# The Stein-log-Sobolev inequality and the exponential convergence of the Stein Variational Gradient Descent algorithm

## From non-local to local

**Jethro Warnett**,
Jose A. Carrillo, Jakub Skrzeczkowski

*Mathematical Institute, University of Oxford*

Oxford Mathematics

UNIVERSITY OF OXFORD
Mathematical Institute

# Table of Contents

Mathematical Institute

# The mobility equation



The following PDE denotes a general mobility equation

$$\partial_t \rho = \mathrm{div}(m(\rho)\mathcal{V}(\rho))$$

where

- $m : \mathbb{R} \to \mathbb{R}$ is some continuous mobility function,
- $\mathcal{V}(\rho)$ is some velocity that can have the form

$$\mathcal{V}(\rho_t) = \nabla U'(\rho_t) + \nabla V + \nabla(W * \rho_t).$$

# Examples [Carrillo et al., 2024b]

Mathematical Institute

- Lubrication model for thin films [Bertozzi, 1998]

$$m(\rho) = \rho^3, \quad \mathcal{V}(\rho) = -\sqrt{\mathrm{Ca}}\,\nabla\rho + \nabla\Delta\rho.$$

- Cahn-Hilliard equation for phase separation in binary alloys [Cahn, 1998]

$$m(\rho) = 1 - \rho^2, \quad \mathcal{V}(\rho) = -\nabla\frac{\delta E}{\delta\rho}, \quad E(\rho) := \int_\Omega H(\rho(x)) + \frac{\epsilon^2}{2}|\nabla\rho(x)|^2 \, \mathrm{d}x,$$

where $\rho : [0, \infty) \times \Omega \to [-1, 1]$.

- Chemotaxis with prevention of overcrowding [Burger et al., 2006]

$$m(\rho) = \rho(1 - \rho).$$

# Table of Contents

# Stein Variational Gradient Descent [Liu and Wang, 2016]



We want to approximate $\rho_\infty \propto e^{-V}$ for a potential $V : \mathbb{R}^d \to \mathbb{R}$.

SVGD is governed by a positive definite interaction kernel $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

Given is some initial distribution of particles $X_0^1, \ldots, X_0^N \in \mathbb{R}^d$.

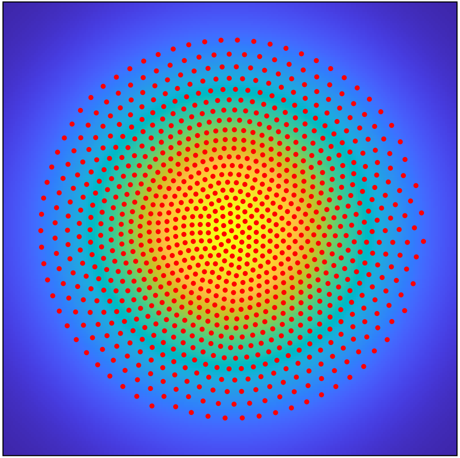
SVGD has deterministic particle dynamics

$$X_{n+1}^i := X_n^i + \varepsilon_n \sum_{j=1}^N (\nabla_y K)(X_n^i,\, X_n^j) - K(X_n^i,\, X_n^j)\nabla V(X_n^j).$$

# Computational Examples



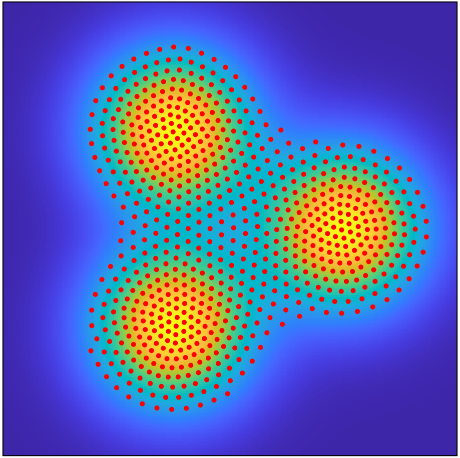
## Normal Gaussian



# Computational Examples

Mathematical Institute

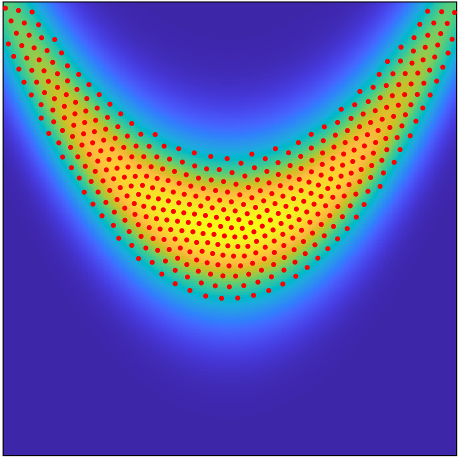
## Mixture of Gaussians



# Computational Examples



## Banana Distribution



# Mean-field limit



Formal computations in [Liu, 2017] show that if $N \to \infty$, $\varepsilon \to 0$ and the kernel is translation invariant $K(x, y) := k(x - y)$, then the following evolution of distribution of particles holds true

$$\partial_t \rho = \mathrm{div}\left(\rho\ k * (\nabla \rho + \rho \nabla V)\right) = \mathrm{div}(\rho\ k * \nabla(\rho e^V) e^{-V}).$$

# SVGD as a gradient flow

In fact we can interpret the previous equation as a gradient flow

$$\frac{d}{dt}\mathrm{KL}(\rho_t \,||\, \rho_\infty) = -\mathbb{D}^2(\rho_t \,||\, \rho_\infty),$$

where we define the dissipation by

$$\mathbb{D}^2(\rho_t \,||\, \rho_\infty) := \int_{\mathbb{R}^d} \nabla(\rho e^V)e^{-V} \cdot k * \nabla(\rho e^V)e^{-V} \,\mathrm{d}x,$$

and the Kullback-Leibler divergence (or relative entropy)

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) := \int_{\mathbb{R}^d} \rho \ln(\rho) \,\mathrm{d}x + \int_{\mathbb{R}^d} \rho V \,\mathrm{d}x.$$

# Stein-log-Sobolev inequality



We achieve an exponential rate of decay in the Kullback-leibler divergence if the **Stein-log-Sobolev inequality** holds for some fixed $\lambda > 0$

$$\lambda \, \mathrm{KL}(\rho_t \,||\, \rho_\infty) \leq \mathbb{D}^2(\rho_t \,||\, \rho_\infty).$$

In this case we get

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) \leq e^{-\lambda t} \mathrm{KL}(\rho_0 \,||\, \rho_\infty).$$

# Table of Contents

Mathematical Institute

# From non-local to local

Take $k \in L^1(\mathbb{R}^d)$ with $k \geq 0$, $\int_{\mathbb{R}^d} k \,\mathrm{d}x = 1$ and $k$ is positive definite. Define the approximation of the identity $k_\sigma := \sigma^{-d} k(x/\sigma)$ for $\sigma \in (0,1)$. Let $\rho^\sigma$ denote the solution of

$$\partial_t \rho^\sigma = \operatorname{div}\left(\rho^\sigma \; k_\sigma * (\nabla \rho^\sigma + \rho^\sigma \nabla V)\right).$$

We observe that $k_\sigma \rightharpoonup \delta_0$ as $\sigma \to 0$ in the sense of distributions. Then, if $\rho^\sigma \rightharpoonup \rho$ as $\sigma \to 0$, then one formally derives a weak solution to the following mobility equation

$$\partial_t \rho = \operatorname{div}\left(\rho \left(\nabla \rho + \rho \nabla V\right)\right) = \operatorname{div}\left(\rho^2 \; \nabla(\ln(\rho) + V)\right).$$

# Exponential convergence rate for the continuity equation

Assume that the Stein-log-Sobolev inequality holds uniformly for all solutions $\{\rho^\sigma\}_\sigma$ with some fixed constant $\lambda > 0$. Then, we derive the uniform bound

$$\mathrm{KL}(\rho_t^\sigma \,||\, \rho_\infty) \leq e^{-\lambda t}\mathrm{KL}(\rho_0 \,||\, \rho_\infty).$$

By taking the limit $\sigma \to 0$ and using the lower semicontinuity of the Kullback-Leibler divergence, we see that the continuity equation would possess an exponential convergence rate

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) \leq e^{-\lambda t}\mathrm{KL}(\rho_0 \,||\, \rho_\infty).$$

# Table of Contents

# Main Theorem [Carrillo et al., 2024a]

We were inspired by [Duncan et al., 2023] to use the following ansatz form

$$K(x,y) = e^{V(x)-\frac{V_0(x)}{2}}k(x-y)e^{V(y)-\frac{V_0(y)}{2}}.$$

A sufficient criteria for the Stein-log-Sobolev inequality is

- $V \geq \mathbb{V}_0$ for $\mathbb{V}_0(x) = \alpha|x|^2 + \beta$,
- $k \in L^1(\mathbb{R}^d) + L^2(\mathbb{R}^d)$,
- There exists two constants $C_0, C_1 \geq 1$ such that

$$\frac{1}{C_0}\frac{1}{1+|\xi|^2} \leq \hat{k}(\xi) \leq C_1\frac{1}{1+|\xi|^2}.$$

# Existence of weak solution [Carrillo et al., 2024a]

Mathematical Institute

If additionally $V \in C^1(\mathbb{R}^d) \cap H^m_{\mathrm{loc}}(\mathbb{R}^d)$ for some $m > \frac{d}{2}$, then there exists a solution $\rho \in C^0([0,\infty), \mathcal{P}(\mathbb{R}^d))$ in the distributional sense to the non-local continuity equation

$$\partial_t \rho_t = \mathrm{div}\left(\rho_t e^{V - \frac{V_0}{2}} \; k * \left[(\nabla \rho_t + \rho_t \nabla V) e^{V - \frac{V_0}{2}}\right]\right). \tag{$*$}$$

Moreover, we have exponential decay in Kullback-Leibler divergence and the energy dissipation inequality

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) \leq e^{-\lambda t} \mathrm{KL}(\rho_0 \,||\, \rho_\infty),$$

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) + \int_0^t \mathbb{D}^2(\rho_s \,||\, \rho_\infty)\, \mathrm{d}s \leq \mathrm{KL}(\rho_0 \,||\, \rho_\infty).$$

# Stein-log-Sobolev constant [Carrillo et al., 2024a]

The Stein-log-Sobolev constant has the form

$$\lambda = \lambda_0 \, (\alpha \wedge 1) \, e^{\beta} \, \frac{1}{C_0^2} \, \frac{1}{\|e^{-V}\|_{L^1}},$$

where $\lambda_0 > 0$ is some constant independent of dimension $d$, and we recall that $\alpha$ and $\beta$ determine $\mathbb{V}_0$, and $C_0$ is used for the lower bound of $\hat{k}$.

If we only vary $k$, then we see that the Stein-log-Sobolev constant only depends on $C_0$.

# Application of result

Mathematical Institute

Let $\rho^\sigma$ be the weak solution to $(*)$ with kernel $k_\sigma(x) := \sigma^{-d} k(x/\sigma)$.

We observe a uniform lower bound

$$\hat{k}_\sigma(x) = \hat{k}(\sigma x) \geq \frac{1}{C_0} \frac{1}{1 + |\sigma \xi|^2} \geq \frac{1}{C_0} \frac{1}{1 + |\xi|^2}.$$

Thus we see that the Stein-log-Sobolev inequality holds uniformly for all $\{k_\sigma\}_\sigma$.

# Table of Contents

# Passing to the limit

Let $\{\rho^\sigma\}_\sigma$ be the weak solutions to $(*)$.

If we assume $k = \omega * \omega$ for some measurable function $\omega : \mathbb{R}^d \to \mathbb{R}$ that satisfies certain constraints, and $\omega_\sigma(x) = \sigma^{-d}\omega(x/\sigma)$, then we need to pass to the limit in the following PDE

$$\partial_t \rho_t^\sigma = \mathrm{div}\left( \left(\rho_t^\sigma e^{V - \frac{V_0}{2}}\right)\ \omega_\sigma * \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V - \frac{V_0}{2}} \right] * \omega_\sigma \right).$$

# Passing to the limit

Oxford Mathematical Institute

$$\partial_t \rho_t^\sigma = \mathrm{div}\left( (\rho_t^\sigma e^{V - \frac{V_0}{2}}) \ \omega_\sigma * \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V - \frac{V_0}{2}} \right] * \omega_\sigma \right)$$

We restrict ourselves to the domain $[0, T] \times B_R$. A commutator argument shows that the above limit coincides with the one below

$$\partial_t \rho_t^\sigma = \mathrm{div}\left( (\rho_t^\sigma e^{V - \frac{V_0}{2}} \chi) * \omega_\sigma \ \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V - \frac{V_0}{2}} \right] * \omega_\sigma \right)$$

where $\chi \in C_c^\infty(\mathbb{R}^d)$ satisfies $0 \leq \chi \leq 1$, $\chi = 1$ on $B_R$ and $\chi = 0$ outside $B_{2R}$.

# Passing to the limit

$$\partial_t \rho_t^\sigma = \operatorname{div}\left( (\rho_t^\sigma e^{V - \frac{V_0}{2}} \chi) * \omega_\sigma \ \ \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V - \frac{V_0}{2}} \right] * \omega_\sigma \right)$$

The energy dissipation gives us weak compactness in $L^2([0,T] \times \mathbb{R}^d)$.

# Passing to the limit

Mathematical Institute

$$\partial_t \rho_t^\sigma = \mathrm{div}\left( (\rho_t^\sigma e^{V-\frac{V_0}{2}} \chi) * \omega_\sigma \quad \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V-\frac{V_0}{2}} \right] * \omega_\sigma \right)$$

Using the Aubin-Lions lemma we show compactness in $L^2([0,T] \times \mathbb{R}^d)$.

# Passing to the limit



$$\color{red}{\partial_t \rho_t^\sigma} = \operatorname{div}\left( \left(\rho_t^\sigma e^{V - \frac{V_0}{2}} \chi\right) * \omega_\sigma \quad \left[ (\nabla \rho_t^\sigma + \rho_t^\sigma \nabla V) e^{V - \frac{V_0}{2}} \right] * \omega_\sigma \right)$$

We show weak compactness in $L^{2\frac{q+2}{q+1}}(0, T; H^{-m}(B_R))$ to derive uniform convergence in $C^0([0, T]; \mathcal{P}(\mathbb{R}^d))$ for $q := d \vee 3$ and some fixed $m > \frac{d}{2}$.

Combining everything allows us to pass to the limit.

# Table of Contents

Mathematical Institute

# Summary

Mathematical Institute

We've shown the existence of a solution $\rho \in C^0([0,\infty);\mathcal{P}(\mathbb{R}^d))$ in the sense of distributions to the following mobility equation

$$\partial_t \rho = \mathrm{div}(\rho^2 \nabla(\ln(\rho) + V)).$$

Moreover, we have shown that this equation possess an exponential decay in the Kullback-Leibler divergence (or relative entropy), for some $\lambda > 0$ depending only on $V$,

$$\mathrm{KL}(\rho_t \,||\, \rho_\infty) \leq e^{-\lambda t}\mathrm{KL}(\rho_0 \,||\, \rho_\infty).$$

# Thank you for your attention

Bertozzi, A. (1998).
The mathematics of moving contact lines in thin liquid films.
*Notices of the American Mathematical Society*, 45:689–697.

Burger, M., Di Francesco, M., and Dolak-Struss, Y. (2006).
The keller–segel model for chemotaxis with prevention of overcrowding: Linear vs. nonlinear diffusion.
*SIAM Journal on Mathematical Analysis*, 38(4):1288–1315.

Cahn, J. W. (1998).
On spinodal decomposition.

Carrillo, J. A., Skrzeczkowski, J., and Warnett, J. (2024a).
The stein-log-sobolev inequality and the exponential rate of convergence for the continuous stein variational gradient descent method.

Carrillo, J. A., Wang, L., and Wei, C. (2024b).

Structure preserving primal dual methods for gradient flows with nonlinear mobility transport distances.
*SIAM Journal on Numerical Analysis*, 62(1):376–399.

Duncan, A., Nüsken, N., and Szpruch, L. (2023).
On the geometry of Stein variational gradient descent.
*J. Mach. Learn. Res.*, 24:Paper No. [56], 39.

Liu, Q. (2017).
Stein variational gradient descent as gradient flow.
*Advances in Neural Information Processing Systems*, 30.

Liu, Q. and Wang, D. (2016).
Stein variational gradient descent: a general purpose Bayesian inference algorithm.
*Proc. 30th Int. Conf. Neural Inf. Proc. Syst.*, page 2378–2386.