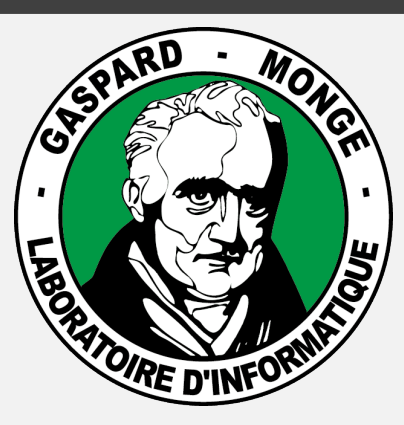




The Wasserstein gradient flow of the Sinkhorn divergence

Gaussian analysis and applications in generative modeling

Entropic optimal transport and the Sinkhorn divergence

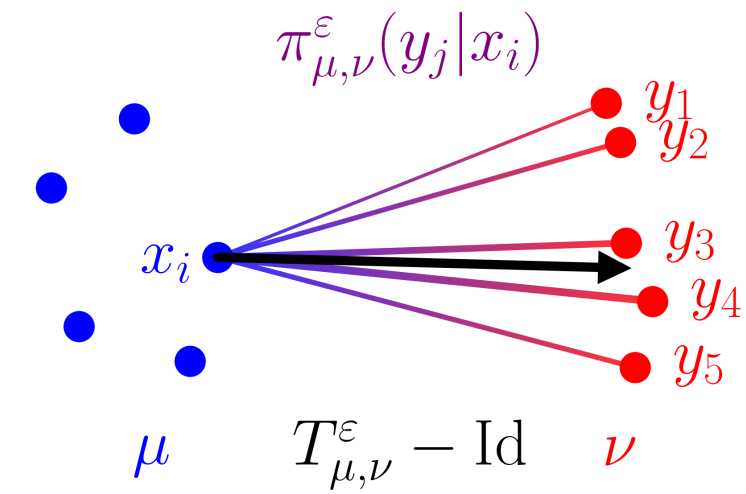
Entropic Optimal Transport (with $c(x, y) := \|x - y\|^2$):

$$\text{OT}_\varepsilon(\mu, \nu) := \min_{\pi \in \Pi(\mu, \nu)} \langle \pi, c \rangle + \varepsilon \text{KL}(\pi | \mu \otimes \nu)$$

$$= \max_{f, g} \langle \mu, f \rangle + \langle \nu, g \rangle - \varepsilon \langle \mu \otimes \nu, \exp(\frac{1}{\varepsilon}(f \oplus g - c)) \rangle + \varepsilon$$

$\pi_{\mu, \nu}^\varepsilon$:= minimizer of the primal, $(f_{\mu, \nu}^\varepsilon, g_{\mu, \nu}^\varepsilon)$:= maximizers of the dual.

$T_{\mu, \nu}^\varepsilon := \text{Id} - \nabla f_{\mu, \nu}^\varepsilon = \int y d\pi_{\mu, \nu}^\varepsilon(y | \cdot)$ is the *barycentric map*.



⚠ $\text{OT}_\varepsilon(\mu, \mu) \neq 0$ and $\arg \min_\nu \text{OT}_\varepsilon(\mu, \nu) \neq \mu$!

→ **Sinkhorn divergence:**

$$S_\varepsilon(\mu, \nu) := \text{OT}_\varepsilon(\mu, \nu) - \frac{1}{2}\text{OT}_\varepsilon(\mu, \mu) - \frac{1}{2}\text{OT}_\varepsilon(\nu, \nu)$$

Positive-definite and smooth loss [4], with a geometric flavor, good computational and statistical properties.

Wasserstein gradient flows (WGFs) and that of S_ε

WGF of a functional F :

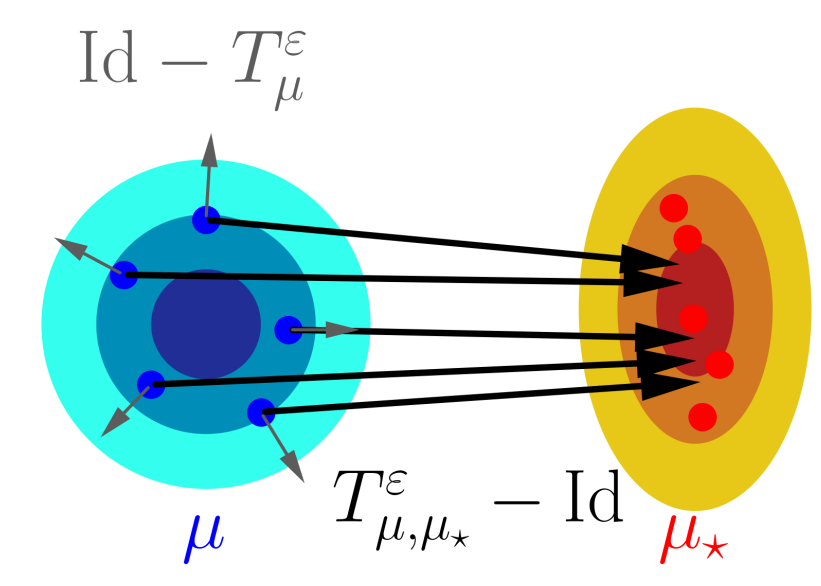
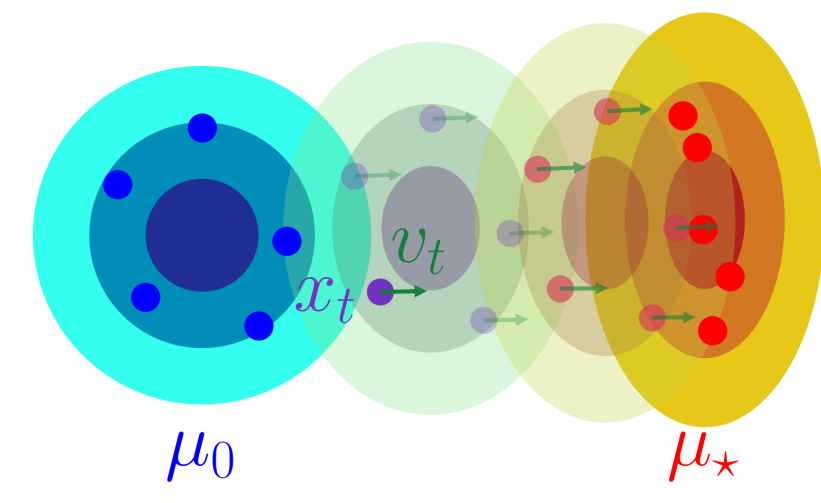
$$\dot{\mu}_t + \text{div}(\mu_t v_t) = 0,$$

$$v_t := -\nabla \frac{\delta F}{\delta \mu_t}$$

For $F = S_\varepsilon(\cdot, \mu_\star)$:

$$-\nabla \frac{\delta F}{\delta \mu} = -\nabla(f_{\mu, \mu_\star}^\varepsilon - f_\mu^\varepsilon) = (T_{\mu, \mu_\star}^\varepsilon - \text{Id}) + (\text{Id} - T_\mu^\varepsilon)$$

$T_{\mu, \mu_\star}^\varepsilon - \text{Id}$ drives the flow towards a shrunk $\mu_\star$, while $\text{Id} - T_\mu^\varepsilon$ pushes the mass away from its barycenter.



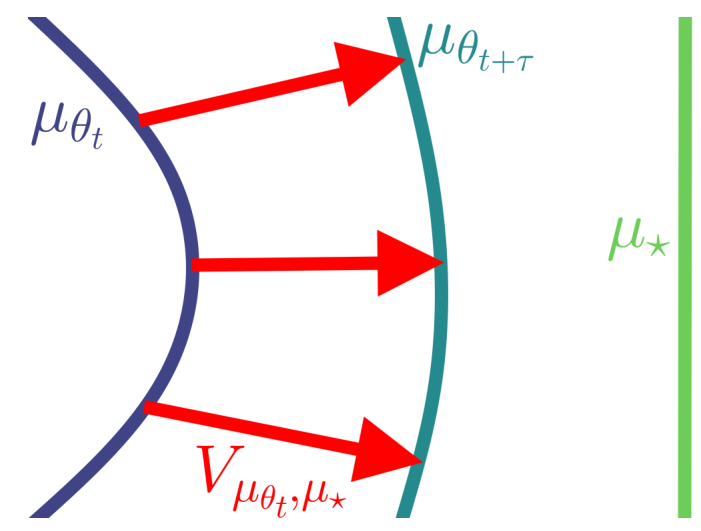
v_t is smooth → can't split mass !

Use in generative modeling

Neural Sinkhorn Flow [8]: Learn the velocity of the WGF using a NN, push samples along the learnt flow at inference.

Drifting models [3]: $\mu_\theta := (\varphi_\theta)_\# \mu^0$, φ_θ a NN.

At training: “drift” towards $\mu_\star$ along a vector field $V_{\mu_\theta, \mu_\star}$.
One-step at inference.



The articles [6, 5] use $V_{\mu, \mu_\star} = -\nabla(f_{\mu, \mu_\star}^\varepsilon - f_\mu^\varepsilon)$, effectively approximating the WGF of $S_\varepsilon(\cdot, \mu_\star)$.

Question (initially raised in [2])

Does the WGF of $S_\varepsilon(\cdot, \mu_\star)$ converge to the target measure $\mu_\star$ as $t \rightarrow \infty$?

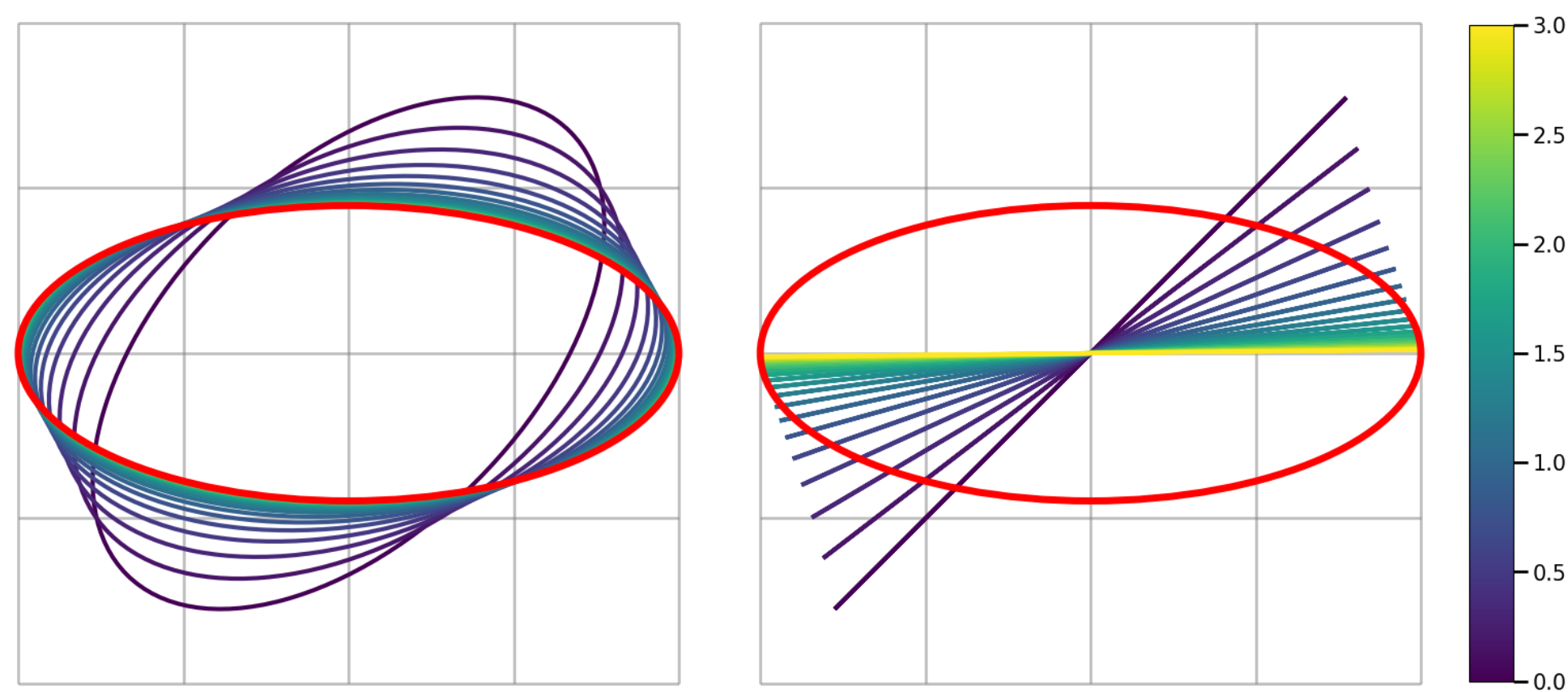
→ We give a **first answer** for Gaussians.

Theorem (H & Lacombe, 2026)

Let $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$, $\mu_\star = \mathcal{N}(m_\star, \Sigma_\star)$ (with $\Sigma, \Sigma_\star \succeq 0$). There exists a curve $t \mapsto \mu_t = \mathcal{N}(m_t, \Sigma_t)$ solution to the PDE $\dot{\mu}_t + \text{div}(\mu_t \nabla(f_{\mu_t} - f_{\mu_\star})) = 0$. It is unique in the class of measures $\mathcal{R} = \{\exp(-V), \alpha_V I \preceq \nabla^2 V \preceq \beta_V I\}$. We always have $m_t \xrightarrow[t \rightarrow \infty]{} m_\star$, and moreover:

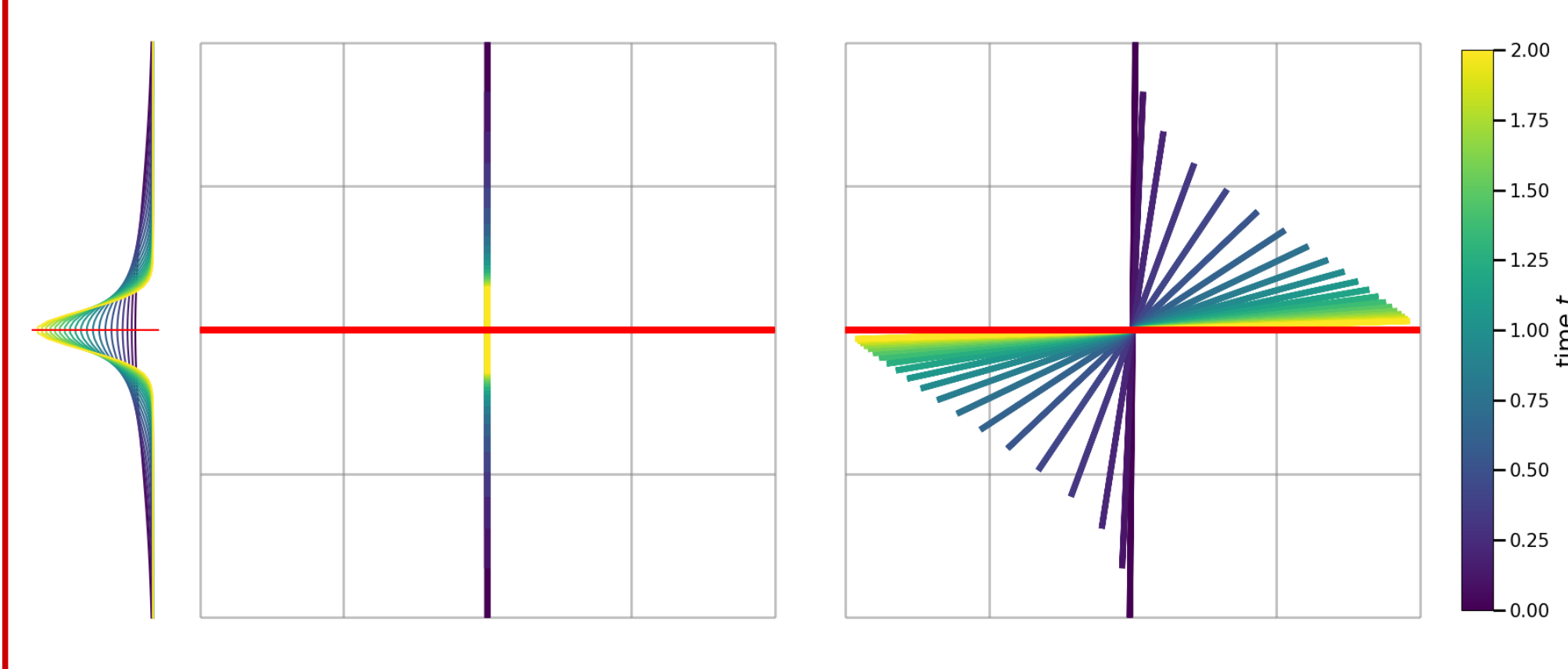
- 1 If μ_0 is non-singular, $\mu_t \xrightarrow[t \rightarrow \infty]{} \mu_\star$.
If μ_0 is singular, then $\forall t, \mu_t$ also singular.
- 2 If $\Sigma_\star = P \begin{pmatrix} \lambda_1^\star & (0) \\ (0) & \lambda_d^\star \end{pmatrix} P^T$, $\Sigma_t \xrightarrow[t \rightarrow \infty]{} P \begin{pmatrix} \lambda_1 & (0) \\ (0) & \lambda_d \end{pmatrix} P^T$, $\lambda_i \in \{0, \lambda_i^\star\}$.
- 3 If $\Sigma_0 \Sigma_\star = \Sigma_\star \Sigma_0$, $\mu_t \xrightarrow[t \rightarrow \infty]{} \mu_\star \iff \text{supp}(\mu_\star) \subseteq \text{supp}(\mu_0)$, at rate $O(e^{-Ct})$ if equality and $O(1/t)$ otherwise.

1 | 2



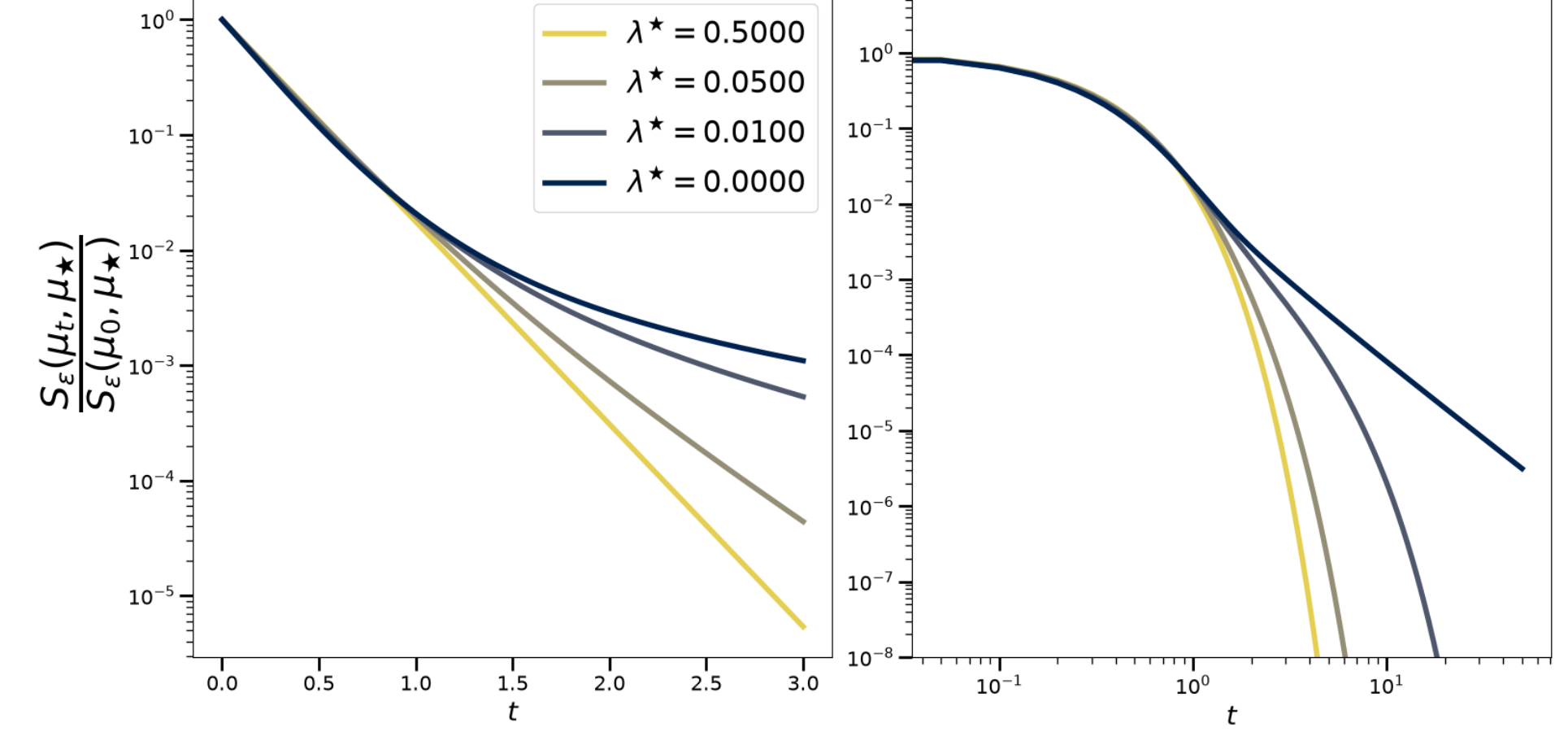
Covariance ellipses along the flow (targets in red, grid spacing $\sqrt{\varepsilon}$).

2 | 3



For singular Gaussians, convergence can fail (left) due to the impossibility to split mass. It makes the flow unstable w.r.t. μ_0 (small perturbation on the right).

3

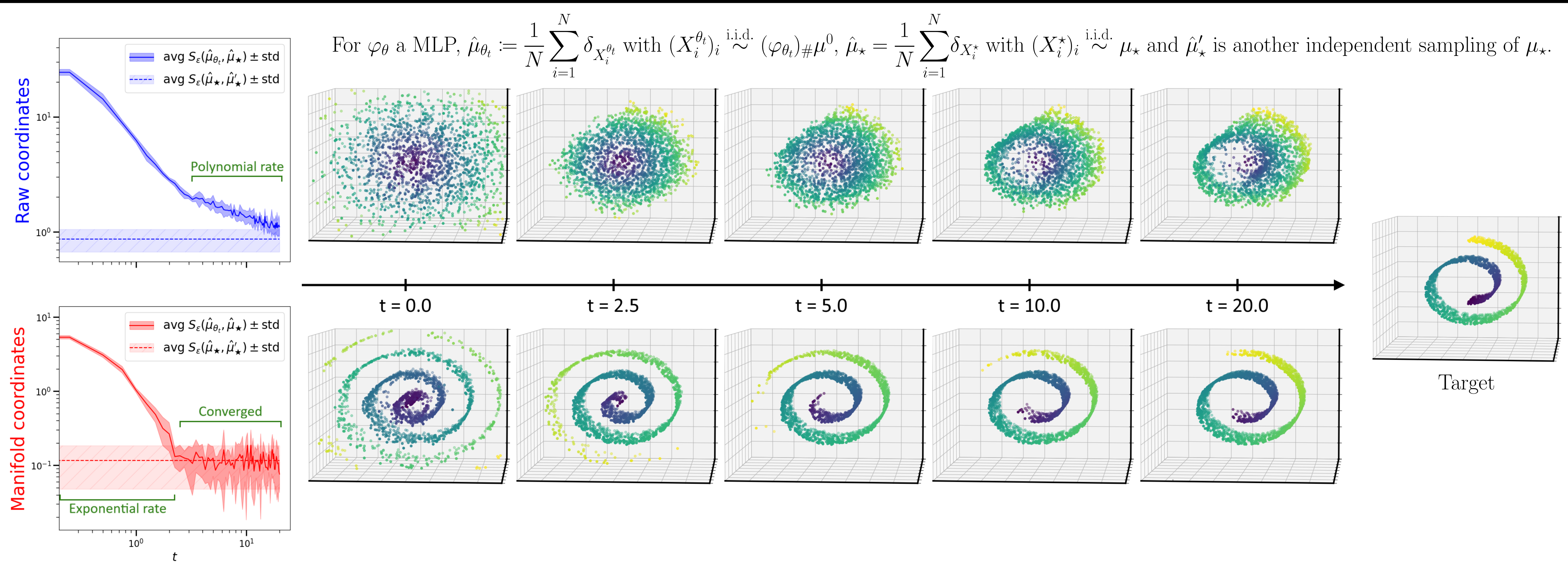


Convergence rates for $\Sigma_\star = \text{diag}((0, \lambda_\star))$ with $\lambda_\star \downarrow 0$ when $\Sigma_0 \succ 0$ and $\Sigma_0 \Sigma_\star \neq \Sigma_\star \Sigma_0$, semi-log vs. log-log.

Key proof elements

- $S_\varepsilon(\cdot, \mu_\star)$ restricted to Gaussians is $-4\lambda_{\max}(\Sigma_\star)/\varepsilon$ -convex along generalized geodesics + apply [1].
- EVI on $\mathcal{R}$.
- Extensive use of the closed-form expressions [7].

Comparison with non-Gaussian data: learning the swiss roll with a Sinkhorn-drifting model in raw vs manifold coordinates



Main takeaway

Convergence is slowed down by the need to collapse mass, e.g. when mapping a full support distribution to a manifold-supported distribution → use latent spaces !

References

- [1] L. Ambrosio et al. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Lectures in Math. ETH Zürich. Birkhäuser Basel, 2nd edition, 2008.
- [2] G. Carlier et al. Displacement Smoothness of Entropic Optimal Transport. *ESAIM: COCV*, 2024.
- [3] M. Deng et al. Generative Modeling via Drifting. 2026. arXiv: 2602.04770 [cs.LG].
- [4] J. Feydy et al. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. In *AISTATS*, 2019.
- [5] J. Han et al. One-Step Generative Modeling via Wasserstein Gradient Flows. 2026. arXiv: 2605.11755 [cs.LG].
- [6] P. He et al. Sinkhorn-Drifting Generative Models. 2026. arXiv: 2603.12366 [cs].
- [7] H. Janati et al. Entropic Optimal Transport between Unbalanced Gaussian Measures Has a Closed Form. In *NeurIPS*, 2020.
- [8] H. Zhu et al. Neural Sinkhorn Gradient Flow. 2024. arXiv: 2401.14069 [cs].

