Stein Variational Gradient Descent

Jethro Warnett,
Jose A. Carrillo, Jakub Skrzeczkowski

Mathematical Institute, University of Oxford

Goal: We want to sample from a probability distribution $\rho_\infty \in \mathcal{P}(\mathbb{R}^d)$.

We assume the distribution is defined by

$$\rho_\infty = \frac{1}{Z}\, e^{-V},$$

where the *potential* $V : \mathbb{R}^d \to \mathbb{R}$ is known, but the *normalization constant* $Z > 0$ is unknown.

Popular Sampling Algorithm

A popular sampling algorithm is the Langevin Dynamics.

You choose a random initial value X_0 and simulate the following SDE until a sampling time $T > 0$

$$\mathrm{d}X_t = \nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t,$$

where B_t is Brownian motion.

The law μ_t of X_t is governed by a Fokker-Planck equation

$$\partial_t \mu_t = \mathrm{div}(\mu_t\, \nabla[\log(\mu_t) + V]).$$

Challenge of Langevin Dynamics

The Log-Sobolev inequality states

$$\underbrace{\lambda \, \mathrm{KL}(\mu_t \,||\, \rho_\infty)}_{= \int_{\mathbb{R}^d} \mu_t \ln\left(\frac{\mu_t}{\rho_\infty}\right) \mathrm{d}x} \leq \int_{\mathbb{R}^d} \frac{|\nabla \mu_t + \mu_t \nabla V|^2}{\mu_t} \, \mathrm{d}x = -\frac{d}{dt} \mathrm{KL}(\mu_t \,||\, \rho_\infty)$$

If it holds, then Langevin Dynamics converges exponentially

$$\mathrm{KL}(\mu_t \,||\, \rho_\infty) \leq e^{-\lambda t} \mathrm{KL}(\mu_0 \,||\, \rho_\infty)$$

The Log-Sobolev inequality holds if V is strongly convex

$$\nabla^2 V \succcurlyeq \lambda \, \mathrm{id}$$

Stein Variational Gradient Descent [Liu and Wang, 2016]

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

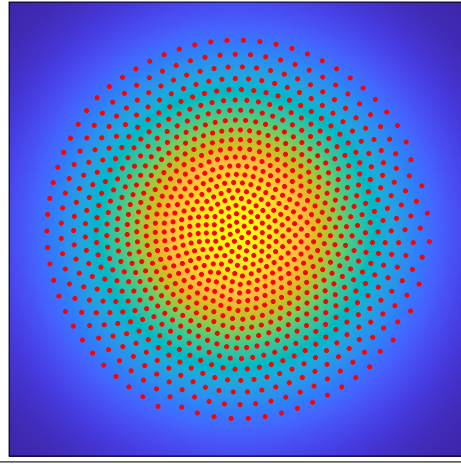
Given is some initial distribution of particles $X_0^1, \ldots, X_0^N \in \mathbb{R}^d$.

SVGD has deterministic particle dynamics

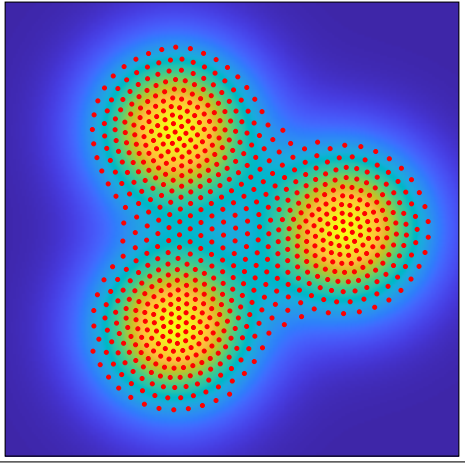
$$X_{n+1}^i := X_n^i + \varepsilon_n \sum_{j=1}^{N} (\nabla_y K)(X_n^i,\ X_n^j) - K(X_n^i,\ X_n^j) \nabla V(X_n^j).$$

Computational Examples

Normal Gaussian

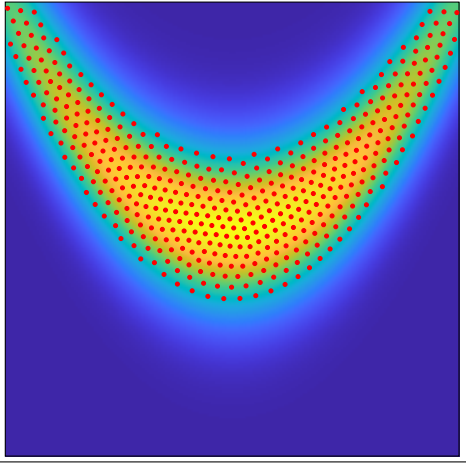


Mixture of Gaussians



Computational Examples

Banana Distribution



Mean-field limit

Formal computations in [Liu, 2017] show that if $N \to \infty$ and $\varepsilon \to 0$, then the following evolution of distribution of particles ρ_t holds true

$$\partial_t \rho(x) = \operatorname{div}\left(\rho(x) \int_{\mathbb{R}^d} K(x,y)(\nabla \rho(y) + \rho(y)\nabla V(y))\,\mathrm{d}y\right).$$

Comparison

Particle evolution for Langevin dynamics

$$\partial_t \mu_t = \operatorname{div}(\mu_t \nabla[\log(\mu_t) + V]).$$

Particle density evolution for SVGD

$$\partial_t \rho(x) = \operatorname{div}\left(\rho(x) \int_{\mathbb{R}^d} K(x, y) \nabla[\log(\rho(y)) + V(y)] \rho(y) \, \mathrm{d}y\right).$$

Stein-log-Sobolev inequality

The **Stein-log-Sobolev inequality** states

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

with dissipation

$$\mathbb{D}^2(\rho_t \,||\, \rho_\infty) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [\nabla \rho_t + \rho_t \nabla V](x) \; K(x,y) \; [\nabla \rho_t + \rho_t \nabla V](y) \, \mathrm{d}y \, \mathrm{d}x,$$

If it holds, then SVGD converges exponentially

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

Main Theorem [Carrillo et al., 2024]

Ansatz form [Duncan et al., 2023]

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

Sufficient criteria for SLSI

- $V \geq \mathbb{V}_0$ 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., 2024]

There exists a distributional solution $\rho \in C^0([0,\infty), \mathcal{P}(\mathbb{R}^d))$, whenever $V \in C^1(\mathbb{R}^d) \cap H^m_{\mathrm{loc}}(\mathbb{R}^d)$ for some $m > \frac{d}{2}$, for

$$\partial_t \rho_t = \operatorname{div}\left(\rho_t e^{V - \frac{\mathbb{V}_0}{2}} \; k * \left[(\nabla \rho_t + \rho_t \nabla V) e^{V - \frac{\mathbb{V}_0}{2}}\right]\right).$$

We have exponential decay and energy dissipation inequality

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

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

From non-local to local

By concentrating k to a Dirac measure, there exists a distributional solution $\varrho \in C^0([0,\infty), \mathcal{P}(\mathbb{R}^d))$ to

$$\partial_t \varrho_t = \operatorname{div}\left(\varrho_t^2 e^{2V-\mathbb{V}_0}\, \nabla(\ln(\varrho_t) + V)\right).$$

We have exponential decay and energy dissipation inequality

$$\mathrm{KL}(\varrho_t \,||\, \varrho_\infty) \leq e^{-\lambda t}\mathrm{KL}(\varrho_0 \,||\, \varrho_\infty),$$

$$\mathrm{KL}(\varrho_t \,||\, \varrho_\infty) + \int_0^t |\nabla \varrho_s + \varrho_s \nabla V|^2 e^{2V-\mathbb{V}_0}\, \mathrm{d}s \leq \mathrm{KL}(\varrho_0 \,||\, \varrho_\infty).$$

Outlook

- Derive and quantize the mean-field limit
- Extend theory to $V \geq \alpha|x|^q + \beta$ for $q \in (0, 2)$
- Implement numerically

Thank you for your attention

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

The Stein-log-Sobolev constant has the form

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

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

Exact Renormalization Groups & Wasserstein Gradient Flows

Polchinski's Equation & Geometric Flows

Connor Marrs

July 24, 2025

Joint work with Prof. Katy Craig

Review of Wasserstein Gradient Flows

Many diffusion PDE's (or equivalently diffusion processes) have a Wasserstein gradient flow structure.

Review of Wasserstein Gradient Flows

Many diffusion PDE's (or equivalently diffusion processes) have a Wasserstein gradient flow structure. The simplest examples describe systems diffusing in the presence or absence of some potential:

	Heat	Fokker Planck
SDE	$dX_t = \sqrt{2}dB_t$	$dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t$
PDE	$\partial_t \mu = \Delta \mu$	$\partial_t \mu = \mathrm{div}(\mu V)$
WGF	$\mathcal{E}(\rho) = \mathrm{Ent}(\rho) = \int \rho \log \rho$	$\mathcal{F}(\rho) = \int \rho \log \rho + \int V d\mu$

Review of Wasserstein Gradient Flows

Many diffusion PDE's (or equivalently diffusion processes) have a Wasserstein gradient flow structure. The simplest examples describe systems diffusing in the presence or absence of some potential:

	Heat	Fokker Planck
SDE	$dX_t = \sqrt{2}dB_t$	$dX_t = -\nabla V(X_t)dt + \sqrt{2}dB_t$
PDE	$\partial_t \mu = \Delta \mu$	$\partial_t \mu = \mathrm{div}(\mu V)$
WGF	$\mathcal{E}(\rho) = \mathrm{Ent}(\rho) = \int \rho \log \rho$	$\mathcal{F}(\rho) = \int \rho \log \rho + \int V d\mu$

Upshot:

- We can construct solutions via the JKO scheme
- We can establish properties of long time behavior for "nice functionals" (e.g. V λ-convex).

Exact Renormalization Groups

Let's consider a statistical model of a fluid, and suppose we model it with

- a lattice $X = (\varepsilon\mathbb{Z}^d)/(L\mathbb{Z}^d)$ (discretized space)
- "states" of our model, $\rho \in \mathbb{R}^X$ (function/density on X)

The states look qualitatively different at different scales.

Exact Renormalization Groups

Let's consider a statistical model of a fluid, and suppose we model it with

- a lattice $X = (\varepsilon\mathbb{Z}^d)/(L\mathbb{Z}^d)$ (discretized space)
- "states" of our model, $\rho \in \mathbb{R}^X$ (function/density on X)

The states look qualitatively different at different scales.

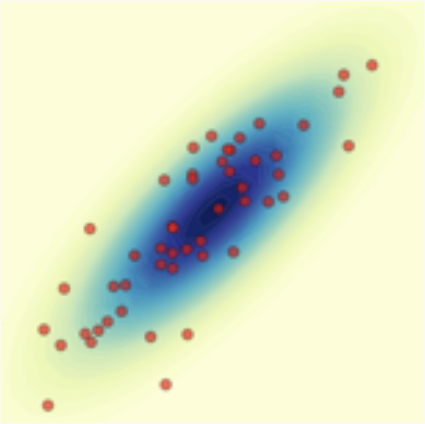


Figure 1: Intuition for small scale vs large scale states of a system of particles described by a Gaussian at large scale. Credit to Amir Masoud Sefidian

Exact Renormalization Groups

Let's consider a statistical model of a fluid, and suppose we model it with

- a lattice $X = (\varepsilon\mathbb{Z}^d)/(L\mathbb{Z}^d)$ (discretized space)
- "states" of our model, $\rho \in \mathbb{R}^X$ (function/density on X)

The states look qualitatively different at different scales.

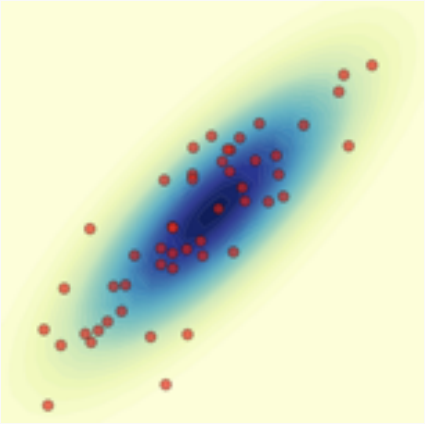


Figure 1: Intuition for small scale vs large scale states of a system of particles described by a Gaussian at large scale. Credit to Amir Masoud Sefidian

1. On the scale of molecules ($\varepsilon \approx 0$), states have largre spikes.
2. At larger scales ($\varepsilon >> 0$) states are averaged $\Rightarrow$ smoother.

Exact Renormalization Groups

Key observation: small fluctuations in density that we'd observe at very small scales will be qualitatively irrelevant at larger scales.

Exact Renormalization Groups

Key observation: small fluctuations in density that we'd observe at very small scales will be qualitatively irrelevant at larger scales.

Let's study the distribution P_Λ of all of possiblestates at a given scale Λ^{-1}.

Exact Renormalization Groups

Key observation: small fluctuations in density that we'd observe at very small scales will be qualitatively irrelevant at larger scales.

Let's study the distribution P_Λ of all of possiblestates at a given scale Λ^{-1}.

The ERG is the "flow" $\Lambda \mapsto P_\Lambda$ (e.g. something like a Fokker-Planck).

Upshot: Knowing how the distribution of states depends on scale $\Rightarrow$ computing phase transitions, etc.

Lattice Polchinski ERG Flow

For our mathematical model (as in [1]), we'll consider

- a lattice $X = (\Lambda(t)^{-1}\mathbb{Z}^d)/(L\mathbb{Z}^d)$, $\Lambda(t) = \Lambda_0 e^{-t}$ (here $\varepsilon = \Lambda^{-1}$).
- corresponding states $\varphi \in \mathbb{R}^X$
- a sequence of symmetric p.d. matrices $C_t = \int_0^t \dot{C}_s ds$ and their associated Gaussian measures $P_{C_t} = \mathcal{N}(0, C_t)$
- A series of inner products $\langle x, y \rangle_{\dot{C}_t} := \langle x, \dot{C}_t^{-1} y \rangle$.
- The differential operators associated to $\langle x, y \rangle_{\dot{C}_t}$: $\Delta_{\dot{C}_t}$, $\nabla_{\dot{C}_t}$, etc.

Lattice Polchinski ERG Flow

For our mathematical model (as in [1]), we'll consider

- a lattice $X = (\Lambda(t)^{-1}\mathbb{Z}^d)/(L\mathbb{Z}^d)$, $\Lambda(t) = \Lambda_0 e^{-t}$ (here $\varepsilon = \Lambda^{-1}$).
- corresponding states $\varphi \in \mathbb{R}^X$
- a sequence of symmetric p.d. matrices $C_t = \int_0^t \dot{C}_s ds$ and their associated Gaussian measures $P_{C_t} = \mathcal{N}(0, C_t)$
- A series of inner products $\langle x, y \rangle_{\dot{C}_t} := \langle x, \dot{C}_t^{-1} y \rangle$.
- The differential operators associated to $\langle x, y \rangle_{\dot{C}_t}$: $\Delta_{\dot{C}_t}$, $\nabla_{\dot{C}_t}$, etc.

Intuition:

- We imagine "flowing" through scales $\Lambda(t)^{-1}$ at time t.
- The measures P_{C_t} are the weights we use to "average" out small fluctations at each scale t.
- The inner product $\langle x, y \rangle_{\dot{C}_t}$ influences the geometry of $\mathbb{R}^X$ at each scale.

Lattice Polchinski ERG Flow

For our mathematical model (as in [1]), we'll consider

- a lattice $X = (\Lambda(t)^{-1}\mathbb{Z}^d)/(L\mathbb{Z}^d)$, $\Lambda(t) = \Lambda_0 e^{-t}$ (here $\varepsilon = \Lambda^{-1}$).
- corresponding states $\varphi \in \mathbb{R}^X$
- a sequence of symmetric p.d. matrices $C_t = \int_0^t \dot{C}_s ds$ and their associated Gaussian measures $P_{C_t} = \mathcal{N}(0, C_t)$
- A series of inner products $\langle x, y \rangle_{\dot{C}_t} := \langle x, \dot{C}_t^{-1} y \rangle$.
- The differential operators associated to $\langle x, y \rangle_{\dot{C}_t}$: $\Delta_{\dot{C}_t}$, $\nabla_{\dot{C}_t}$, etc.

Intuition:

- We imagine "flowing" through scales $\Lambda(t)^{-1}$ at time t.
- The measures P_{C_t} are the weights we use to "average" out small fluctations at each scale t.
- The inner product $\langle x, y \rangle_{\dot{C}_t}$ influences the geometry of $\mathbb{R}^X$ at each scale.

Example: the heat flow corresponds to $C_t = t\,\mathrm{Id}$, so $\dot{C}_t = \mathrm{Id}$.

Lattice Polchinski ERG Flow

Consider an initial potential $V_0 : \mathbb{R}^X \to \mathbb{R}$ and the assoiated measure σ_0 given by

$$\mathbb{E}_{\sigma_0}[F] \propto \mathbb{E}_{C_\infty}[e^{-V_0(\zeta)} F(\zeta)]$$

The Polchinski ERG evolves the potential & measure through different scales:

Lattice Polchinski ERG Flow

Consider an initial potential $V_0 : \mathbb{R}^X \to \mathbb{R}$ and the assoiated measure σ_0 given by

$$\mathbb{E}_{\sigma_0}[F] \propto \mathbb{E}_{C_\infty}[e^{-V_0(\zeta)} F(\zeta)]$$

The Polchinski ERG evolves the potential & measure through different scales: Suppose that $0 < s < t$ and $F : \mathbb{R}^N \to \mathbb{R}$ is bounded & measurable.

- The renormalized potential at time t is given by

$$V_t(\varphi) = -\log \mathbb{E}_{C_t}[e^{-V_0(\varphi+\zeta)}]$$

- The Polchinski semigroup (starting at time s and ending at time t) is given by

$$P_{s,t}F(\varphi) = e^{V_t(\varphi)} \mathbb{E}_{C_t - C_s}[e^{-V_s(\varphi+\zeta)} F(\varphi + \zeta)]$$

- The renormalized measure ν_t given by

$$\mathbb{E}_{\sigma_t}[F] = P_{t,\infty}F(0) = e^{V_\infty(0)} \mathbb{E}_{C_\infty - C_t}[e^{-V_t(\zeta)F(\zeta)}]$$

Insight: Connection with Fokker-Planck Variant

Recall that for a usual Wasserstein gradient flow, $\sigma : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$, the measures solves the following continuity eqation in a weak sense:

$$\partial_t \sigma_t = \text{div}\left(\sigma_t \nabla \frac{\delta \mathcal{F}}{\delta p}(\sigma_t)\right) = -\nabla^{W_2} \mathcal{F}(\sigma_t)$$

Insight: Connection with Fokker-Planck Variant

Recall that for a usual Wasserstein gradient flow, $\sigma : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$, the measures solves the following continuity eqation in a weak sense:

$$\partial_t \sigma_t = \operatorname{div}\left(\sigma_t \nabla \frac{\delta \mathcal{F}}{\delta p}(\sigma_t)\right) = -\nabla^{W_2} \mathcal{F}(\sigma_t)$$

By Itô's formula, we can show that the renormalized measures in the Polchinski flow satisfy Polchinski's equation:

$$\begin{aligned}\partial_t \sigma_t &= \operatorname{div}_{\dot{C}_t}\left(\sigma_t \nabla_{\dot{C}_t} \frac{\delta \mathcal{F}_t}{\delta p}(\sigma_t)\right) \\ &= -\nabla^{W_{2,\dot{C}_t}} \mathcal{F}_t(\sigma_t)\end{aligned}$$

where $\mathcal{F}_t(\rho) = \mathrm{KL}(\rho || \pi_t)$ for some curve of measures π_t.

Insight: Connection with Fokker-Planck Variant

Recall that for a usual Wasserstein gradient flow, $\sigma : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$, the measures solves the following continuity eqation in a weak sense:

$$\partial_t \sigma_t = \text{div}\left(\sigma_t \nabla \frac{\delta \mathcal{F}}{\delta p}(\sigma_t)\right) = -\nabla^{W_2} \mathcal{F}(\sigma_t)$$

By Itô's formula, we can show that the renormalized measures in the Polchinski flow satisfy Polchinski's equation:

$$\begin{aligned}\partial_t \sigma_t &= \text{div}_{\dot{C}_t}\left(\sigma_t \nabla_{\dot{C}_t} \frac{\delta \mathcal{F}_t}{\delta p}(\sigma_t)\right) \\ &= -\nabla^{W_{2,\dot{C}_t}} \mathcal{F}_t(\sigma_t)\end{aligned}$$

where $\mathcal{F}_t(\rho) = \text{KL}(\rho || \pi_t)$ for some curve of measures π_t.

Upshot: the Polchinski flow has a **gradient flow structure** where our metric and functional both depend on time.

Construction of Gradient Flows via JKO

Given

- initial data σ_0
- $\tau > 0$ and a partition of $[0, T]$ $\{0, \tau, 2\tau, \dots n\tau \leq T < (n+1)\tau\}$

we iteratively solve

$$\sigma_0^\tau = \sigma_0$$

$$\sigma_{k+1}^\tau = \arg\min_\sigma \mathcal{F}(\sigma) + \frac{1}{2\tau} W_2^2(\sigma, \sigma_k^\tau)$$

Construction of Gradient Flows via JKO

Given

- initial data σ_0
- $\tau > 0$ and a partition of $[0, T]$ $\{0, \tau, 2\tau, \dots n\tau \leq T < (n+1)\tau\}$

we iteratively solve

$$\sigma_0^\tau = \sigma_0$$

$$\sigma_{k+1}^\tau = \arg\min_\sigma \mathcal{F}(\sigma) + \frac{1}{2\tau} W_2^2(\sigma, \sigma_k^\tau)$$

For "nice" functionals such as

$$\mathcal{F}(\rho) = \int \rho \log \rho + \int V d\mu, \text{ with } V \ \lambda \text{ convex}$$

- σ^τ converges to a unique solution to our PDE as $\tau \to 0$
- converge to their stationary state exponentially in time

Gradient Flows on Time Dependent Metric Measure Spaces

Despite the added time dependencies, we can tweak the JKO scheme in a simple way. Given

- initial data σ_0
- $\tau > 0$ and a partition of $[0, T]$ $\{0, \tau, 2\tau, \dots n\tau \leq T < (n+1)\tau\}$

we can iteratively solve

$$\sigma_0^\tau = \sigma_0$$

$$\sigma_{k+1}^\tau = \arg\min_\sigma \mathcal{F}_{(k+1)\tau}(\sigma) + \frac{1}{2\tau} W^2_{2,(k+1)\tau}(\sigma, \sigma_k^\tau)$$

where $\mathcal{F}_t(\rho) = \mathrm{KL}(\rho||\pi_t)$

Gradient Flows on Time Dependent Metric Measure Spaces

Despite the added time dependencies, we can tweak the JKO scheme in a simple way. Given

- initial data σ_0
- $\tau > 0$ and a partition of $[0, T]$ $\{0, \tau, 2\tau, \dots n\tau \leq T < (n+1)\tau\}$

we can iteratively solve

$$\sigma_0^\tau = \sigma_0$$

$$\sigma_{k+1}^\tau = \arg\min_\sigma \mathcal{F}_{(k+1)\tau}(\sigma) + \frac{1}{2\tau} W^2_{2,(k+1)\tau}(\sigma, \sigma_k^\tau)$$

where $\mathcal{F}_t(\rho) = \mathrm{KL}(\rho||\pi_t)$ If our potentials V_t and metrics $\dot{C}_t$ evolve "nicely" (e.g. some uniform Lipschitz conditions in terms of time and boundedness, convexity of potentials, etc.) [2]

- **the curve limit exists, is unique, and solves Polchinski's equation**
- **long time behavior can be determined (may not converge to equilibrium)**

Long Term Goals

The previous results only apply under extremely restrictive assumptions (e.g. the so called "free fields") on a lattice.

Long Term Goals

The previous results only apply under extremely restrictive assumptions (e.g. the so called "free fields") on a lattice.

- Physicists want to understand more complicated potentials such as the φ^4 model or continuum Sine-Gordon model. [3]
- The dream is to understand what happens as the lattice approaches $\mathbb{R}^d$ e.g. $\Lambda, L \to \infty$. [4] Can we say something in this context?
- The continuum case has only been successfully studied by singular SPDE techniques; can we relate our approach to theirs?

References

[1] R. Bauerschmidt, T. Bodineau, and B. Dagallier, **"Stochastic dynamics and the Polchinski equation: An introduction,"** *Probability Surveys*, arXiv:2307.07619 [math].

[2] E. Kopfer and K.-T. Sturm, **Boltzman Entropy Gradient Flow on Time Dependent Metric Spaces,** arXiv:1611.02570 [math].

[3] J. Cotler and S. Rezchikov, **"Renormalization group flow as optimal transport,"** *Physical Review D*,

[4] S. Sheffield, **Gaussian free fields for mathematicians,** arXiv:math/0312099.

Learning empowered structure-preserving particle method for homogeneous Landau equation

Yan Huang (joint with Li Wang)

School of Mathematics, University of Minnesota

OT summer school, UCSB
Jul 24, 2025

Landau equation

- The Landau equation models the distribution of charged particles in collisional plasmas:

$$\partial_t f + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f + F \cdot \nabla_{\boldsymbol{v}} f = Q_L(f, f)\,,$$

$$Q_L(f, f) = \nabla_{\boldsymbol{v}} \cdot \int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*)\left(f(\boldsymbol{v}_*)\nabla_{\boldsymbol{v}} f(\boldsymbol{v}) - f(\boldsymbol{v})\nabla_{\boldsymbol{v}_*} f(\boldsymbol{v}_*)\right) \mathrm{d}\boldsymbol{v}_*\,.$$

  with the collision kernel $A(\boldsymbol{z}) = C_\gamma |\boldsymbol{z}|^{\gamma+2}\left(I_d - \frac{\boldsymbol{z}\otimes\boldsymbol{z}}{|\boldsymbol{z}|^2}\right)$. The physically relevant case is $d = 3, \gamma = -3$, often referred to as the Coulomb case.
- The Landau operator $Q_L(f, f)$ conserves mass, momentum, energy, and is entropy dissipative.
- Computational difficulty of $Q_L(f, f)$: high-dimensionality, multi-scale, strong nonlinearity and non-locality, structure-preserving.

Blob method [Carrillo et'al 20']

- Continuity equation form: $\partial_t f + \nabla \cdot (f \boldsymbol{U}[f]) = 0$, where the velocity field

$$\boldsymbol{U}[f] = -\int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) \left(\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*} \right) f_* \mathrm{d}\boldsymbol{v}_* \,, \quad \mathcal{H} = \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} \,.$$

- A particle representation: $f^N(\boldsymbol{v}) = \sum_{i=1}^N w_i \delta(\boldsymbol{v} - \boldsymbol{v}_i(t))$.
- Entropy regularization: $\mathcal{H}_\varepsilon = \int_{\mathbb{R}^d} (f * \psi_\varepsilon) \log(f * \psi_\varepsilon) \mathrm{d}\boldsymbol{v}$, where ψ_ε is a mollifier.
- Evolution of particles:

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{v}_i(t) = -\sum_{j=1}^N w_j A(\boldsymbol{v}_i(t) - \boldsymbol{v}_j(t)) \left(\nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_i(t)) - \nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_j(t)) \right) \,,$$

$$\nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_i(t)) = \int_{\mathbb{R}^d} \nabla \psi_\varepsilon(\boldsymbol{v}_i(t) - \boldsymbol{v}) \log \left(\sum_k w_k \psi_\varepsilon(\boldsymbol{v} - \boldsymbol{v}_k(t)) \right) \mathrm{d}\boldsymbol{v} \,.$$

Pros: structure-preserving.
Cons: explicit scheme; kernel density estimation; computational cost is $\mathcal{O}(N^2)$.

Score-based particle method

- A "Log" form of continuity equation:

$$\partial_t f + \nabla_{\boldsymbol{v}} \cdot (\boldsymbol{U}[f] f) = 0\,,$$

$$\boldsymbol{U}[f] = -\int_{\mathbb{R}^d} A(\boldsymbol{v}-\boldsymbol{v}_*)(\underbrace{\nabla_{\boldsymbol{v}} \log f(\boldsymbol{v})}_{\textit{score}} - \nabla_{\boldsymbol{v}_*} \log f(\boldsymbol{v}_*)) f_* \mathrm{d}\boldsymbol{v}_*\,.$$

- Learn score via the score-matching loss:

$$\boldsymbol{s}_\theta^n(\boldsymbol{v}) \in \arg\min_\theta \frac{1}{N}\sum_{i=1}^N |\boldsymbol{s}_\theta(\boldsymbol{v}_i^n)|^2 + 2\nabla \cdot \boldsymbol{s}_\theta(\boldsymbol{v}_i^n)$$

- Update particles: $\boldsymbol{v}_i^{n+1} = \boldsymbol{v}_i^n - \Delta t \frac{1}{N}\sum_{j=1}^N A(\boldsymbol{v}_i^n - \boldsymbol{v}_j^n)[\boldsymbol{s}_\theta^n(\boldsymbol{v}_i^n) - \boldsymbol{s}_\theta^n(\boldsymbol{v}_j^n)]$.
- Update density (no kernel density estimation):

$$I_i^{n+1} = -\Delta t \frac{1}{N}\sum_{j=1}^N \nabla_{\boldsymbol{v}_i} \cdot \{A(\boldsymbol{v}_i^n - \boldsymbol{v}_j^n)[\boldsymbol{s}_\theta(\boldsymbol{v}_i^n) - \boldsymbol{s}_\theta(\boldsymbol{v}_j^n)]\}\,, f^{n+1}(\boldsymbol{v}_i^{n+1}) = f^n(\boldsymbol{v}_i^n)/\exp(I_i^{n+1})\,.$$

Landau equation as a gradient flow [Carrillo et'al 24']

- Heat equation: $\partial_t f = \Delta f = \nabla \cdot \left(f \nabla \frac{\delta \mathcal{H}}{\delta f} \right)$. The entropy dissipation rate is

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} = - \int_{\mathbb{R}^d} \left| \nabla \frac{\delta \mathcal{H}}{\delta f} \right|^2 f \mathrm{d}\boldsymbol{v} \,.$$

- 2-Wasserstein metric:

$$\begin{aligned} d_{W_2}^2(f_0, f_1) := & \inf_{f, \boldsymbol{u}} \int_0^1 \int_{\mathbb{R}^d} |\boldsymbol{u}|^2 f \mathrm{d}\boldsymbol{v} \mathrm{d}t \,, \\ s.t. \;\; & \partial_t f + \nabla \cdot (\boldsymbol{u} f) = 0 \,, \;\; f(0, \cdot) = f_0 \,, \;\; f(1, \cdot) = f_1 \,. \end{aligned}$$

- Landau equation: $\partial_t f = \nabla \cdot \left(f \int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) f_* \mathrm{d}\boldsymbol{v}_* \right)$. The entropy dissipation rate is

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} = -\frac{1}{2} \iint_{\mathbb{R}^{2d}} (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) A(\boldsymbol{v} - \boldsymbol{v}_*) (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) f f_* \mathrm{d}\boldsymbol{v} \mathrm{d}\boldsymbol{v}_* \,.$$

- Landau metric:

$$\begin{aligned} d_L^2(f_0, f_1) := & \inf_{f, \boldsymbol{u}} \frac{1}{2} \int_0^1 \iint_{\mathbb{R}^{2d}} (\boldsymbol{u} - \boldsymbol{u}_*) A(\boldsymbol{v} - \boldsymbol{v}_*) (\boldsymbol{u} - \boldsymbol{u}_*) f f_* \mathrm{d}\boldsymbol{v} \mathrm{d}\boldsymbol{v}_* \mathrm{d}t \,, \\ s.t. \;\; & \partial_t f + \nabla \cdot \left[f \left(\int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) (\boldsymbol{u} - \boldsymbol{u}_*) f_* \mathrm{d}\boldsymbol{v}_* \right) \right] = 0 \,, \;\; f(0, \cdot) = f_0 \,, \;\; f(1, \cdot) = f_1 \,. \end{aligned}$$

Dynamic JKO scheme

As a result, the Landau equation can be viewed as the gradient flow of entropy $\mathcal{H}$ with respect to the metric d_L. Therefore, one can construct a weak solution by the Jordan-Kinderlehrer-Otto (JKO) scheme:

$$f^0 = f(0,\cdot)\,,\ \ f^{n+1} \in \arg\min_f \left[d_L^2(f, f^n) + 2\Delta t \mathcal{H}(f) \right] .$$

Numerically, we use the dynamic JKO scheme: given f^n, solve $f^{n+1} := f(1,\cdot)$ by

$$\begin{cases} \inf\limits_{f,\boldsymbol{u}} \ \dfrac{1}{2}\displaystyle\int_0^1 \iint_{\mathbb{R}^{2d}} |\boldsymbol{u}-\boldsymbol{u}_*|_A^2 f f_* \mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{v}_*\mathrm{d}t + 2\Delta t \mathcal{H}(f(1,\cdot))\,, \\ s.t.\ \partial_t f + \nabla\cdot\left[f\left(\displaystyle\int_{\mathbb{R}^d} A(\boldsymbol{v}-\boldsymbol{v}_*)(\boldsymbol{u}-\boldsymbol{u}_*) f_* \mathrm{d}\boldsymbol{v}_* \right)\right] = 0\,,\ \ f(0,\cdot) = f^n\,. \end{cases}$$

Lagrangian dynamic JKO scheme

Given f^n, solve the optimal flow map T_t^{n+1} by

$$\begin{cases} \inf_{\boldsymbol{u}} \frac{1}{2}\int_0^1 \iint_{\mathbb{R}^{2d}} |\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))|_A^2 f^n f_*^n \mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{v}_*\mathrm{d}t - 2\Delta t \int_{\mathbb{R}^d} \log|\det \nabla_{\boldsymbol{v}} T_1(\boldsymbol{v})| f^n \mathrm{d}\boldsymbol{v} \\ s.t.\ \frac{\mathrm{d}}{\mathrm{d}t} T_t(\boldsymbol{v}) = \int_{\mathbb{R}^d} A(T_t(\boldsymbol{v})-T_t(\boldsymbol{v}_*))[\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))] f_*^n \mathrm{d}\boldsymbol{v}_* \\ \frac{\mathrm{d}}{\mathrm{d}t}\log|\det \nabla_{\boldsymbol{v}} T_t(\boldsymbol{v})| = \nabla_{T_t(\boldsymbol{v})} \cdot \int_{\mathbb{R}^d} A(T_t(\boldsymbol{v})-T_t(\boldsymbol{v}_*))[\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))] f_*^n \mathrm{d}\boldsymbol{v}_* \end{cases}$$

Then we obtain $f^{n+1} := {T_1}_{\#}^{n+1} f^n$.

JKO-based particle method

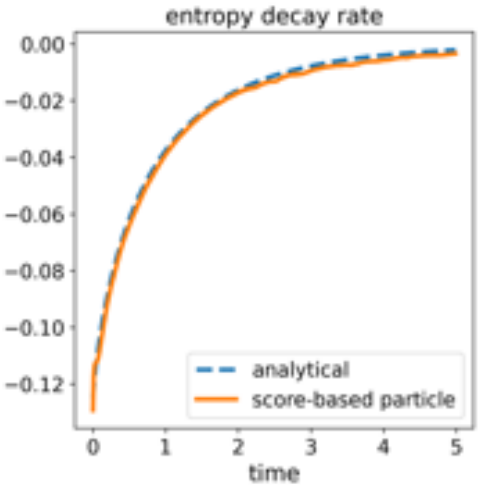
- Algorithm: Given $\{\boldsymbol{v}_i^n\}_{i=1}^N$ and $\{f^n(\boldsymbol{v}_i^n)\}_{i=1}^N$, solving the following variational problem:

$$\begin{cases} \inf\limits_{\boldsymbol{u}} \ \dfrac{1}{2N^2}\displaystyle\sum_{i=1}^N\sum_{j=1}^N\int_0^1 |\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))|_A^2 \mathrm{d}t - \dfrac{2\tau}{N}\displaystyle\sum_{i=1}^N \ell_i(1)\,, \\ s.t.\ \dfrac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \dfrac{1}{N}\displaystyle\sum_{j=1}^N A(\boldsymbol{v}_i(t)-\boldsymbol{v}_j(t))[\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))]\,,\ \boldsymbol{v}_i(0)=\boldsymbol{v}_i^n\,, \\ \dfrac{\mathrm{d}\ell_i}{\mathrm{d}t} = \dfrac{1}{N}\displaystyle\sum_{j=1}^N \nabla_{\boldsymbol{v}_i}\cdot\{A(\boldsymbol{v}_i(t)-\boldsymbol{v}_j(t))[\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))]\}\ ,\ \ell_i(0)=0\,. \end{cases}$$

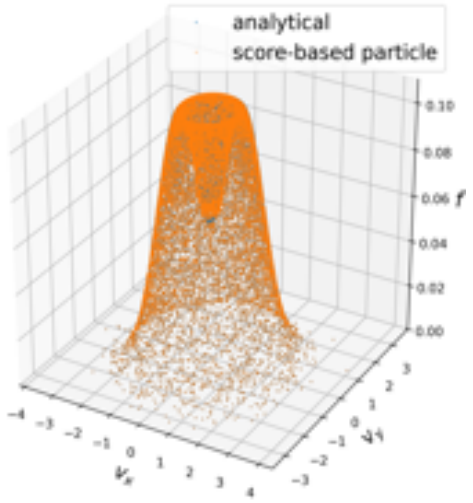
  Then compute $\boldsymbol{v}_i^{n+1}=\boldsymbol{v}_i(1)$ and $f^{n+1}(\boldsymbol{v}_i^{n+1})=f^n(\boldsymbol{v}_i^n)/\exp(\ell_i(1))$.

- Implementation: Neural network approximation for $\boldsymbol{u}$; flexible inner time discretization; stochastic optimization and random batch particle method

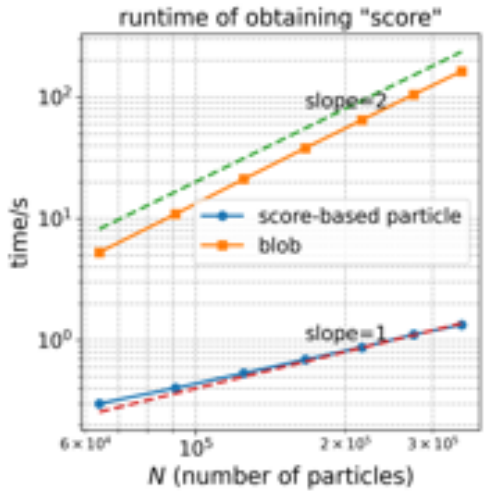
Numerical experiments—score



(a) Time evolution of the entropy decay rate.

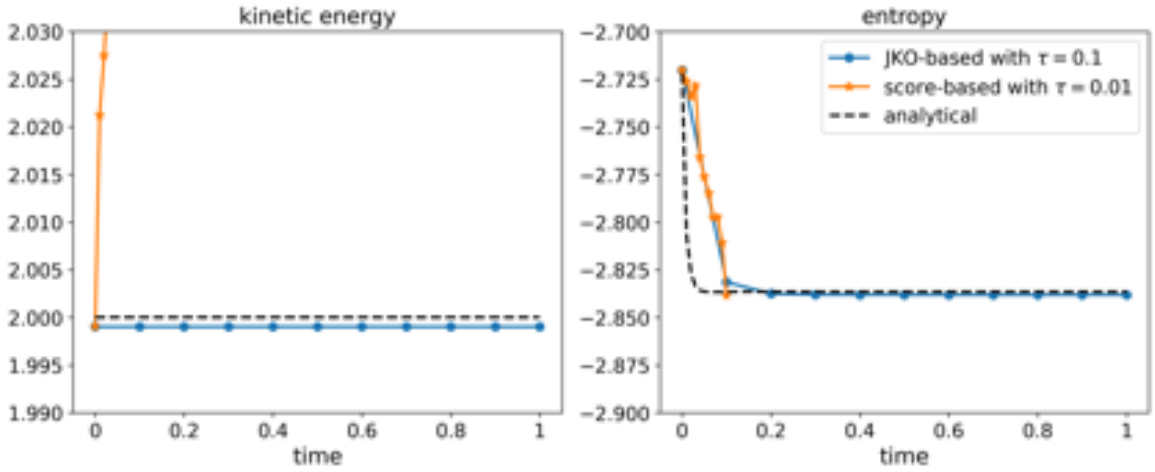


(b) Density visualization at particle locations.



(c) Computational time of obtaining "score" on GPU

Numerical experiments–JKO



(a) Comparison between JKO-based and score-based particle methods with varying time step sizes in strong collision regime.

References:
[1] Y. Huang and L. Wang, A score-based particle method for homogeneous Landau equation, Journal of Computational Physics, (2025), p. 114053.
[2] Y. Huang and L. Wang, JKO for Landau: a variational particle method for homogeneous Landau equation, https://arxiv.org/abs/2409.12296.

Phase Transitions and Linear Stability for the mean-field Kuramoto-Daido Model

Kyunghoo Mun
Joint work with Matthew Rosenzweig

Carnegie Mellon University

July 24, 2025

Mean-field interacting particle systems

We consider a system of N particles on the one-dimensional torus $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ by,

$$d\theta_i = \frac{K}{N} \sum_{j=1}^{N} \nabla W(\theta_i - \theta_j)\, dt + dB_i, \quad 1 \leq i \leq N,$$

where W is a $\mathbb{T}$-periodic interaction function, $(B_i)_{i=1}^N$ are independent Brownian motions, and $K > 0$ is the interaction strength parameter.

Statistical perspective

We aim to understand the behavior of this interacting particle systems where $N \gg 1$. For example, when modeling particles in gas, we have $N \approx 10^{23}$ (Avogadro's number). Due to the large system size N, its nonlinearity, and the coupling between particles, tracking all individual trajectories become computationally infeasible.

Key idea

Analyze the **approximate dynamics** of a single particle from a statistical point of view.

We study the limiting behavior of the empirical measure

$$d\nu_{N,t} = \frac{1}{N}\sum_{i=1}^{N} d\delta_{\theta_i(t)},$$

as $N \to \infty$.

Mean-field equation

For a smooth interaction $W \in \mathcal{C}^\infty(\mathbb{T})$, we obtain the mean-field limit

$$\nu_{N,t} \to \nu_t, \quad \text{weakly as } N \to \infty,$$

where ν_t is absolutely continuous with a smooth density $q_t(\theta)$, which solves the McKean-Vlasov equation (mean-field equation)

$$\partial_t q_t = \frac{1}{2}\Delta q_t - K\,\nabla\Big(q_t\,\nabla(W * q_t)\Big).$$

Linearized McKean-Vlasov equation

Suppose $q \in \mathcal{C}^2(\mathbb{T})$ is a stationary solution of the McKean-Vlasov equation, *i.e.*

$$0 = \frac{1}{2} q'' - K\Big(q(W' * q)\Big)'.$$

Goal

Understand in which situation a solution converges to a stationary state q.

The first step is to study dynamics when the initial condition is a perturbation of q. Consider the evolution of the perturbation $u_t(\theta) = q_t(\theta) - q(\theta)$. Then it follows

$$\partial_t u_t(\theta) = L_q u_t(\theta) - K\Big(u_t(W' * u_t)\Big)'.$$

Here, L_q is the linear operator defined by

$$L_q u = \frac{1}{2} u'' - K\Big(q(W' * u) + u(W' * q)\Big)'.$$

When the initial perturbation u_0 is sufficiently small, then we may only consider the linear term in the equation as,

$$\partial_t u_t(\theta) = L_q u_t(\theta).$$

This is the **linearized McKean-Vlasov equation** at the stationary solution q, which is the primary object in this talk.

Free energy

The McKean-Vlasov equation can be understood as a gradient flow for the free energy functional $\mathcal{F}$ with respect to the Wasserstein-2 distance on the space of probability measures with finite second moments as,

$$\partial_t q_t = \nabla \left[q_t \, \nabla \Big(\frac{\delta \mathcal{F}(q_t)}{\delta q_t} \Big) \right],$$

where $\delta \mathcal{F} / \delta q$ is L^2 Fréchet derivative of the functional $\mathcal{F}$. The free energy $\mathcal{F}$ is defined by

$$\mathcal{F}(q) = \frac{1}{2} \int q \log q \; d\theta - \frac{K}{2} \int q (W * q) \; d\theta.$$

Therefore, a critical point of $\mathcal{F}$ is equivalent to a stationary solution of the McKean-Vlasov equation.

Phase transition

For smooth attractive interaction W, a *phase transition* occurs, *i.e.* there exists a critical interaction strength $K_c > 0$ such that

1. For $K < K_c$, the uniform distribution $q_{\text{unif}} = (2\pi)^{-1}$ is the unique minimizer of the free energy $\mathcal{F}$.
2. For $K > K_c$, there exists a non-uniform minimizer $q \neq q_{\text{unif}}$ of $\mathcal{F}$.

Remark. By the continuity of $\mathcal{F}$, the uniform distribution q_{unif} remains a minimizer when $K = K_c$.

Kuramoto model

The Kuramoto model corresponds to the the interaction potential

$$W(\theta) = \cos\theta.$$

Let $q(\theta)$ be a stationary solution of the McKean-Vlasov equation. Defining its first Fourier coefficient as $r = \int q(\varphi)\cos\varphi \; d\varphi$, the stationary solution takes the form

$$q(\theta) = \frac{\exp(2Kr\cos\theta)}{\int \exp(2Kr\cos\theta) \; d\theta},$$

where $r = r(K)$ solves the self-consistency equation

$$r = \frac{\int \cos\theta \exp(2Kr\cos\theta) \; d\theta}{\int \exp(2Kr\cos\theta) \; d\theta}.$$

Equivalently, r is a critical point of (parameterized) free energy functional

$$F_K(r) = Kr^2 - \log\left(\frac{1}{2\pi}\int \exp(2Kr\cos\theta) \; d\theta\right)$$

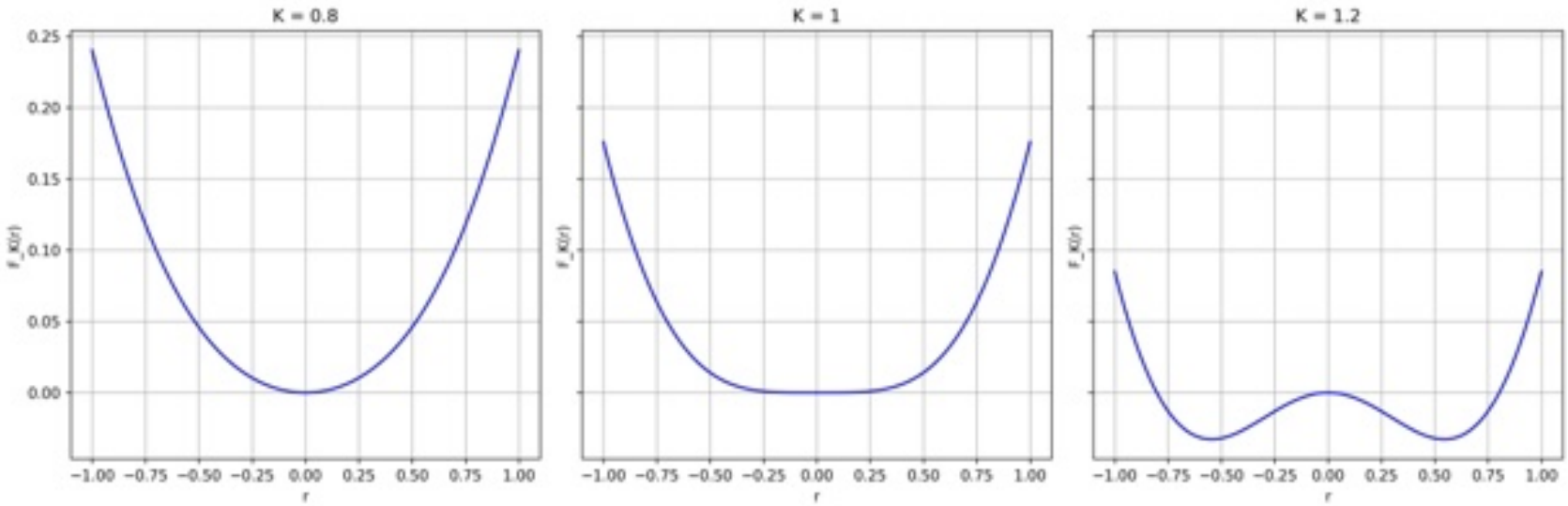


Figure: Parameterized free energy $F_K(r)$ for different $K \in \{0.8, 1, 1.2\}$.

Kuramoto-Daido model

Kuramoto-Daido model was introduced by Daido (1995); Daido (1996), with the interaction potential

$$W(\theta) = \cos\theta + m\cos 2\theta, \quad m > 0.$$

Daido proposed the model to study multibranch synchronization phenomena, which involve interactions through multiple Fourier modes. Through numerical simulations and heuristic arguments, he demonstrated the occurrence of phase transitions. Since then, the Kuramoto-Daido model has been studied in physics literature Hansel-Mato-Meunier (1993); Komarov-Pikovsky (2014); Cobero-Politi-Rosenblum (2016).

Remark. For $m < 0$, this model has $K_c = 1$, as in the Kuramoto model, analyzed by Vukadinovic (2023).

Phase Transition in the uniqueness of the stationary solution

There exists a uniform constant $m_* \in (1, 2)$ such that

1. For $m \in [0, 1/2]$, $K_c = 1$ and the transition is continuous.
2. For $m \in (1/2, 1]$, $K_c < 1$ and the transition is discontinuous.
3. For $m \in (1, m_*)$, $K_c < m^{-1}$ and the transition is discontinuous.
4. For $m = m_*$, $K_c = m^{-1}$ and the transition is discontinuous.
5. For $m \in (m_*, \infty)$, $K_c = m^{-1}$ and the transition is continuous.

Linear stability at the non-uniform stationary solution q

Assume $K > 1$ and $m \in (0, 8.568 \times 10^{-4}]$. Then the spectrum of linearized McKean-Vlasov operator L_q consists solely of pure points lying in $(-\infty, 0]$, and it includes the eigenvalue 0 whose one-dimensional eigenspace is spanned by q'. Moreover, the spectral gap is bounded below by

$$\mathrm{gap}(L_q) \geq C,$$

where the positive constant C is a function of K, m.

Thank you for your attention.

Learning empowered structure-preserving particle method for homogeneous Landau equation

Yan Huang (joint with Li Wang)

School of Mathematics, University of Minnesota

OT summer school, UCSB
Jul 24, 2025

Landau equation

- The Landau equation models the distribution of charged particles in collisional plasmas:

$$\partial_t f + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f + F \cdot \nabla_{\boldsymbol{v}} f = Q_L(f, f)\,,$$

$$Q_L(f, f) = \nabla_{\boldsymbol{v}} \cdot \int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*)\left(f(\boldsymbol{v}_*)\nabla_{\boldsymbol{v}} f(\boldsymbol{v}) - f(\boldsymbol{v})\nabla_{\boldsymbol{v}_*} f(\boldsymbol{v}_*)\right) \mathrm{d}\boldsymbol{v}_*\,.$$

  with the collision kernel $A(\boldsymbol{z}) = C_\gamma |\boldsymbol{z}|^{\gamma+2}\left(I_d - \frac{\boldsymbol{z}\otimes\boldsymbol{z}}{|\boldsymbol{z}|^2}\right)$. The physically relevant case is $d = 3, \gamma = -3$, often referred to as the Coulomb case.
- The Landau operator $Q_L(f, f)$ conserves mass, momentum, energy, and is entropy dissipative.
- Computational difficulty of $Q_L(f, f)$: high-dimensionality, multi-scale, strong nonlinearity and non-locality, structure-preserving.

Blob method [Carrillo et'al 20']

- Continuity equation form: $\partial_t f + \nabla \cdot (f \boldsymbol{U}[f]) = 0$, where the velocity field

$$\boldsymbol{U}[f] = -\int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) \left(\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*} \right) f_* \mathrm{d}\boldsymbol{v}_* \,, \quad \mathcal{H} = \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} \,.$$

- A particle representation: $f^N(\boldsymbol{v}) = \sum_{i=1}^N w_i \delta(\boldsymbol{v} - \boldsymbol{v}_i(t))$.
- Entropy regularization: $\mathcal{H}_\varepsilon = \int_{\mathbb{R}^d} (f * \psi_\varepsilon) \log(f * \psi_\varepsilon) \mathrm{d}\boldsymbol{v}$, where ψ_ε is a mollifier.
- Evolution of particles:

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{v}_i(t) = -\sum_{j=1}^N w_j A(\boldsymbol{v}_i(t) - \boldsymbol{v}_j(t)) \left(\nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_i(t)) - \nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_j(t)) \right) \,,$$

$$\nabla \frac{\delta \mathcal{H}_\varepsilon^N}{\delta f}(\boldsymbol{v}_i(t)) = \int_{\mathbb{R}^d} \nabla \psi_\varepsilon(\boldsymbol{v}_i(t) - \boldsymbol{v}) \log \left(\sum_k w_k \psi_\varepsilon(\boldsymbol{v} - \boldsymbol{v}_k(t)) \right) \mathrm{d}\boldsymbol{v} \,.$$

Pros: structure-preserving.
Cons: explicit scheme; kernel density estimation; computational cost is $\mathcal{O}(N^2)$.

Score-based particle method

- A "Log" form of continuity equation:

$$\partial_t f + \nabla_{\boldsymbol{v}} \cdot (\boldsymbol{U}[f] f) = 0\,,$$

$$\boldsymbol{U}[f] = -\int_{\mathbb{R}^d} A(\boldsymbol{v}-\boldsymbol{v}_*)(\underbrace{\nabla_{\boldsymbol{v}} \log f(\boldsymbol{v})}_{\textit{score}} - \nabla_{\boldsymbol{v}_*} \log f(\boldsymbol{v}_*)) f_* \mathrm{d}\boldsymbol{v}_*\,.$$

- Learn score via the score-matching loss:

$$\boldsymbol{s}_\theta^n(\boldsymbol{v}) \in \arg\min_\theta \frac{1}{N}\sum_{i=1}^N |\boldsymbol{s}_\theta(\boldsymbol{v}_i^n)|^2 + 2\nabla \cdot \boldsymbol{s}_\theta(\boldsymbol{v}_i^n)$$

- Update particles: $\boldsymbol{v}_i^{n+1} = \boldsymbol{v}_i^n - \Delta t \frac{1}{N}\sum_{j=1}^N A(\boldsymbol{v}_i^n - \boldsymbol{v}_j^n)[\boldsymbol{s}_\theta^n(\boldsymbol{v}_i^n) - \boldsymbol{s}_\theta^n(\boldsymbol{v}_j^n)]$.
- Update density (no kernel density estimation):

$$I_i^{n+1} = -\Delta t \frac{1}{N}\sum_{j=1}^N \nabla_{\boldsymbol{v}_i} \cdot \{A(\boldsymbol{v}_i^n - \boldsymbol{v}_j^n)[\boldsymbol{s}_\theta(\boldsymbol{v}_i^n) - \boldsymbol{s}_\theta(\boldsymbol{v}_j^n)]\}\,, f^{n+1}(\boldsymbol{v}_i^{n+1}) = f^n(\boldsymbol{v}_i^n)/\exp(I_i^{n+1})\,.$$

Landau equation as a gradient flow [Carrillo et'al 24']

- Heat equation: $\partial_t f = \Delta f = \nabla \cdot \left(f \nabla \frac{\delta \mathcal{H}}{\delta f} \right)$. The entropy dissipation rate is

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} = - \int_{\mathbb{R}^d} \left| \nabla \frac{\delta \mathcal{H}}{\delta f} \right|^2 f \mathrm{d}\boldsymbol{v} \,.$$

- 2-Wasserstein metric:

$$\begin{aligned} d_{W_2}^2(f_0, f_1) := & \inf_{f, \boldsymbol{u}} \int_0^1 \int_{\mathbb{R}^d} |\boldsymbol{u}|^2 f \mathrm{d}\boldsymbol{v} \mathrm{d}t \,, \\ s.t. \;\; & \partial_t f + \nabla \cdot (\boldsymbol{u} f) = 0 \,, \;\; f(0, \cdot) = f_0 \,, \;\; f(1, \cdot) = f_1 \,. \end{aligned}$$

- Landau equation: $\partial_t f = \nabla \cdot \left(f \int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) f_* \mathrm{d}\boldsymbol{v}_* \right)$. The entropy dissipation rate is

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} f \log f \mathrm{d}\boldsymbol{v} = -\frac{1}{2} \iint_{\mathbb{R}^{2d}} (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) A(\boldsymbol{v} - \boldsymbol{v}_*) (\nabla \frac{\delta \mathcal{H}}{\delta f} - \nabla_* \frac{\delta \mathcal{H}_*}{\delta f_*}) f f_* \mathrm{d}\boldsymbol{v} \mathrm{d}\boldsymbol{v}_* \,.$$

- Landau metric:

$$\begin{aligned} d_L^2(f_0, f_1) := & \inf_{f, \boldsymbol{u}} \frac{1}{2} \int_0^1 \iint_{\mathbb{R}^{2d}} (\boldsymbol{u} - \boldsymbol{u}_*) A(\boldsymbol{v} - \boldsymbol{v}_*) (\boldsymbol{u} - \boldsymbol{u}_*) f f_* \mathrm{d}\boldsymbol{v} \mathrm{d}\boldsymbol{v}_* \mathrm{d}t \,, \\ s.t. \;\; & \partial_t f + \nabla \cdot \left[f \left(\int_{\mathbb{R}^d} A(\boldsymbol{v} - \boldsymbol{v}_*) (\boldsymbol{u} - \boldsymbol{u}_*) f_* \mathrm{d}\boldsymbol{v}_* \right) \right] = 0 \,, \;\; f(0, \cdot) = f_0 \,, \;\; f(1, \cdot) = f_1 \,. \end{aligned}$$

Dynamic JKO scheme

As a result, the Landau equation can be viewed as the gradient flow of entropy $\mathcal{H}$ with respect to the metric d_L. Therefore, one can construct a weak solution by the Jordan-Kinderlehrer-Otto (JKO) scheme:

$$f^0 = f(0,\cdot)\,,\ \ f^{n+1} \in \arg\min_f \left[d_L^2(f, f^n) + 2\Delta t \mathcal{H}(f) \right] .$$

Numerically, we use the dynamic JKO scheme: given f^n, solve $f^{n+1} := f(1,\cdot)$ by

$$\begin{cases} \inf\limits_{f,\boldsymbol{u}} \ \dfrac{1}{2}\displaystyle\int_0^1 \iint_{\mathbb{R}^{2d}} |\boldsymbol{u}-\boldsymbol{u}_*|_A^2 f f_* \mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{v}_*\mathrm{d}t + 2\Delta t \mathcal{H}(f(1,\cdot))\,, \\ s.t.\ \partial_t f + \nabla\cdot\left[f\left(\displaystyle\int_{\mathbb{R}^d} A(\boldsymbol{v}-\boldsymbol{v}_*)(\boldsymbol{u}-\boldsymbol{u}_*) f_* \mathrm{d}\boldsymbol{v}_* \right)\right] = 0\,,\ \ f(0,\cdot) = f^n\,. \end{cases}$$

Lagrangian dynamic JKO scheme

Given f^n, solve the optimal flow map T_t^{n+1} by

$$\begin{cases} \inf_{\boldsymbol{u}} \frac{1}{2}\int_0^1 \iint_{\mathbb{R}^{2d}} |\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))|_A^2 f^n f_*^n \mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{v}_*\mathrm{d}t - 2\Delta t \int_{\mathbb{R}^d} \log|\det \nabla_{\boldsymbol{v}} T_1(\boldsymbol{v})| f^n \mathrm{d}\boldsymbol{v} \\ s.t.\ \frac{\mathrm{d}}{\mathrm{d}t} T_t(\boldsymbol{v}) = \int_{\mathbb{R}^d} A(T_t(\boldsymbol{v})-T_t(\boldsymbol{v}_*))[\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))] f_*^n \mathrm{d}\boldsymbol{v}_* \\ \frac{\mathrm{d}}{\mathrm{d}t}\log|\det \nabla_{\boldsymbol{v}} T_t(\boldsymbol{v})| = \nabla_{T_t(\boldsymbol{v})} \cdot \int_{\mathbb{R}^d} A(T_t(\boldsymbol{v})-T_t(\boldsymbol{v}_*))[\boldsymbol{u}(t,T_t(\boldsymbol{v}))-\boldsymbol{u}(t,T_t(\boldsymbol{v}_*))] f_*^n \mathrm{d}\boldsymbol{v}_* \end{cases}$$

Then we obtain $f^{n+1} := {T_1}_{\#}^{n+1} f^n$.

JKO-based particle method

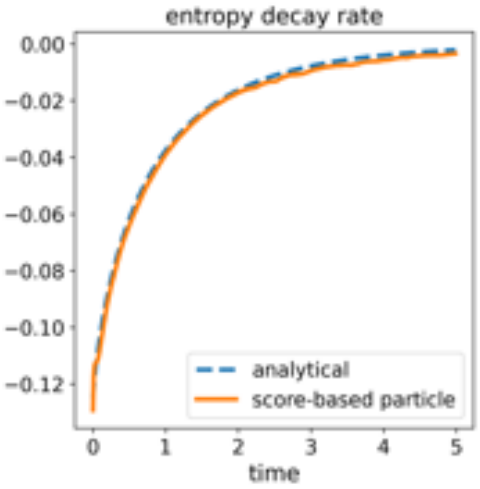
- Algorithm: Given $\{\boldsymbol{v}_i^n\}_{i=1}^N$ and $\{f^n(\boldsymbol{v}_i^n)\}_{i=1}^N$, solving the following variational problem:

$$\begin{cases} \inf\limits_{\boldsymbol{u}} \ \dfrac{1}{2N^2}\displaystyle\sum_{i=1}^N\sum_{j=1}^N\int_0^1 |\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))|_A^2 \mathrm{d}t - \dfrac{2\tau}{N}\displaystyle\sum_{i=1}^N \ell_i(1)\,, \\ s.t.\ \dfrac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} = \dfrac{1}{N}\displaystyle\sum_{j=1}^N A(\boldsymbol{v}_i(t)-\boldsymbol{v}_j(t))[\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))]\,,\ \boldsymbol{v}_i(0)=\boldsymbol{v}_i^n\,, \\ \dfrac{\mathrm{d}\ell_i}{\mathrm{d}t} = \dfrac{1}{N}\displaystyle\sum_{j=1}^N \nabla_{\boldsymbol{v}_i}\cdot\{A(\boldsymbol{v}_i(t)-\boldsymbol{v}_j(t))[\boldsymbol{u}(t,\boldsymbol{v}_i(t))-\boldsymbol{u}(t,\boldsymbol{v}_j(t))]\}\ ,\ \ell_i(0)=0\,. \end{cases}$$

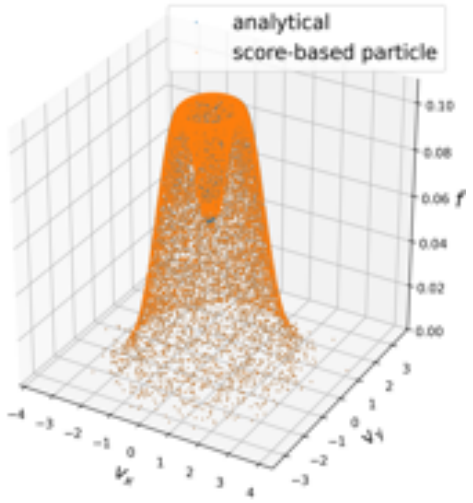
  Then compute $\boldsymbol{v}_i^{n+1}=\boldsymbol{v}_i(1)$ and $f^{n+1}(\boldsymbol{v}_i^{n+1})=f^n(\boldsymbol{v}_i^n)/\exp(\ell_i(1))$.

- Implementation: Neural network approximation for $\boldsymbol{u}$; flexible inner time discretization; stochastic optimization and random batch particle method

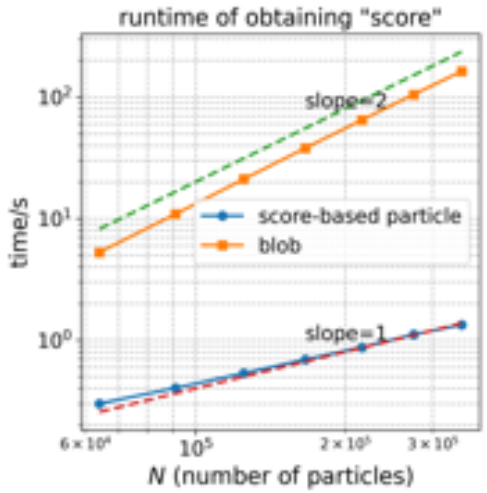
Numerical experiments—score



(a) Time evolution of the entropy decay rate.

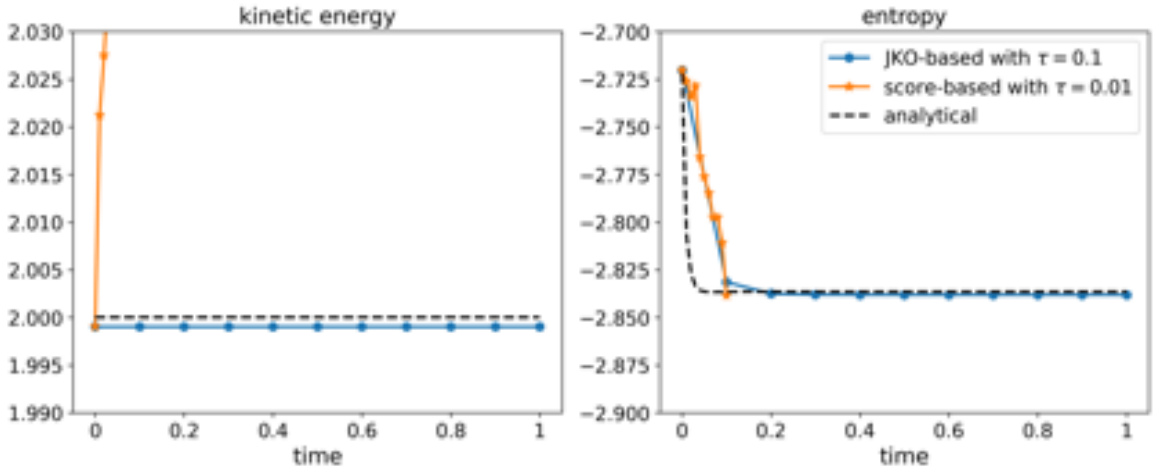


(b) Density visualization at particle locations.



(c) Computational time of obtaining "score" on GPU

Numerical experiments–JKO



(a) Comparison between JKO-based and score-based particle methods with varying time step sizes in strong collision regime.

References:
[1] Y. Huang and L. Wang, A score-based particle method for homogeneous Landau equation, Journal of Computational Physics, (2025), p. 114053.
[2] Y. Huang and L. Wang, JKO for Landau: a variational particle method for homogeneous Landau equation, https://arxiv.org/abs/2409.12296.

Phase Transitions and Linear Stability for the mean-field Kuramoto-Daido Model

Kyunghoo Mun
Joint work with Matthew Rosenzweig

Carnegie Mellon University

July 24, 2025

Mean-field interacting particle systems

We consider a system of N particles on the one-dimensional torus $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ by,

$$d\theta_i = \frac{K}{N} \sum_{j=1}^{N} \nabla W(\theta_i - \theta_j)\, dt + dB_i, \quad 1 \leq i \leq N,$$

where W is a $\mathbb{T}$-periodic interaction function, $(B_i)_{i=1}^N$ are independent Brownian motions, and $K > 0$ is the interaction strength parameter.

Statistical perspective

We aim to understand the behavior of this interacting particle systems where $N \gg 1$. For example, when modeling particles in gas, we have $N \approx 10^{23}$ (Avogadro's number). Due to the large system size N, its nonlinearity, and the coupling between particles, tracking all individual trajectories become computationally infeasible.

Key idea

Analyze the **approximate dynamics** of a single particle from a statistical point of view.

We study the limiting behavior of the empirical measure

$$d\nu_{N,t} = \frac{1}{N}\sum_{i=1}^{N} d\delta_{\theta_i(t)},$$

as $N \to \infty$.

Mean-field equation

For a smooth interaction $W \in \mathcal{C}^\infty(\mathbb{T})$, we obtain the mean-field limit

$$\nu_{N,t} \to \nu_t, \quad \text{weakly as } N \to \infty,$$

where ν_t is absolutely continuous with a smooth density $q_t(\theta)$, which solves the McKean-Vlasov equation (mean-field equation)

$$\partial_t q_t = \frac{1}{2}\Delta q_t - K\,\nabla\Big(q_t\,\nabla(W * q_t)\Big).$$

Linearized McKean-Vlasov equation

Suppose $q \in \mathcal{C}^2(\mathbb{T})$ is a stationary solution of the McKean-Vlasov equation, *i.e.*

$$0 = \frac{1}{2} q'' - K\Big(q(W' * q)\Big)'.$$

Goal

Understand in which situation a solution converges to a stationary state q.

The first step is to study dynamics when the initial condition is a perturbation of q. Consider the evolution of the perturbation $u_t(\theta) = q_t(\theta) - q(\theta)$. Then it follows

$$\partial_t u_t(\theta) = L_q u_t(\theta) - K\Big(u_t(W' * u_t)\Big)'.$$

Here, L_q is the linear operator defined by

$$L_q u = \frac{1}{2} u'' - K\Big(q(W' * u) + u(W' * q)\Big)'.$$

When the initial perturbation u_0 is sufficiently small, then we may only consider the linear term in the equation as,

$$\partial_t u_t(\theta) = L_q u_t(\theta).$$

This is the **linearized McKean-Vlasov equation** at the stationary solution q, which is the primary object in this talk.

Free energy

The McKean-Vlasov equation can be understood as a gradient flow for the free energy functional $\mathcal{F}$ with respect to the Wasserstein-2 distance on the space of probability measures with finite second moments as,

$$\partial_t q_t = \nabla \left[q_t \, \nabla \Big(\frac{\delta \mathcal{F}(q_t)}{\delta q_t} \Big) \right],$$

where $\delta \mathcal{F} / \delta q$ is L^2 Fréchet derivative of the functional $\mathcal{F}$. The free energy $\mathcal{F}$ is defined by

$$\mathcal{F}(q) = \frac{1}{2} \int q \log q \; d\theta - \frac{K}{2} \int q (W * q) \; d\theta.$$

Therefore, a critical point of $\mathcal{F}$ is equivalent to a stationary solution of the McKean-Vlasov equation.

Phase transition

For smooth attractive interaction W, a *phase transition* occurs, *i.e.* there exists a critical interaction strength $K_c > 0$ such that

1. For $K < K_c$, the uniform distribution $q_{\text{unif}} = (2\pi)^{-1}$ is the unique minimizer of the free energy $\mathcal{F}$.
2. For $K > K_c$, there exists a non-uniform minimizer $q \neq q_{\text{unif}}$ of $\mathcal{F}$.

Remark. By the continuity of $\mathcal{F}$, the uniform distribution q_{unif} remains a minimizer when $K = K_c$.

Kuramoto model

The Kuramoto model corresponds to the the interaction potential

$$W(\theta) = \cos\theta.$$

Let $q(\theta)$ be a stationary solution of the McKean-Vlasov equation. Defining its first Fourier coefficient as $r = \int q(\varphi)\cos\varphi \; d\varphi$, the stationary solution takes the form

$$q(\theta) = \frac{\exp(2Kr\cos\theta)}{\int \exp(2Kr\cos\theta) \; d\theta},$$

where $r = r(K)$ solves the self-consistency equation

$$r = \frac{\int \cos\theta \exp(2Kr\cos\theta) \; d\theta}{\int \exp(2Kr\cos\theta) \; d\theta}.$$

Equivalently, r is a critical point of (parameterized) free energy functional

$$F_K(r) = Kr^2 - \log\left(\frac{1}{2\pi}\int \exp(2Kr\cos\theta) \; d\theta\right)$$

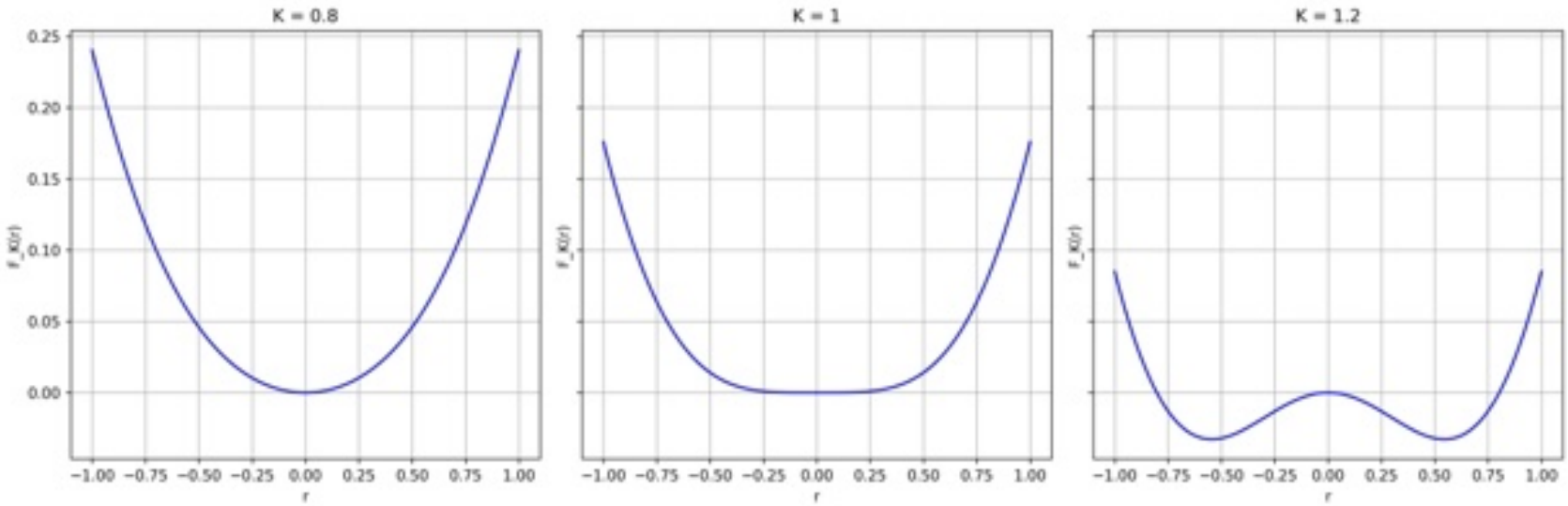


Figure: Parameterized free energy $F_K(r)$ for different $K \in \{0.8, 1, 1.2\}$.

Kuramoto-Daido model

Kuramoto-Daido model was introduced by Daido (1995); Daido (1996), with the interaction potential

$$W(\theta) = \cos\theta + m\cos 2\theta, \quad m > 0.$$

Daido proposed the model to study multibranch synchronization phenomena, which involve interactions through multiple Fourier modes. Through numerical simulations and heuristic arguments, he demonstrated the occurrence of phase transitions. Since then, the Kuramoto-Daido model has been studied in physics literature Hansel-Mato-Meunier (1993); Komarov-Pikovsky (2014); Cobero-Politi-Rosenblum (2016).

Remark. For $m < 0$, this model has $K_c = 1$, as in the Kuramoto model, analyzed by Vukadinovic (2023).

Phase Transition in the uniqueness of the stationary solution

There exists a uniform constant $m_* \in (1, 2)$ such that

1. For $m \in [0, 1/2]$, $K_c = 1$ and the transition is continuous.
2. For $m \in (1/2, 1]$, $K_c < 1$ and the transition is discontinuous.
3. For $m \in (1, m_*)$, $K_c < m^{-1}$ and the transition is discontinuous.
4. For $m = m_*$, $K_c = m^{-1}$ and the transition is discontinuous.
5. For $m \in (m_*, \infty)$, $K_c = m^{-1}$ and the transition is continuous.

Linear stability at the non-uniform stationary solution q

Assume $K > 1$ and $m \in (0, 8.568 \times 10^{-4}]$. Then the spectrum of linearized McKean-Vlasov operator L_q consists solely of pure points lying in $(-\infty, 0]$, and it includes the eigenvalue 0 whose one-dimensional eigenspace is spanned by q'. Moreover, the spectral gap is bounded below by

$$\mathrm{gap}(L_q) \geq C,$$

where the positive constant C is a function of K, m.

Thank you for your attention.