DEPARTMENT OF
STATISTICS

Guiding Diffusions Towards Singular Rewards

The Case of Diffusion Bridges

Jakiw Pidstrigach, Gridmatic (Work mostly at Univ. of Oxford)

March 19, 2026

Acknowledgements

I would like to thank my collaborators:

Libby Baker
DTU

Sam Howard
Oxford

Nik Nüsken
King's College London

Carles Domingo-Enrich
Microsoft Research

Andreas Bergmeister
TU Munich

George Deligiannidis
QRT

Stefanie Jegelka
TU Munich

Setup

Given a **reference system** X_t (weather, molecular dynamics, diffusion models, flow matching) together with an observation Y:

$$\begin{aligned} \mathrm{d}X_t &= b_t(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \\ Y &= G(X_T), \end{aligned}$$

Setup

Given a **reference system** X_t (weather, molecular dynamics, diffusion models, flow matching) together with an observation Y:

$$\begin{aligned} \mathrm{d}X_t &= b_t(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \\ Y &= G(X_T), \end{aligned}$$

Goal

We want to learn the conditional distribution

$$X \mid Y = y.$$

We will focus on the case $Y = X_T$, i.e. **diffusion bridges**.

Setup

Given a **reference system** X_t (weather, molecular dynamics, diffusion models, flow matching) together with an observation Y:

$$\begin{aligned} \mathrm{d}X_t &= b_t(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \\ Y &= G(X_T), \end{aligned}$$

Goal

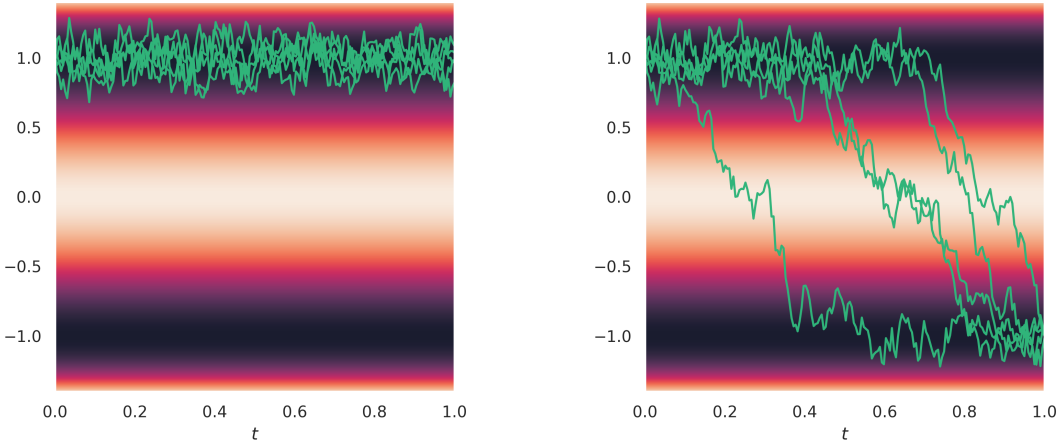
We want to learn the conditional distribution

$$X \mid Y = y.$$

We will focus on the case $Y = X_T$, i.e. **diffusion bridges**.

Example: Transition path sampling in molecular dynamics; interpolating between weather states from a coarser simulation; guidance/reinforcement learning for diffusion models.

A simple diffusion bridge example



Left: unconditioned dynamics in the double-well potential. Right: the conditioned bridge law.

Doob's h-Transform

Naive approach: simulate reference system and only choose paths where $X_T = x$.

Doob's h-Transform

~~Naive approach: simulate reference system and only choose paths where $X_T = x$.~~

Doob's h-Transform

~~Naive approach: simulate reference system and only choose paths where $X_T = x$.~~

Lemma (Doob's h-transform)

We can simulate paths from $X|X_T = x_T$ with a controlled system

$$\mathrm{d}X_t^u = b_t(X_t^u) + u_t(X_t^u; x_T)\mathrm{d}t + \mathrm{d}B_t.$$

with

$$u_t^*(x; x_T) = \nabla_x \log p(X_T = x_T \mid X_t = x).$$

Doob's h-Transform

~~Naive approach: simulate reference system and only choose paths where $X_T = x$.~~

Lemma (Doob's h-transform)

We can simulate paths from $X|X_T = x_T$ with a controlled system

$$\mathrm{d}X_t^u = b_t(X_t^u) + u_t(X_t^u; x_T)\mathrm{d}t + \mathrm{d}B_t.$$

with

$$u_t^*(x; x_T) = \nabla_x \log p(X_T = x_T \mid X_t = x).$$

How do we approximate u^*?

Generalized Tweedie / Score Formula

Theorem (Pidstrigach, Baker, Domingo-Enrich, Deligiannidis, Nüsken (2025))

Assume α_t *integrates to* 1. *Then*

$$u_s^*(x_s; x_T) = \nabla_{x_s} \log p(X_T = x_T \mid X_s = x_s) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \mid X_T = x_T, X_s = x_s\right].$$

Generalized Tweedie / Score Formula

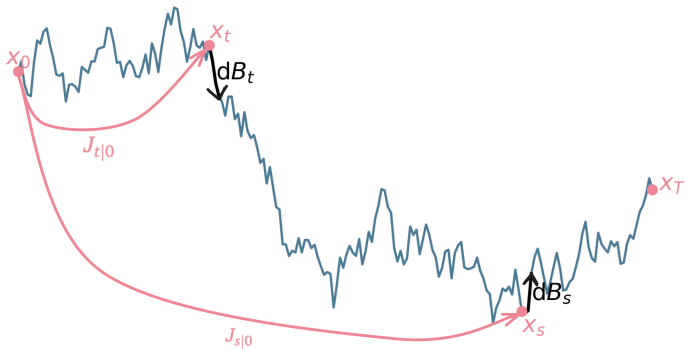
Theorem (Pidstrigach, Baker, Domingo-Enrich, Deligiannidis, Nüsken (2025))

Assume α_t *integrates to* 1. *Then*

$$u_s^*(x_s; x_T) = \nabla_{x_s} \log p(X_T = x_T \mid X_s = x_s) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \mid X_T = x_T, X_s = x_s\right].$$

Theorem (Pidstrigach, Baker, Domingo-Enrich, Deligiannidis, Nüsken (2025))

$$u_s^*(x_s; x_T) = \nabla_{x_s} \log p(X_T = x_T \mid X_s = x_s) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \mid X_T = x_T, X_s = x_s\right].$$



(Very) Rough Proof Sketch: Malliavin Calculus

1. Start from $p_{T|s}(x_T \mid x_s) = \mathbb{E}[\delta_{x_T}(X_T) \mid X_s = x_s]$.
2. $\nabla \log p_{T|s}(x_T \mid x_s) = \frac{1}{p_{T|s}} \nabla \mathbb{E}[\delta_{x_T}(X_T) \mid X_s = x_s]$.
3. Apply Malliavin integration by parts.
4. The derivative leaves the singular object δ_y, producing the stochastic integral $\mathcal{S}_s$.

Very handwavy intuition

Move differentiation off the singular reward δ and onto the path measure. So

$$\mathbb{E}[\nabla g(X)] = \int \nabla g(x) p(x) \mathrm{d}x = -\int g(x) \nabla \log p(x) p(x) \mathrm{d}x = -\mathbb{E}[g \nabla \log p(X)]$$

in infinite dimensions.

Intuition: Self-Consistency

Under the target measure, the original Brownian motion has drift u^*.

$$\begin{aligned} \mathrm{d}X_t &= b_t(X_t)\,\mathrm{d}t + \mathrm{d}B_t, \\ &= b_t(X_t)\,\mathrm{d}t + u_t^*(X_t)\,\mathrm{d}t - u_t^*(X_t)\,\mathrm{d}t + \mathrm{d}B_t \\ &= b_t(X_t)\,\mathrm{d}t + u_t^*(X_t)\,\mathrm{d}t + \mathrm{d}W_t. \end{aligned}$$

with $\mathrm{d}W_t = \mathrm{d}B_t - u_t^*(X_t)\mathrm{d}t$ a Brownian motion under the target measure. Therefore

$$\begin{aligned} u^*(x_s) =& \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \mid X_T = x_T, X_s = x_s\right] \\ =& \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}W_t + u_t^*(X_t)\mathrm{d}t \mid X_T = x_T, X_s = x_s\right] \\ =& \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top u_t^*(X_t)\mathrm{d}t \mid X_T = x_T, X_s = x_s\right] \end{aligned}$$

Intuition: Self-Consistency

For $\alpha = \delta_t$ we get that the optimal control satisfies the following self-consistency property:

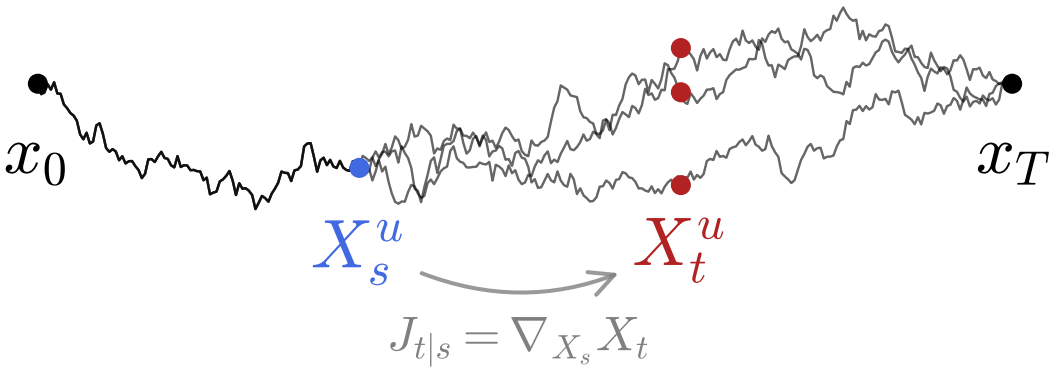
Theorem (Self-consistency: Howard, Nüsken, Pidstrigach (2026))

For any $t \geq s$

$$u^*(s,x) = \mathbb{E}_{\mathbb{P}^{u^*}}\left[\left(\nabla_{X_s} X_t\right)^\top u^*(t, X_t^{u^*}) \mid X_s^{u^*} = x \right].$$

The Optimal Control is Self-Consistent

$$u(s, X_s^u) = \mathbb{E}\left[J_{t|s}^\top \, u(t, X_t^u) \mid X_s^u \right]$$



Big Problem: We cannot simulate from the target (conditioned) measure.

2025: BEL-Algorithm

- Train on data from the reference law $\mathbb{P}$.
- Use **amortization**.
- Pro: Get a clean regression problem.
- Con: support mismatch for rare events.

2026: Consistency Bridges

- Train on data from the current controlled law $\mathbb{P}^{u_\theta}$.
- Transform to a **fixed-point problem**.
- Pro: Better on-policy coverage of rare bridges.
- Con: Training is less stable.

Route 1: Regress on the Score Under the Base System

BEL training objective / Amortization

Train against the stochastic score estimator under the reference diffusion:

$$\mathcal{L}_{\mathrm{BEL}}(\theta) = \mathbb{E}\left[\int_0^T \left\| u_\theta(s, X_s; X_T) - \int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \right\|^2 \mathrm{d}s\right],$$

$$u_\theta(s, x_s, x_T) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \,\middle|\, X_T = x_T, X_s = x_s\right] = u_s^{*,x_T}(x_s).$$

Route 1: Regress on the Score Under the Base System

BEL training objective / Amortization

Train against the stochastic score estimator under the reference diffusion:

$$\mathcal{L}_{\mathrm{BEL}}(\theta) = \mathbb{E}\left[\int_0^T \left\| u_\theta(s, X_s; X_T) - \int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \right\|^2 \mathrm{d}s\right],$$

$$u_\theta(s, x_s, x_T) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \,\middle|\, X_T = x_T, X_s = x_s\right] = u_s^{*,x_T}(x_s).$$

Algorithmic picture: simulate the base process, use X_T as the endpoint label, and amortize one network over all terminal values.

Route 1: Regress on the Score Under the Base System

BEL training objective / Amortization

Train against the stochastic score estimator under the reference diffusion:

$$\mathcal{L}_{\mathrm{BEL}}(\theta) = \mathbb{E}\left[\int_0^T \left\| u_\theta(s, X_s; X_T) - \int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \right\|^2 \mathrm{d}s\right],$$

$$u_\theta(s, x_s, x_T) = \mathbb{E}\left[\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top \mathrm{d}B_t \,\middle|\, X_T = x_T, X_s = x_s\right] = u_s^{*,x_T}(x_s).$$

Algorithmic picture: simulate the base process, use X_T as the endpoint label, and amortize one network over all terminal values.

What it buys and what it misses

Elegant and fully amortized, but still trained on the *uncontrolled* system. For rare events, the bridge support is poorly covered exactly where accuracy matters most.

Route 2: Turn the Control Equation Into a Fixed Point

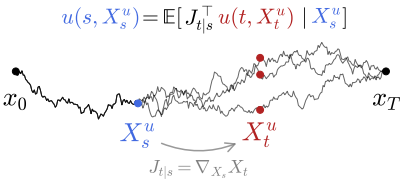
Self-consistency loss from Paper 2

$$\mathcal{L}_{\mathrm{FP}}(\theta) = \mathbb{E}\left[\left\| u_\theta(s, X_s^{\bar{\theta}}) - \mathcal{U}_s \right\|^2\right],$$

$$\mathcal{U}_s = \frac{1}{A_s}\int_s^T \alpha_t \left(\nabla_{X_s} X_t\right)^\top u_{\bar{\theta}}(t, X_t^{\bar{\theta}})\,\mathrm{d}t.$$

Interpretation

- Simulate the *current controlled process* $X^{\bar{\theta}}$.
- Enforce **self-consistency** by regressing against the future values of the current control.
- **Enforce the bridge endpoint** by construction at the final step.



Earlier control values are matched to later control values transported back along the trajectory.

Unique Fixed Point

The following justifies that the fixed point of the above iteration is unique:

Theorem (Howard, Nüsken, Pidstrigach (2026))

Assume u is a control which has the following properties:

1. *Endpoint condition: $X_T^u = x_T$ almost surely.*
2. *Self-consistency: For all $t \geq s$: $u_s(x_s) = \mathbb{E}_{\mathbb{P}^{u^*}}\left[(\nabla_{X_s} X_t)^\top u(t, X_t^u) \mid X_s^u = x\right]$*
3. *It is of gradient form at some point $u_t = \nabla F$ for any t.*

Then $u_s(x_s) = \nabla \log p(X_T = x_T \mid X_s = x_s)$ is the optimal diffusion bridge control.

Proof Sketch: Self-Consistency

For a smooth terminal weight $F(X_T)$, define the tilted path measure

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}(X) \propto F(X_T).$$

Then the optimal control satisfies

$$u^*(s,x) = \nabla_x \log \mathbb{E}_{\mathbb{P}}[F(X_T) \mid X_s = x] = \mathbb{E}_{\mathbb{Q}}\Big[(\nabla_{X_s} X_T)^\top \nabla \log F(X_T) \mid X_s = x\Big].$$

For any $s < t < T$, insert the tower property and the chain rule

$$\nabla_{X_s} X_T = (\nabla_{X_s} X_t)\,(\nabla_{X_t} X_T):$$

$$\begin{aligned} u^*(s,x) &= \mathbb{E}_{\mathbb{Q}}\Big[(\nabla_{X_s} X_t)^\top \mathbb{E}_{\mathbb{Q}}\Big[(\nabla_{X_t} X_T)^\top \nabla \log F(X_T) \mid X_t\Big] \,\Big|\, X_s = x\Big] \\ &= \mathbb{E}_{\mathbb{Q}}\Big[(\nabla_{X_s} X_t)^\top u^*(t, X_t) \mid X_s = x\Big]. \end{aligned}$$

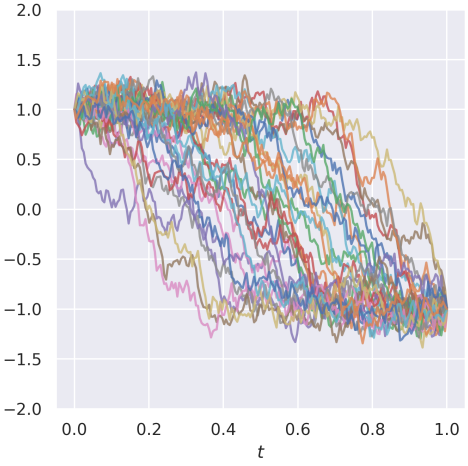
Theorem (Self-consistency: Howard, Nüsken, Pidstrigach)

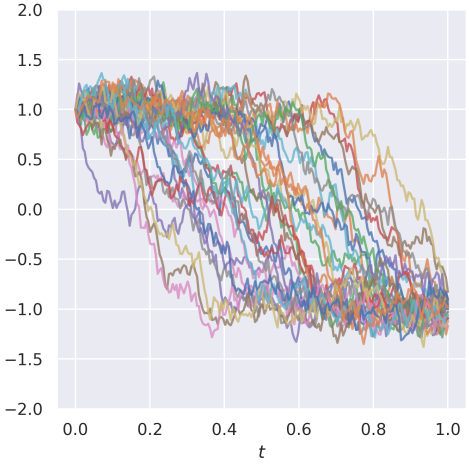
$$u^*(s,x) = \mathbb{E}_{\mathbb{P}^{u^*}}\Big[(\nabla_{X_s} X_t)^\top u^*(t, X_t^{u^*}) \mid X_s^{u^*} = x\Big].$$

And now: Nice images

three figures I like from the projects

Double-Well Marginals

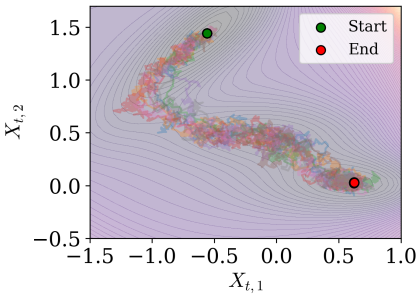


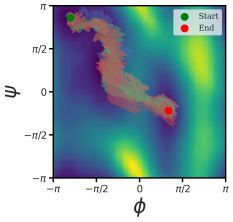


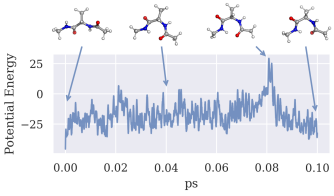
Fashion-MNIST Conditioning

Fashion-MNIST example from *Conditioning Diffusions Using Malliavin Calculus*. Left: conditioned generations. Right: conditioning inputs.

Müller-Brown Potential







Full three-panel figure from the consistency-loss paper: Müller-Brown trajectories, Ramachandran plot, and alanine transition-path energy.

Fine-Tuning Example


astronaut crossing an alien portal


cyberpunk concept vehicle


mid-century modern revival mansion


blue cow and black keyboard


pizza and bowl


four boats

Recreated from Figure 1 in the `fine_tuning_non_diff` project. In each pair: base model on the left, fine-tuned model on the right.

References

- Jakiw Pidstrigach, Elizabeth Baker, Carles Domingo-Enrich, George Deligiannidis, and Nikolas Nüsken.
  Conditioning Diffusions Using Malliavin Calculus.
  arXiv:2504.03461, 2025.
  `https://arxiv.org/abs/2504.03461`
- Samuel Howard, Nikolas Nüsken, and Jakiw Pidstrigach.
  Control Consistency Losses for Diffusion Bridges.
  arXiv:2512.05070, 2025, submitted to ICML
  `https://arxiv.org/abs/2512.05070`
- Andreas Bergmeister, Jakiw Pidstrigach, Nik Nüsken, Stefanie Jegelka, Carles Domingo-Enrich
  Reinforce Adjoint Matching: Fine-tuning Diffusion and Flow Matching Models without Reward Gradients.
  Submitted to ICML.