UNIVERSITY OF CAMBRIDGE

UCLA

UNIVERSITY OF South Carolina

Deterministic Sampling with Wasserstein Proximals

Hong Ye Tan
Joint work with: Stanley Osher (UCLA), Wuchen Li (USC)
Department of Applied Mathematics and Theoretical Physics, University of Cambridge

1. Wasserstein Proximals

- **Goal:** Particle-based sampling without added noise

$$\underbrace{\frac{\partial \rho}{\partial t} = \nabla\cdot(\nabla V(x)\rho) + \beta\Delta\rho}_{\text{Fokker-Planck Equation}}, \quad \rho(x,0)=\rho_0(x) \implies \underbrace{\rho(x) \sim \exp(-V(x)/\beta)}_{\text{Stationary Solution}}$$

- Approximate version of the Jordan-Kinderlehrer-Otto scheme

$$\rho_{k+1} = \arg\min_{\rho\in\mathcal{P}_2} \underbrace{\int_{\mathbb{R}^d} (\beta\rho\log\rho + V\rho)dx}_{\text{Free energy functional}} + \frac{1}{2h}\mathcal{W}(\rho_k,\rho)^2,$$

- **Key**: Implicit computation of the Wasserstein proximal using backwards Euler scheme
- Computational machinery: computable approximation of the Wasserstein proximal, Monte Carlo integration, clever ODE discretization

2. Approximating the Proximal

$dX(t) = -\nabla V(X(t))dt + \sqrt{2\beta}dW(t), \quad X(0)=X_0$

Fokker-Planck SDE ⟺ (Liouville's Equation) **Score-based ODE**

(1) ≀ ↓ Motivated by kernel formulation

$\frac{dX}{dt} = -\nabla V(X) - \beta\nabla\log\rho(t,X)$ ≀ $\frac{dX}{dt} = \nabla\Phi(t,X) - \beta\nabla\log\rho(t,X)$

Regularized FP-SDE ⟺ **Regularized score-based ODE**

Solving for one time step:

Proposed Method:

(2) **Backward Euler discretization** — $\frac{dX}{dt} = \nabla\Phi(t,X) - \beta\nabla\log\rho(t,X)$

$$X_{k+1} = X_k + \eta[\nabla\Phi(T,X_k) - \beta\nabla\log\rho_{k,0}(T,X_k)]$$
$$= X_k - \eta\nabla V(X_k) - \eta\beta\nabla\log\rho_{k,T}(X_k).$$

Remove dependence on dual variable Φ → **Compute regularized proximal with kernel formula** → (4) **Evolve particles**

Regularized Benamou-Brenier
$$\begin{cases} \partial_t\rho(t,x) + \nabla_x\cdot(\rho(t,x)\nabla_x\Phi(t,x)) = \beta\Delta_x\rho(t,x), \\ \partial_t\Phi(t,x) + \frac{1}{2}\|\nabla_x\Phi(t,x)\|^2 = -\beta\Delta_x\Phi(t,x), \\ \rho(0,x)=\rho_0(x), \quad \Phi(T,x) = -V(x). \end{cases} \Rightarrow \rho(T,x) = \int_{\mathbb{R}^d} K(x,y)\rho_0(y)\,dy,$$
$$K(x,y) = \frac{\exp(-\frac{1}{2\beta}(V(x)+\frac{\|x-y\|^2}{2T}))}{\int_{\mathbb{R}^d}\exp(-\frac{1}{2\beta}(V(z)+\frac{\|z-y\|^2}{2T}))\,dz}. \quad (3)$$

3. Convergence

Theorem. Applied to the d-dimensional Ornstein-Uhlenbeck process with condition number κ, the worst-case (TV)-mixing time is

$$t_{\text{mix}}(\delta) = \mathcal{O}\left(\kappa^{3/2}\log\left(\kappa\sqrt{d}\right)/\delta\right).$$

Moreover, the covariance has a closed form, and the inverse covariance matrix converges linearly to the (biased) stationary distribution.

(For quadratic V.)

Comparisons: **ULA:** $\mathcal{O}((d^3 + d\log^2(1/\delta))\kappa^2\delta^{-2})$

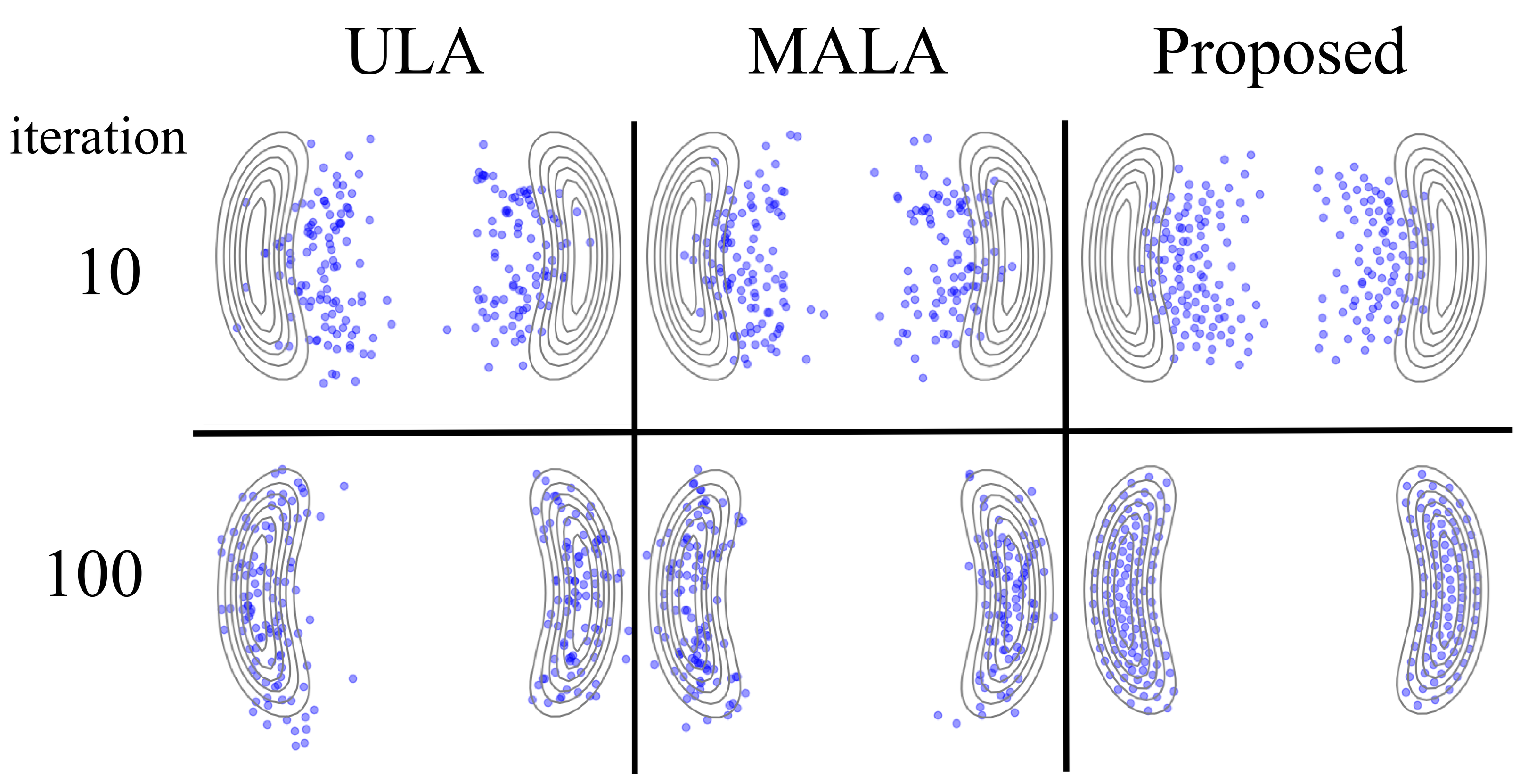
MALA: $\mathcal{O}(d^2\kappa\log(\kappa/\delta))$

Better dependence on problem dimension d!

(Implicitly hidden in the Monte-Carlo step)

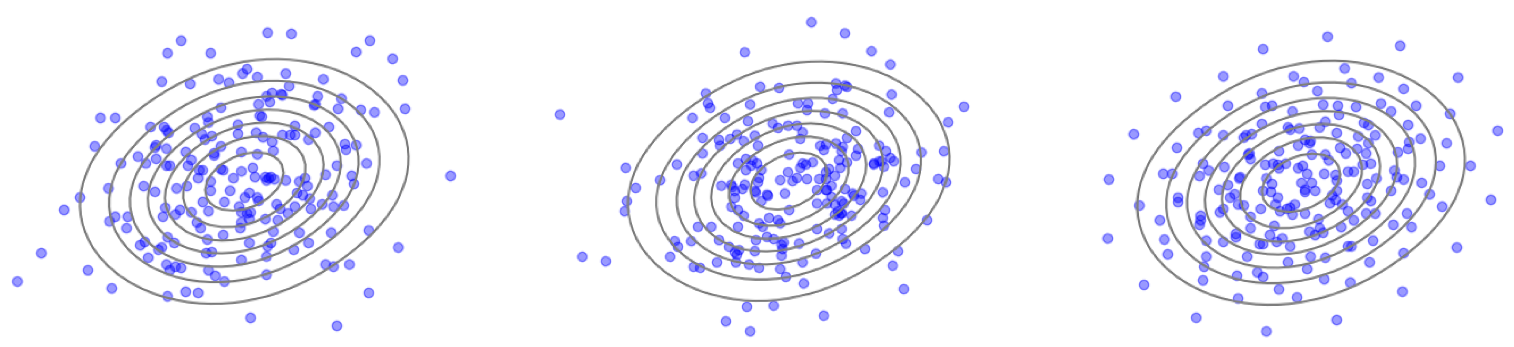
Coming soon for more general V

4. Experiments



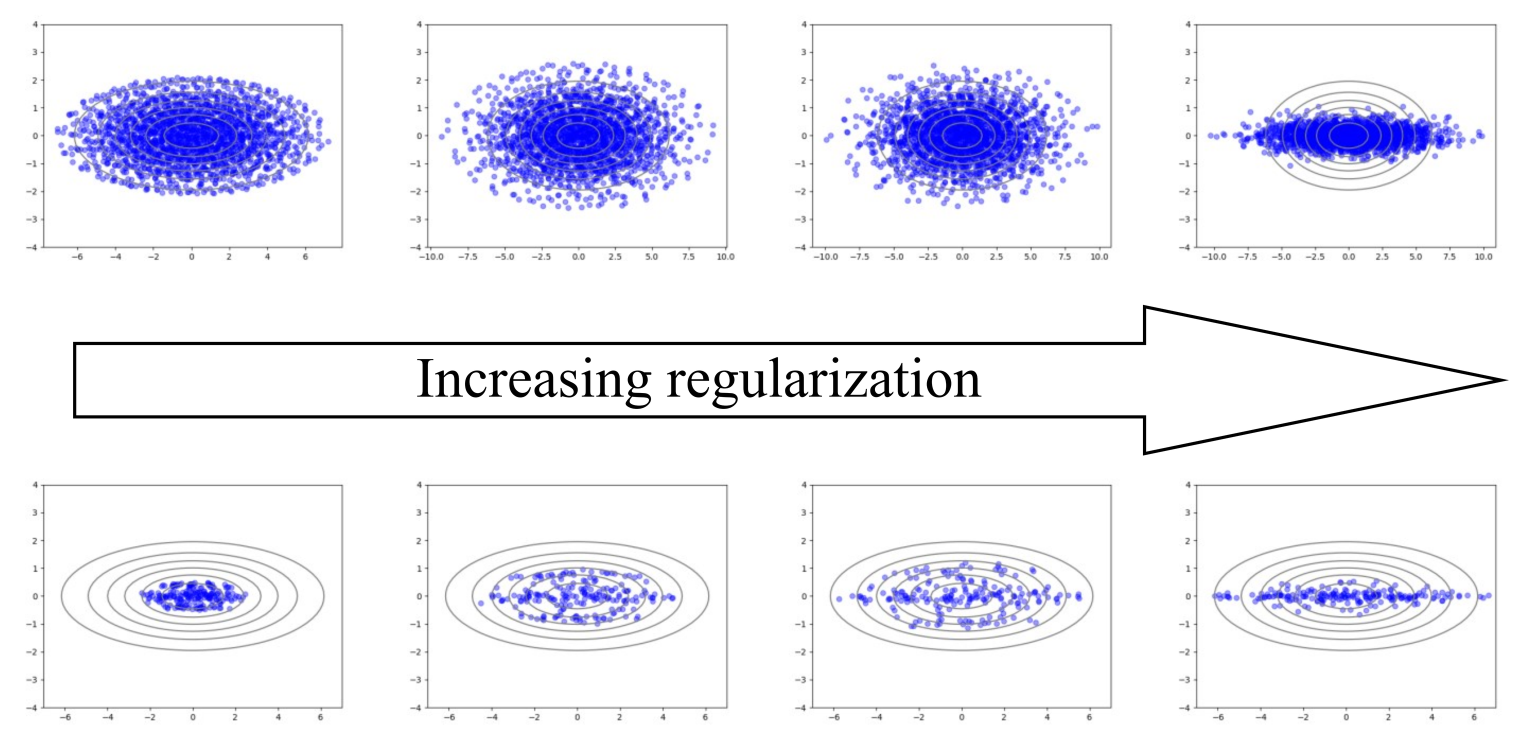
2D bi-modal double banana distribution

Interesting structure!



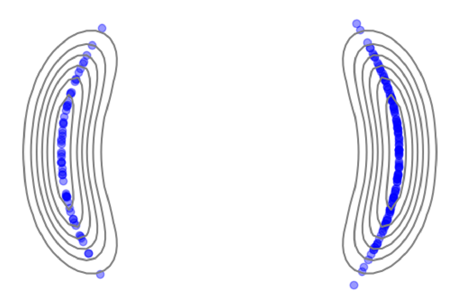
2D Gaussian mixture

2D Gaussian



10D Gaussian

Variance reduction from regularization



Regularization allows for larger step-size

Key Approximations

1. Approximating the potential in the FP-SDE with the Kantorovich dual variable
2. Discrete time approximation of ODE
3. Monte-Carlo computation in kernel formula
4. Empirical distribution approximation

References

[1] HYT, S. Osher, and W. Li. "Noise-Free Sampling Algorithms via Regularized Wasserstein Proximals". arXiv Oct 2023.

[2] W. Li, S. Liu, and S. Osher. "A kernel formula for regularized Wasserstein proximal operators". RMS 2023.

[3] F. Han, S. Osher, and W. Li. "Tensor train based sampling algorithms for approximating regularized Wasserstein proximal operators". arXiv Jan 2024.
(Followup work that resolves the bias issues and allows for faster computation.)

Acknowledgements: Work supported by NoMADS program and AFOSR YIP award No. FA9550-23-1-0087.
Workshop on Optimal Transport from Theory to Applications, Berlin, March 11-15th 2024