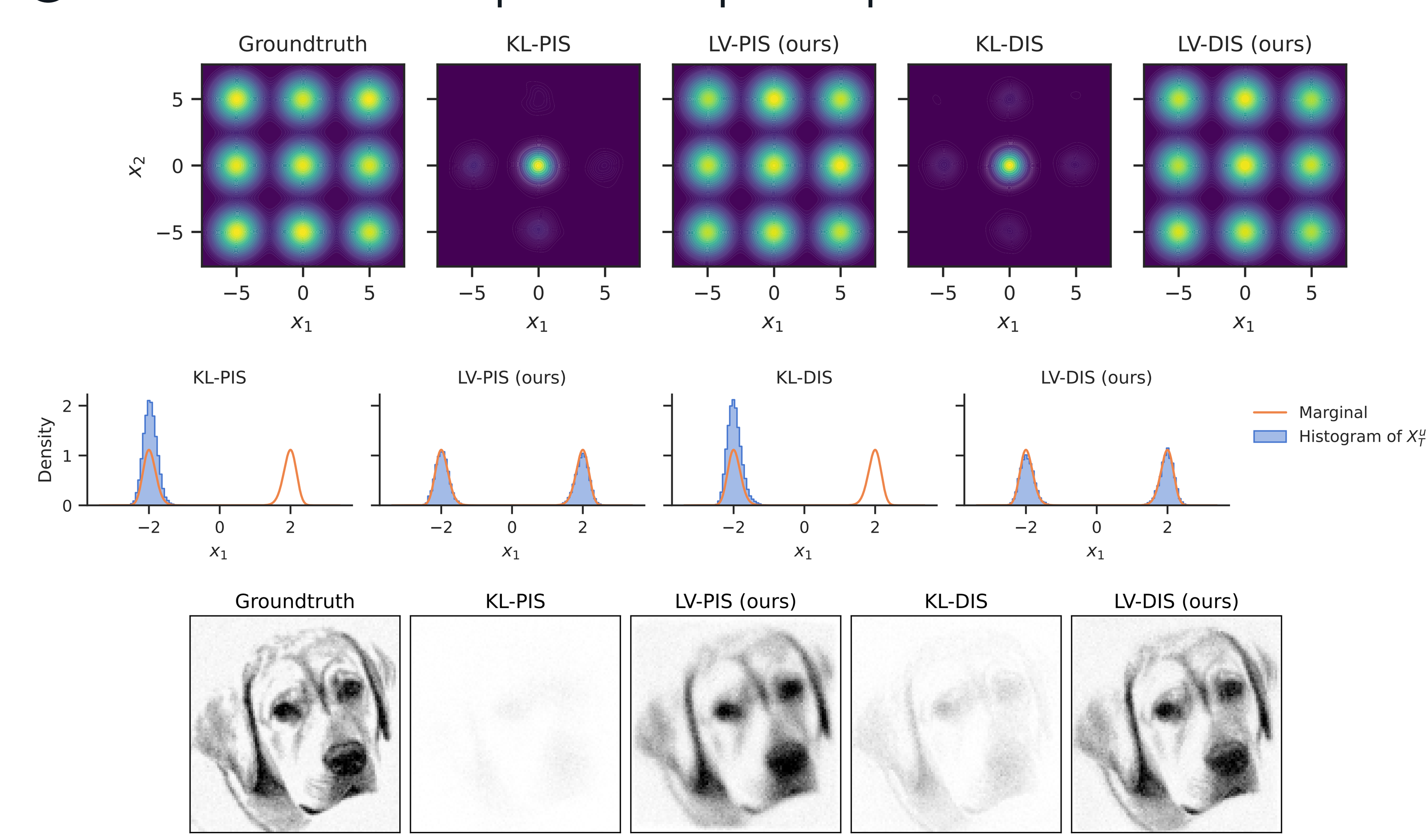


Numerical experiments

The **log-variance divergence** significantly **improves performance** for all considered methods.

Problem	Method	Loss	$\Delta \log Z \downarrow$	$\mathcal{W}_\gamma^2 \downarrow$	ESS $\uparrow$	Δstd $\downarrow$
Gaussian Mixture	PIS	KL	1.094	0.467	0.0051	1.937
$(d=2)$		**LV**	**0.046**	**0.020**	**0.9093**	**0.023**
	DIS	KL	1.551	0.064	0.0226	2.522
		LV	**0.056**	**0.020**	**0.8660**	**0.004**
Funnel	PIS	KL	0.288	5.639	**0.1333**	6.921
$(d=10)$		**LV**	**0.277**	**5.593**	0.0746	**6.850**
	DIS	KL	0.433	5.120	0.1383	5.254
		LV	**0.430**	**5.062**	**0.2261**	**5.220**
Double Well	PIS	KL	3.567	1.699	0.0004	1.409
$(d=5)$		**LV**	**0.214**	**0.121**	**0.6744**	**0.001**
	DIS	KL	1.462	1.175	0.0012	0.431
		LV	**0.375**	**0.120**	**0.4519**	**0.001**
Double Well	PIS	KL	0.101	**6.821**	0.8172	0.001
$(d=50)$		**LV**	**0.087**	6.823	**0.8453**	**0.000**
	DIS	KL	1.785	**6.854**	0.0225	0.009
		LV	**1.783**	6.855	**0.0227**	0.009

Prevents mode collapse and improves performance.



Reference

arxiv.org/abs/2307.01198
github.com/juliusberner/sde_sampler