UNIVERSIDAD DE GRANADA
THE UNIVERSITY OF BRITISH COLUMBIA
Universidad Autónoma de Madrid
UNIVERSITY OF CAMBRIDGE

Conditional Diffusion Sampling

F.M. Castro-Macías[1] P. Morales-Álvarez[1] S. Syed[2] D. Hernández-Lobato[3] R. Molina[1] J. M. Hernández-Lobato[4]

[1]University of Granada [2]University of British Columbia [3]Universidad Autónoma de Madrid [4]University of Cambridge

The Sampling Problem

Goal. Generate independent samples from a distribution with unnormalized density given by:

$$\pi(x) = Z^{-1}\tilde{\pi}(x), \quad Z = \int_{\mathcal{X}} \tilde{\pi}(x)\mathrm{d}x.$$

Challenge. Evaluating $\tilde{\pi}$ is costly: the number of density evaluations is limited.

Why is it difficult? In most cases, π is multimodal. Standard MCMC methods suffer from poor mixing: they become trapped in local modes, failing to explore the distribution.

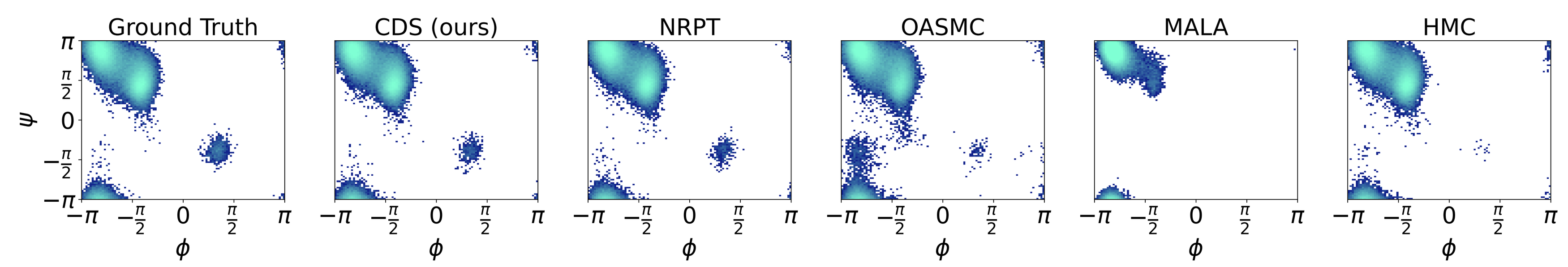


Figure 1. **Ramachandran histograms of Alanine Dipeptide (ALDP) in vacuum at $T = 300$K.** All methods utilize a fixed budget of $2 \cdot 10^5$ density evaluations.

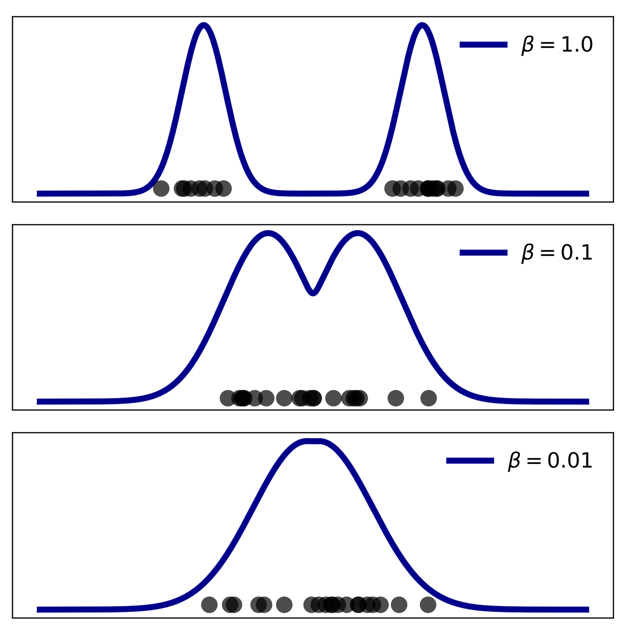
Background: Building Bridges Between Distributions

Idea: Transport samples from a tractable reference distribution π_{ref} to the target distribution π.

Annealing-based methods. Interpolate using a sequence of annealing distributions with density:

$$\pi_\beta(x) \propto \pi_{\mathrm{ref}}(x)^{1-\beta}\pi(x)^\beta, \quad \beta \in [0,1]$$

- **Parallel Tempering (PT, [2]).** Given an annealing schedule $0 = \beta_0 < \cdots < \beta_N = 1$, PT runs several chains in parallel (one for each β_n). It alternates between local MCMC updates and swap moves between adjacent chains.
- **Drawback:** Performance is highly sensitive to the annealing schedule and degrades when the overlap between the reference and the target is poor.



Stochastic Interpolants [1]. Interpolate directly between random variables rather than densities:

$$x_t = F_t(z, x), \quad z \sim \nu_{\mathrm{ref}}, \quad x \sim \nu,$$

- F_t is a differentiable map satisfying the boundary conditions $F_0(z,x) = z$ and $F_1(z,x) = x$.
- The marginal distribution of x_t induces the following **transport dynamics**:

$$\begin{cases} x_0 \sim \pi_{\mathrm{ref}}, \\ \mathrm{d}x_t = a_t(x_t)\mathrm{d}t + \sigma_t \mathrm{d}w_t, \\ x_1 \sim \pi \end{cases} \xleftrightarrow{\mathrm{Law}(x_t) = p_t} \begin{cases} p_0 = \pi_{\mathrm{ref}}, \\ \frac{\partial}{\partial t} p_t = -\operatorname{div}(p_t a_t) + \frac{\sigma_t^2}{2}\Delta p_t, \\ p_1 = \pi \end{cases}$$

- Samples from π can be generated by integrating the dynamics from $x_0 \sim \pi_{\mathrm{ref}}$.
- **Drawback:** a_t needs to be learned using samples from π.

References

[1] 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.

[2] Saifuddin Syed, Alexandre Bouchard-Côté, George Deligiannidis, and Arnaud Doucet. Non-Reversible Parallel Tempering: A Scalable Highly Parallel mcmc Scheme. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):321–350, 2022.

Check our code!

```
github.com/Franblueee/
conditional_diffusion_sampling
```

Conditional Diffusion Sampling (CDS): Overview

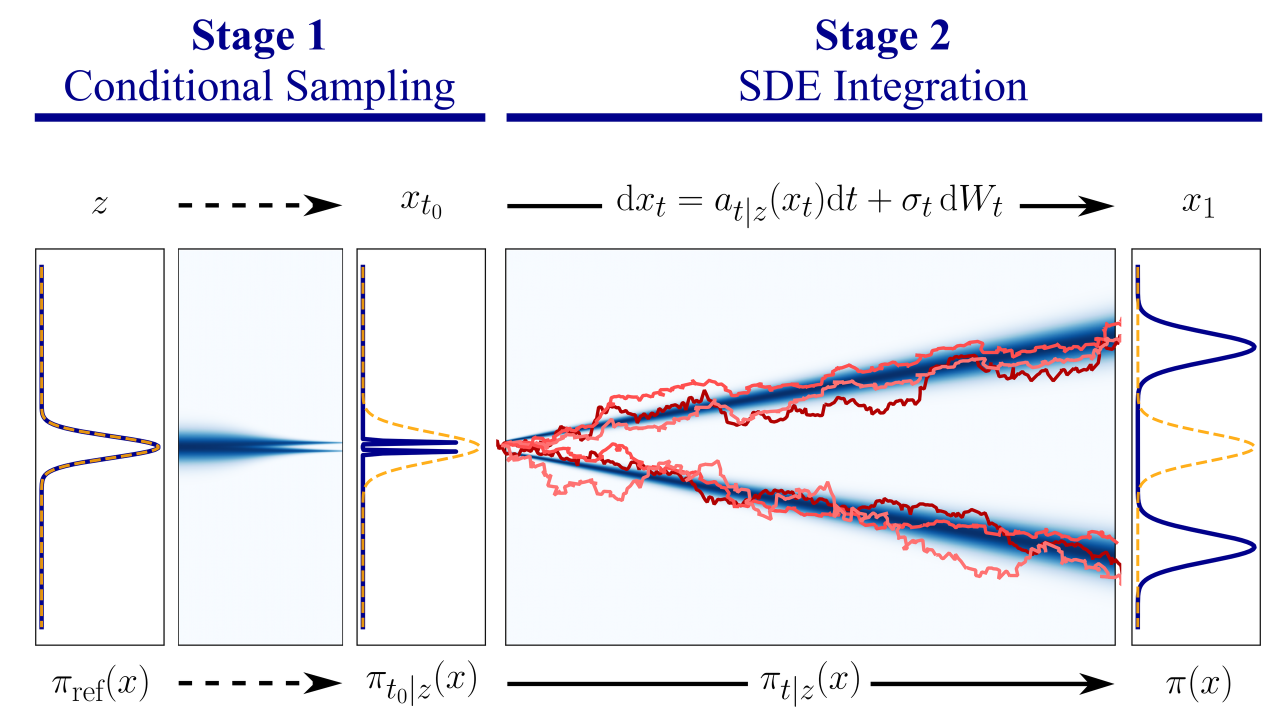
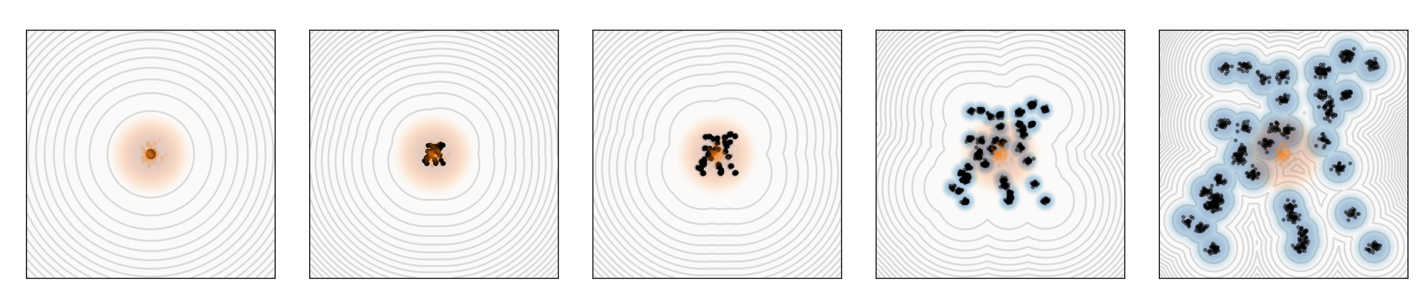


Figure 2. **Overview of CDS.** In the first stage, Parallel Tempering (PT) transforms initial samples $z \sim \pi_{\mathrm{ref}}$ (orange) into samples from the initialization distribution $\pi_{t_0|z}$ (blue). In the second stage, these are transported to the target distribution π (blue) by integrating the closed-form SDE.

- **Conditional Interpolants:** A *conditional path* of distributions $\{\pi_{t|z}\}_{t\in(0,1]}$ evolves from a point mass at $z \sim \pi_{\mathrm{ref}}$ to the target distribution π.
- **Closed-form Dynamics:** Samples are transported along the conditional path by simulating an SDE with a closed-form drift.
- **Dynamics Initialization:** For a small t_0, sampling from $\pi_{t_0|z}$ for a small $t_0 > 0$ ensures substantial overlap with π_{ref}, making global exploration highly efficient.

Conditional Interpolants & Transport Dynamics



(a) $t = 0.0$ (b) $t = 0.1$ (c) $t = 0.2$ (d) $t = 0.5$ (e) $t = 1.0$

Figure 3. **Evolution of $\pi_{t|z}$.** As $t \to 0$, the target distribution (blue) increasingly concentrates inside the reference (orange).

Idea. Consider the conditional distribution of x_t given a reference variable $z \sim \pi_{\mathrm{ref}}$:

$$x_t = F_{t|z}(x) := F_t(z,x), \quad x \sim \nu.$$

Conditional Density. Due to the change of variables:

$$\pi_{t|z}(x) = \left|\det \mathrm{J}F_{t|z}\left(F_{t|z}^{-1}(x)\right)\right|^{-1} \pi\left(F_{t|z}^{-1}(x)\right),$$

Transport Dynamics. It obeys the following SDE:

$$\mathrm{d}x_t = \left(u_{t|z}(x_t) + \frac{\sigma_t^2}{2}\nabla \log \pi_{t|z}(x_t)\right)\mathrm{d}t + \sigma_t\,\mathrm{d}w_t,$$

with $u_{t|z}(x) = \frac{\partial F_{t|z}}{\partial t}\left(F_{t|z}^{-1}(x)\right)$.

Example: Linear Interpolant

$$F_t(z,x) = (1-t)z + tx,$$

$$\pi_{t|z}(x) = t^{-D}\pi\left(\frac{x-(1-t)z}{t}\right),$$

$$u_{t|z}(x) = \frac{x-z}{t},$$

Vanishing Time ($t \to 0$)

Dynamics are ill-defined at $t = 0$. What about sampling?

- **Reduced Transport Distance:** As the interpolant contracts the state space around z, the sampling error becomes negligible for small t (theoretical analysis in the paper).
- **However, in practice:** The multimodal structure of π is preserved, so poor mixing may still occur.
- Overlap with the reference increases → annealing-based methods' performance increases.

(Stage 1) Impact of t_0 in Communication and Sample Quality

Question: How to choose the right initial time t_0?

- We analyze Round Trips (RTs) in PT targeting $\pi_{t_0|z}$: traversals between reference and target distributions. Higher RTs indicate better mixing.
- Decreasing t_0 from 1.0 increases RTs and reduces sampling error, showing that sampling from $\pi_{t_0|z}$ is more efficient than sampling ν directly. As $t_0 \to 0$, performance degrades due to excessive concentration.

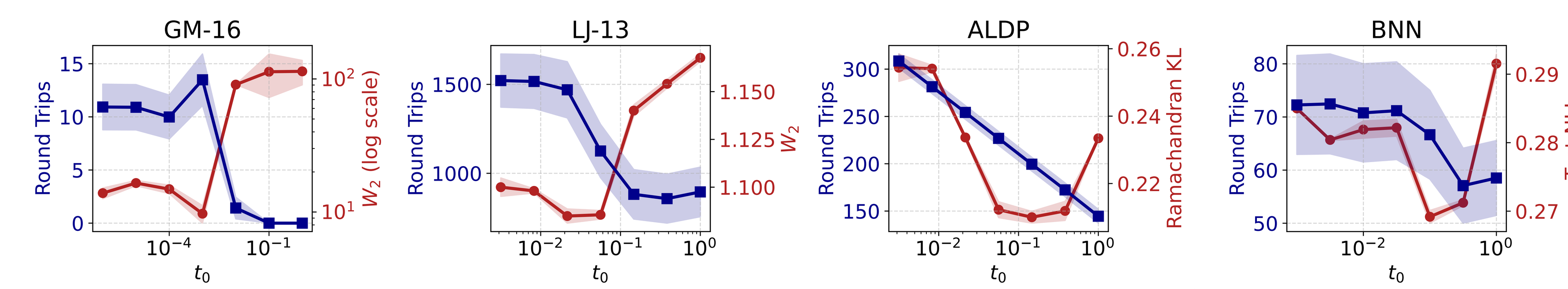


Figure 4. **Round Trips** (RTs, higher is better) and **sampling error** (lower is better) as a function of t_0. Decreasing t_0 from 1.0 generally increases RT counts and reduces sampling error, indicating improved mixing and sample quality.

(Stage 2) Analysis of the Transport Mechanism

Question: What if we replace the SDE with the inverse deterministic transport?

- If $\pi_{t_0|z}$ were sampled perfectly, $F_{t_0|z}^{-1}$ would recover perfect samples from π. In practice, approximation errors are amplified.
- In contrast, the SDE dynamics continuously refine the samples during transport.

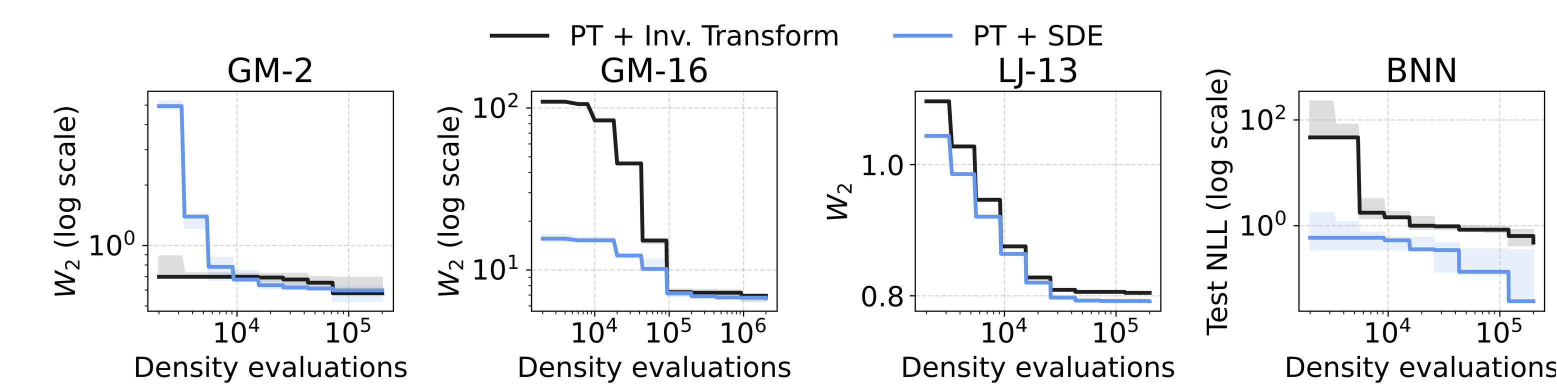


Figure 5. **Comparison between SDE-based transport and the inverse interpolation map.** The SDE-based approach consistently achieves better performance across tasks.

Comparison with other Sampling Methods

- **Sampling Tasks:** Eight target distributions across four tasks spanning synthetic benchmarks (GM-D and GMNU-D), physical systems (LJ-13 and LJ-55), molecular dynamics (ALDP), and high-dimensional Bayesian inference problems (BNN).
- CDS demonstrates higher efficiency by requiring fewer density evaluations for the same level of accuracy.

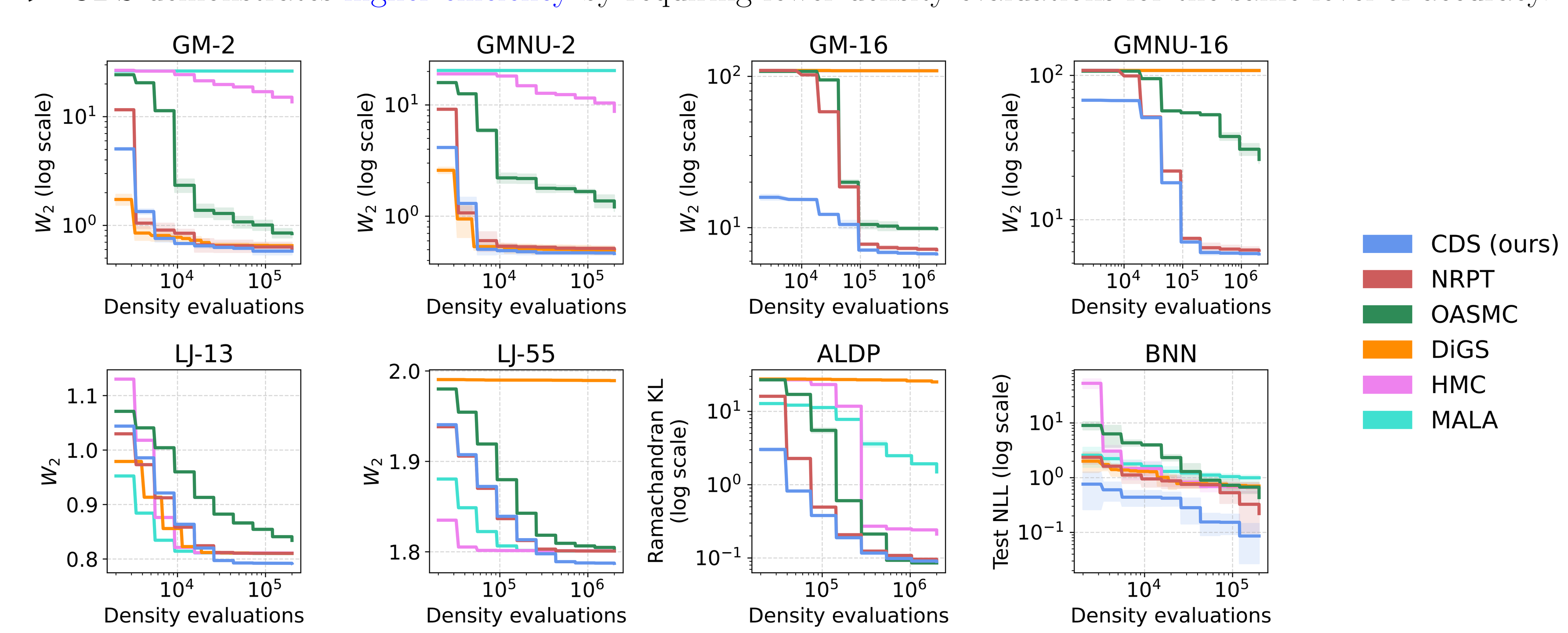


Figure 6. **Pareto fronts for sampling performance across eight target distributions.** The proposed CDS achieves competitive or superior performance compared to state-of-the-art samplers.

fcastro@ugr.es