ENTROPY-DISSIPATION INFORMED NEURAL NETWORK FOR MCKEAN-VLASOV TYPE PDES

Zebang Shen*
ETH Zürich
zebang.shen@inf.ethz.ch

Zhenfu Wang
Peking University
zwang@bicmr.pku.edu.cn

ABSTRACT

We extend the concept of self-consistency for the Fokker-Planck equation (FPE) [Shen et al., 2022] to the more general McKean-Vlasov equation (MVE). While FPE describes the macroscopic behavior of particles under drift and diffusion, MVE accounts for the additional inter-particle interactions, which are often highly singular in physical systems. Two important examples considered in this paper are the MVE with Coulomb interactions and the vorticity formulation of the 2D Navier-Stokes equation. We show that a generalized self-consistency potential controls the KL-divergence between a hypothesis solution to the ground truth, through entropy dissipation. Built on this result, we propose to solve the MVEs by minimizing this potential function, while utilizing the neural networks for function approximation. We validate the empirical performance of our approach by comparing with state-of-the-art NN-based PDE solvers on several example problems.

1 Introduction

Researchers are utilizing AI to push the boundaries of science research, following their success in the fields of computer vision and natural language processing. For centuries, scientists have discovered natural laws through real-world observations, using these laws to describe and predict the dynamics of real-world systems through Partial Differential Equations (PDEs). As PDEs are of fundamental importance, a growing area in machine learning is the use of neural networks (NN) to solve these equations [Han et al., 2017, 2018, Zhang et al., 2018, Wang et al., 2018, Raissi et al., 2020, Cai et al., 2021, Karniadakis et al., 2021, Li et al., 2021, Cuomo et al., 2022].

An important category of PDEs is the McKean-Vlasov equation (MVE), which models the dynamics of a stochastic particle system with mean-field interactions

$$\mathrm{d}\mathbf{X}_t = -\nabla V(\mathbf{X}_t)\mathrm{d}t + K * \bar{\rho}_t(\mathbf{X}_t)\mathrm{d}t + \sqrt{2\nu}\mathrm{d}\mathbf{B}_t, \quad \bar{\rho}_t = \mathrm{Law}(\mathbf{X}_t). \tag{1}$$

Here $\mathbf{X}_t \in \mathcal{X}$ denotes the (phase) space position of a random particle, $\mathcal{X}$ is either $\mathbb{R}^d$ or the torus $\Pi^d = [-L, L]^d$ (a cube with periodic boundary condition), $V : \mathbb{R}^d \to \mathbb{R}$ denotes a *known* exterior potential, $K : \mathbb{R}^d \to \mathbb{R}^d$ denotes some interaction kernel function and the convolution operation $*$ is defined as $h * \phi = \int_{\mathcal{X}} h(\boldsymbol{x} - \boldsymbol{y})\phi(\boldsymbol{y})d\boldsymbol{y}$, $\{\mathbf{B}_t\}_{t\geq 0}$ is the standard d-dimensional Wiener process with $\nu \geq 0$ being some diffusion coefficient, and $\bar{\rho}_t : \mathcal{X} \to \mathbb{R}$ is the law or the probability density function of the random variable $\mathbf{X}_t$ and the initial data $\bar{\rho}_0$ is given.

Under mild regularity conditions, the density function $\bar{\rho}_t$ satisfies the MVE

$$\partial_t \bar{\rho}_t + \ \mathrm{div}\left(\bar{\rho}_t(-\nabla V + K * \bar{\rho}_t)\right) = \nu\Delta\bar{\rho}_t, \tag{2}$$

where div denotes the divergence operator, i.e. $\mathrm{div}\ h(\boldsymbol{x}) = \sum_{i=1}^d \frac{\partial h_i}{\partial x_i}$ for a velocity field $h : \mathbb{R}^d \to \mathbb{R}^d$, Δ denotes the Laplacian operator on the spatial variables defined as $\Delta\phi = \mathrm{div}(\nabla\phi)$, where $\nabla\phi$ denotes the gradient of a scalar function $\phi : \mathbb{R}^d \to \mathbb{R}$. Note that all these operators are applied only on the spatial variable $\boldsymbol{x}$.

It is important to note that in order to accurately describe dynamics in real-world phenomena such as electromagnetism [Golse, 2016] and fluid mechanics [Majda et al., 2002], the interaction kernels K in the MVE are usually chosen to be

*Authors are listed in an alphabetic order.

highly *singular*. Two of the most notable examples are the MVE with Coulomb interactions where

$$K(\boldsymbol{x}) = -\nabla g(\boldsymbol{x}), \text{ with } g(\boldsymbol{x}) = \begin{cases} ((d-2)S_{d-1}(1))^{-1} \|\boldsymbol{x}\|^{-(d-2)}, & d \geq 3, \\ -(2\pi)^{-1} \log \|\boldsymbol{x}\|, & d = 2, \end{cases} \tag{3}$$

with $S_{d-1}(1)$ denoting the surface area of the unit sphere in $\mathbb{R}^d$, and the vorticity formulation of the 2D Navier-Stokes equation ($d = 2$) where the interaction kernel K is given by the Biot-Savart law

$$K(\boldsymbol{x}) = \frac{1}{2\pi} \frac{\boldsymbol{x}^\perp}{\|\boldsymbol{x}\|^2} = \frac{1}{2\pi} \left(-\frac{x_2}{\|\boldsymbol{x}\|^2}, \frac{x_1}{\|\boldsymbol{x}\|^2} \right), \tag{4}$$

where $\boldsymbol{x} = (x_1, x_2)$ and $\|\boldsymbol{x}\|$ denotes the Euclidean norm of a vector. These choices of K are called singular since $\|K(\boldsymbol{x})\| \to \infty$ when $\|\boldsymbol{x}\| \to 0$. The presence of singularity in the interaction kernel poses significant difficulties in solving the MVE using NN-based approaches. Furthermore, there is a significant lack of theoretical guarantees in the literature for NN-based algorithms. Specifically, prior to our research, there was no established method for determining the quality of a candidate solution through an easily computable quantity.

By noting $\Delta \bar{\rho}_t = \operatorname{div}(\bar{\rho}_t \nabla \log \bar{\rho}_t)$, we can rewrite the MVE in the form of a continuity equation

$$\partial_t \bar{\rho}_t + \operatorname{div}\Big(\bar{\rho}_t \Big(-\nabla V + K * \rho_t - \nu \nabla \log \bar{\rho}_t \Big) \Big) = 0, \tag{5}$$

and for simplicity throughout this paper we will denote $\mathcal{A}[\bar{\rho}_t](x) = -\nabla V + K * \bar{\rho}_t - \nu \nabla \log \bar{\rho}_t$ and refer to $\mathcal{A}[\bar{\rho}_t]$ as the *underlying velocity*. Consider another time-varying *hypothesis velocity* field $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$. The goal of this paper is to show that, given some time interval $[0, T]$ with $T > 0$, when $f(t, \cdot)$ is sufficiently close to $\mathcal{A}[\bar{\rho}_t]$ in an appropriate sense, one can recover $\{\bar{\rho}_t\}_{t \in [0,T]}$, i.e. the solution to the MVE in the time interval $[0, T]$, by solving the following continuity equation

$$\partial_t \rho_t^f + \nabla \cdot (\rho_t^f \, f) = 0, \tag{6}$$

for $t \in [0, T]$ with $\rho_0^f = \bar{\rho}_0$ (recall that we assume the initial law $\bar{\rho}_0$ to be given). The superscript in ρ_t^f is to emphasize its dependence on the hypothesis velocity field f and we will refer to ρ_t^f as the *hypothesis solution*. Our results can be informally summarized as follows.

Theorem 1. *Suppose that the initial density function $\bar{\rho}_0$ is sufficiently regular and the hypothesis velocity field $f_t(\cdot) = f(t, \cdot)$ is at least three-times continuously differentiable both in t and x. Define the* self-consistency *potential/loss function to be*

$$R(f) = \int_0^T \left\| f_t - (-\nabla V + K * \rho_t^f - \nu \nabla \log \rho_t^f) \right\|_{\rho_t^f}^2 \mathrm{d}t. \tag{7}$$

Here $\|h\|_\phi$ denotes the weighted $\mathcal{L}^2$ norm for the function h, $\|h\|_\phi^2 = \int \|h(\boldsymbol{x})\|^2 \phi(\boldsymbol{x}) d\boldsymbol{x}$. We have for the MVE with the Coulomb interactions (3) or the Biot-Savart law (4), the KL divergence between the hypothesis solution ρ_t^f and the ground truth solution $\bar{\rho}_t$ is controlled by the self-consistency potential for any $t \in [0, T]$, i.e. for some constant $C > 0$,

$$\sup_{t \in [0,T]} \mathbf{KL}(\rho_t^f, \rho_t) \leq C R(f). \tag{8}$$

Our work generalizes the result in [Shen et al., 2022], where the authors consider the much simpler Fokker-Planck equation, an specific instance of our work (simply take $K \equiv 0$). Consequently, as a by-product, we improve the Wasserstein-type bound therein to the KL type bound, and remove the requirement of the complicated Sobolev norm in the definition of self-consistency therein. Shen et al. [2022] further propose that the self-consistency of a PDE can be used to test the quality of an existing solution and to guide the design of some objective functions for a neural network parameterization of the PDE solution. Inspired by their work, we generalize their framework to the more complicated MVE and provide algorithms that can solve the MVE with singular Coulomb interaction and the 2D Navier-Stokes equation.

Contributions In summary, we propose a novel neural network (NN) based MVE solver by exploiting the stability property of the corresponding system via entropy dissipation, which admits rigorous theoretical guarantees even in the presence of singular interaction kernels. We elaborate the contributions of our work from theory, algorithm and empirical perspectives as follows.

1. (Theory-wise) By studying the stability of the MVE with singular interaction kernels in the Coulomb case (3) and the 2D Navier-Stokes equation (the Biot-Savart case (4)) via entropy dissipation, we establish the self-consistency of the MVE. Specifically, we design a potential function $R(f)$ of a hypothesis velocity f such that $R(f)$ controls the KL divergence between the hypothesis solution ρ_t^f (defined in equation 6) and the ground truth solution $\bar{\rho}_t$ for any time stamp within a given time interval $[0, T]$. A direct consequence of this result is that ρ_t^f exactly recovers $\bar{\rho}_t$ in the KL sense given that $R(f) = 0$.

2. (Algorithm-wise) When the hypothesis velocity field is parameterized by a neural network (NN), i.e. $f = f_\theta$ with θ being some finite-dimensional parameters, the self-consistency potential $R(f_\theta)$ can be used as the loss function of the neural network parameters θ. We discuss in details on how an estimator of the gradient $\nabla_\theta R(f_\theta)$ can be computed, so that gradient-based optimizers can be utilized to train the NN. In particular, for the 2D Navier-Stokes equation (the Biot-Savart case (4)), we show that the singularity in the gradient computation can be removed by exploiting an anti-derivative of the Biot-Savart kernel.

3. (Empirical-wise) We compare the proposed approach, derived from our novel theoretical guarantees, with SOTA NN-based algorithms for solving MVE with the Coulomb interaction and the 2D Navier-Stokes equation (the Biot-Savart interaction). We pick specific instances of the initial density $\bar{\rho}_0$, under which explicit solutions are known and can be used as the ground truth to test the quality of the hypothesis ones. Using NNs with the same complexity (depth, width, and structure), we observe that the proposed method significantly outperforms the included baselines.

2 Self-consistency of the McKean-Vlasov Equation

In this section, we present the generalized self-consistency potential for the MVE. To understand the intuition behind our design, we first write the continuity equation 6 in a similar form as the MVE:

$$\partial_t \rho_t^f + \operatorname{div}\Big(\rho_t^f\Big(- \nabla V + K * \rho_t^f - \nu \log \rho_t^f + \delta_t\Big)\Big) = 0, \tag{9}$$

where f is the hypothesis velocity (recall that $f_t(\cdot) = f(t, \cdot)$) and

$$\delta_t = f_t - (-\nabla V + K * \rho_t^f - \nu \log \rho_t^f), \tag{10}$$

can be regarded as a perturbation to the original MVE system. Taking this perturbation perspective, it is natural to study the deviation of the hypothesis solution ρ_t^f from the true solution $\bar{\rho}_t$ using an appropriate Lyapunov function $L(\rho_t^f, \bar{\rho}_t)$. Clearly this deviation will depend on the perturbation δ_t, and such a dependence is often termed as the *stability* of the underlying dynamical system. Moreover, the aforementioned relation between the perturbation and the deviation allows us to derive the self-consistency potential of the MVE.

Following this idea, the design of the self-consistency potential can be determined by the choice of the Lyapunov function L used in the stability analysis. In the following, we describe the Lyapunov function used for the MVE with the Coulomb interaction and the vorticity formulation of the 2D Navier-Stokes equation (MVE with Biot-Savart interaction). The proof of the following results are the major theoretical contributions of this paper and will be elaborated in the analysis section 3.

- For the MVE with the Coulomb interaction, we choose L to be the *modulated free energy* (defined in equation 26) which is originally proposed in [Bresch et al., 2019a] to establish the mean-field limit of a corresponding interacting particle system. We have (setting $L = E$)

$$\frac{\mathrm{d}}{\mathrm{d}t} E(\rho_t^f, \bar{\rho}_t) \le \frac{1}{2} \int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 \mathrm{d}x + C E(\rho_t^f, \bar{\rho}_t), \tag{11}$$

  where C is a universal constant depending on ν and $(\bar{\rho}_t)_{t\in[0,T]}$.

- For the 2D Navier-Stokes equation (MVE with the Biot-Savart interaction), we choose L to be the KL divergence. Our analysis is inspired by [Jabin and Wang, 2018] which for the first time establishes the quantitative mean-field limit of the stochastic interacting particle systems where the interaction kernel can be in some negative Sobolev space. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{KL}(\rho_t^f, \bar{\rho}_t) \le -\frac{\nu}{2} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + C \mathbf{KL}(\rho_t^f, \bar{\rho}_t) + \frac{1}{\nu} \int \rho_t^f |\delta_t|^2, \tag{12}$$

  where again C is a universal constant depending on ν and $(\bar{\rho}_t)_{t\in[0,T]}$.

From the above discussion, we can see that the self-consistency potential 7 is exactly the term derived by stability analysis of the MVE system with an appropariate Lyapunov function, after applying the Grönwall's inequality. However, the potential function 7 remains elusive from a computational perspective. Moreover, when utilized as the objective loss for training a parameterized hypothesis velocity field, we must be able to compute the gradient w.r.t. the parameters, so that gradient-based optimizer can be utilized. This is elaborated in the next section.

2.1 Stochastic Gradient Computation with Neural Network Parameterization

While the choice of self-consistency potential 7 is theoretically justified through the above stability study, in this section we show that it admits an estimator which can be efficiently computed. Given an initial data point $\boldsymbol{x}_0$, define the trajectory $\{\boldsymbol{x}(t)\}_{t=0}^T$ via the initial value problem $\frac{d\boldsymbol{x}(t)}{dt} = f_t(\boldsymbol{x}(t);\theta)$, $\boldsymbol{x}(0) = \boldsymbol{x}_0$ (suppose that f_t is Lipschitz continuous for all $t \in [0,T]$ so the trajectory exists and is unique). Define the map X_t such that $\boldsymbol{x}(t) = X_t(\boldsymbol{x}_0)$. From the definition of the push-forward measure, one has $\rho_t^f = X_t \sharp \bar{\rho}_0$, where ρ_t^t is defined in equation 6. Recall the definitions of the potential $R(f)$ in equation 7 and the perturbation δ_t in equation 10. Use the change of variable formula of the push-forward measure in (a) and the Fubini's theorem in (b). We have

$$R(f) = \int_0^T \|\delta_t\|^2_{\rho_t^f} dt \overset{(a)}{=} \int_0^T \|\delta_t \circ X_t\|^2_{\bar{\rho}_0} dt \overset{(b)}{=} \int \int_0^T \|\delta_t \circ X_t(\boldsymbol{x}_0)\|^2 dt d\bar{\rho}_0(\boldsymbol{x}_0). \tag{13}$$

Consequently, by defining the trajectory-wise loss

$$R(f;\boldsymbol{x}_0) = \int_0^T \|\delta_t \circ X_t(\boldsymbol{x}_0)\|^2 dt, \tag{14}$$

we can write the potential function 7 as an expectation $R(f) = \mathbb{E}_{\boldsymbol{x}_0 \sim \bar{\rho}_0}[R(f;\boldsymbol{x}_0)]$. Suppose that the hypothesis velocity field is parameterized by a neural network $f = f_\theta$. Using the above expectation formulation, we obtain an unbiased estimation of $\nabla_\theta R(f_\theta)$ via the Monte-Carlo integration w.r.t. $\boldsymbol{x}_0 \sim \bar{\rho}_0$, given that we can compute $\nabla_\theta R(f_\theta;\boldsymbol{x}_0)$.

We show $\nabla_\theta R(f_\theta;\boldsymbol{x}_0)$ can be computed, at least up to a high accuracy, via the adjoint method (for completeness see the derivation of the adjoint method in appendix B). As a recap, suppose that we can write $R(f_\theta;x_0)$ in a standard ODE-constrained form

$$R(f_\theta;\boldsymbol{x}_0) = \ell(\theta) \ = \int_0^T g(t,\boldsymbol{s}(t),\theta)dt, \tag{15}$$

where $\{\boldsymbol{s}(t)\}_{t\in[0,T]}$ is the solution to the initial value problem $\frac{d}{dt}\boldsymbol{s}(t) = \psi(t,\boldsymbol{s}(t);\theta)$ with $\boldsymbol{s}(0) = \boldsymbol{s}_0$, and ψ is a known transition function. The adjoint method states that the gradient $\frac{d}{d\theta}\ell(\theta)$ can be computed as[2]

$$\frac{d\ell}{d\theta} = \mathbb{E}_{t\sim\mathrm{Uniform}[0,T]}\left[a(t)^\top \frac{\partial\psi}{\partial\theta}(t,\boldsymbol{s}(t);\theta) + \frac{\partial g}{\partial\theta}(t,\boldsymbol{s}(t);\theta)\right]. \tag{16}$$

where $a(t)$ is solution to the final value problems $\frac{d}{dt}a(t)^\top + a(t)^\top \frac{\partial\psi}{\partial s}(t,\boldsymbol{s}(t);\theta) + \frac{\partial g}{\partial s}(t,\boldsymbol{s}(t);\theta) = 0, a(T) = 0$. In the following, we focus on how $R(f_\theta;\boldsymbol{x}_0)$ can be written in the above ODE-constrained form.

Write $R(f_\theta;\boldsymbol{x}_0)$ in ODE-constrained Form Expanding the definition of δ_t in equation 10 gives

$$\delta_t \circ X_t(\boldsymbol{x}_0) = \delta_t(\boldsymbol{x}(t)) = f_t(\boldsymbol{x}(t)) - \left(-\nabla V(\boldsymbol{x}(t)) + K * \rho_t^f(\boldsymbol{x}(t)) - \nu\nabla\log\rho_t^f(\boldsymbol{x}(t))\right). \tag{17}$$

Note that in the above quantity, f and V are known functions. Moreover, it is known that $\nabla\log\rho_t^f(\boldsymbol{x}(t))$ admits a closed form dynamics (e.g. see Proposition 2 in [Shen et al., 2022])

$$\frac{d}{dt}\nabla\log\rho_t^f(\boldsymbol{x}(t)) = -\nabla\left(\nabla\cdot f_t(\boldsymbol{x}(t);\theta)\right) - \left(\mathcal{J}_{f_t}(\boldsymbol{x}(t);\theta)\right)^\top \nabla\log\rho_t^f(\boldsymbol{x}(t)), \tag{18}$$

which allows it to be explicitly computed by starting from $\nabla\log\bar{\rho}_0(x_0)$ and integrating over time (recall that $\bar{\rho}_0$ is known). Here $\mathcal{J}_{f_t}$ denotes the Jacobian matrix of f_t. Consequently, all we need to handle is the convolution term $K * \rho_t^f(\boldsymbol{x}(t))$, which in general cannot be exactly computed since it depends on the global configuration of the hypothesis distribution ρ_t^f.

A common choice to avoid the difficulty of the convolution operation is via empirical approximation: Let $\{\boldsymbol{y}_i(t)\}_{i=1}^N$ be a batch of i.i.d. samples distributed according to ρ_t^f and denote an empirical approximation of ρ_t^f by $\mu_N^{\rho_t^f} = \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{y}_i(t)}$, where $\delta_{\boldsymbol{y}_i(t)}$ denotes the Dirac measure at $\boldsymbol{y}_i(t)$. We approximate the convolution term in equation 17 in different ways for the Coulomb and the Biot-Savart interactions:

[2]This implies that we can obtain an unbiased estimator of $\frac{d\ell}{d\theta}$ by first sample $t \sim \mathrm{Uniform}[0,T]$ and simply compute $T * \left(a(t)^\top \frac{\partial\psi}{\partial\theta}(t,\boldsymbol{s}(t);\theta) + \frac{\partial g}{\partial\theta}(t,\boldsymbol{s}(t);\theta)\right)$. In practice, we sample multiple t uniformly from $[0,T]$ to reduce the variance. Note that to obtain $\boldsymbol{s}(t)$ and $\boldsymbol{a}(t)$ for all sampled time stamp t, we still need to solve the ODEs involved in the definition of $\boldsymbol{x}(t)$ and $\boldsymbol{a}(t)$ once, however their dimensions are now independent of the size of the neural network.

1. For the Coulomb interaction, we directly approximate the convolution term in equation 17 by $K * \mu_N^{\rho_t^f}(\boldsymbol{x}(t)) = \frac{1}{N}\sum_{i=1}^N K(\boldsymbol{x}(t) - \boldsymbol{y}_i(t))$. In practice, we choose N sufficiently large so that the above empirical approximation is accurate. Indeed, at least for the whole space case, i.e. the underlying space $\mathcal{X}$ is $\mathbb{R}^d$, one has that

$$\int_{\mathbb{R}^d} |K * \mu_N^{\rho^f}(x) - K * \rho^f(x)|^2 \mathrm{d}x = \int_{x\neq y} g(x-y)\mathrm{d}(\mu_N^{\rho^f} - \rho^f)^{\otimes 2}(x,y) = F(\mu_N^{\rho^f}, \rho^f),$$

where $F(\mu_N^{\rho^f}, \rho^f)$ is the modulated (interaction) energy defined as in Serfaty [2020]. We expect $F(\mu_N^{\rho^f}, \rho^f) \to 0$ almost surely as $N \to \infty$.

2. For Biot-Savart interaction (2D Navier-Stokes equation), there are more structure to exploit and we can completely avoid the singularity: As noted by Jabin and Wang [2018], the convolution kernel K can be written in a divergence form:

$$K = \nabla \cdot U, \text{ with } U(\boldsymbol{x}) = \frac{1}{2\pi}\begin{bmatrix} -\arctan(\frac{\boldsymbol{x}_1}{\boldsymbol{x}_2}), & 0 \\ 0, & \arctan(\frac{\boldsymbol{x}_2}{\boldsymbol{x}_1}) \end{bmatrix}, \tag{19}$$

where the divergence of a matrix function is applied row-wisely, i.e. $[K(\boldsymbol{x})]_i = \text{ div } U_i(\boldsymbol{x})$. Using integration by parts, one has (assuming that the boundary integration vanishes, e.g. when the underlying space is a torus)

$$\begin{aligned} K * \rho_t^f(\boldsymbol{x}) &= \int K(\boldsymbol{y})\rho_t^f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} = \int \nabla\cdot U(\boldsymbol{y})\rho_t^f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} = \int U(\boldsymbol{y})\nabla\rho_t^f(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{y} \\ &= \int U(\boldsymbol{x}-\boldsymbol{y})\rho_t^f(\boldsymbol{y})\nabla\log\rho_t^f(\boldsymbol{y})d\boldsymbol{y} = \mathbb{E}_{\boldsymbol{y}\sim\rho_t^f(\boldsymbol{y})}[U(\boldsymbol{x}-\boldsymbol{y})\nabla\log\rho_t^f(\boldsymbol{y})]. \end{aligned}$$

If the score function $\nabla\log\rho_t^f$ is bounded, then the integrand in the expectation is also bounded. Therefore, we can avoid integrating singular functions and the Monte Carlo-type estimation $\frac{1}{N}\sum_{i=1}^N U(\boldsymbol{x}-\boldsymbol{y}_i(t))\nabla\log\rho_t^f(\boldsymbol{y}_i(t))$ is accurate for a sufficiently large value of N.

With the above discussion, we are now ready to write $R(f_\theta; \boldsymbol{x}_0)$ in an ODE-constrained form. Define the state $\boldsymbol{s}(t)$, the initial condition $\boldsymbol{s}_0$ and the transition function ψ as follows: Let

$$\boldsymbol{s}(t) = \left[\boldsymbol{x}(t), \xi(t), \{\boldsymbol{y}_i(t)\}_{i=1}^N, \{\zeta_i(t)\}_{i=1}^N\right], \tag{20}$$

with $\xi(t) = \nabla\log\rho_t^f(\boldsymbol{x}(t))$ and $\zeta_i(t) = \nabla\log\rho_t^f(\boldsymbol{y}_i(t))$. Take the initial condition

$$\boldsymbol{s}_0 = \left[\boldsymbol{x}_0, \xi_0, \{\boldsymbol{y}_i(0)\}_{i=1}^N, \{\zeta_i(0)\}_{i=1}^N\right] \tag{21}$$

with $\xi_0 = \nabla\log\bar{\rho}_0(\boldsymbol{x}_0)$, $\zeta_i(0) = \nabla\log\bar{\rho}_0(\boldsymbol{y}_i(0))$, and $\boldsymbol{y}_i(0) \overset{iid}{\sim} \bar{\rho}_0$; and define the function

$$\psi(t, s(t); \theta) = [f_t(\boldsymbol{x}(t);\theta), h_t(\boldsymbol{x}(t), \xi(t);\theta), \{f_t(\boldsymbol{y}_i(t);\theta)\}_{i=1}^N, \{h_t(\boldsymbol{y}_i(t), \zeta_i(t);\theta)\}_{i=1}^N], \tag{22}$$

where $h(\boldsymbol{a}, \boldsymbol{b}; \theta) = -\nabla\left(\nabla\cdot f_t(\boldsymbol{a};\theta)\right) - \mathcal{J}_{f_t}^\top(\boldsymbol{a};\theta)\boldsymbol{b}$ (derived from equation 18). Finally, define

$$g(t, \boldsymbol{s}(t); \theta) = \|f(t, \boldsymbol{x}(t);\theta) - (-\nabla V(\boldsymbol{x}(t)) + E(t, \boldsymbol{s}(t)) - \nu\xi(t))\|^2, \tag{23}$$

where the estimator $E(t, \boldsymbol{s})$ of the convolution term is defined as

$$E(t, \boldsymbol{s}(t)) = \begin{cases} \frac{1}{N}\sum_{i=1}^N K(\boldsymbol{x}(t) - \boldsymbol{y}_i(t)) & \text{the Coulomb case,} \\ \frac{1}{N}\sum_{i=1}^N U(\boldsymbol{x} - \boldsymbol{y}_i(t))\zeta_i(t) & \text{the Biot-Savart case.} \end{cases} \tag{24}$$

We recall the definition of U in equation 19.

3 Analysis

We start by defining the KL divergence / relative entropy, modulated (interaction) energy, and modulated free energy for two probability densities $\rho, \bar{\rho}$. Again $\mathcal{X}$ denotes the underlying space which can be the whole space $\mathbb{R}^d$ or the torus Π^d which can be identified with $[-L, L]^d$ with the periodic boundary condition and $L > 0$. Firstly, the KL divergence is defined as $\mathbf{KL}(\rho, \bar{\rho}) = \int_{\mathcal{X}} \rho(x)\log\frac{\rho(x)}{\bar{\rho}(x)}\mathrm{d}x$ if ρ is absolutely continuous with respect to $\bar{\rho}$, otherwise we set $\mathbf{KL}(\rho, \bar{\rho}) = +\infty$. By Jensen's inequality, $\mathbf{KL}(\rho, \bar{\rho}) \geq 0$ for any $\rho, \bar{\rho}$. Given the McKean-Vlasov equation 2, if K is bounded, it is sufficiently to choose the Lyapunov functional $L(\rho_t^f, \bar{\rho}_t)$ as the KL divergence (please see Theorem 5 in the appendix). But for more singular kernels, we need also to consider the Coulomb type energy or modulated energy as in [Serfaty, 2020]

$$F(\rho, \bar{\rho}) = \frac{1}{2}\int_{\mathcal{X}^2} g(x-y)\mathrm{d}(\rho-\bar{\rho})(x)\mathrm{d}(\rho-\bar{\rho})(y), \tag{25}$$

where g is the fundamental solution to the Laplacian equation in $\mathbb{R}^d$, i.e. $-\Delta g = \delta_0$, and the Coulomb interaction reads $K = -\nabla g$ (see its closed form expression in equation 3). If we are only interested in the deterministic dynamics with Coulomb interactions, i.e. $\nu = 0$ in equation 2, it suffices to choose $L(\rho_t^f, \bar{\rho})$ as $F(\rho_t^f, \bar{\rho}_t)$ (please see Theorem 3). But if we consider the system with Coulomb interactions and diffusions, i.e. $\nu > 0$, we shall combine the KL divergence and the modulated energy to form the modulated free energy as in Bresch et al. [2019b], which reads

$$E(\rho, \bar{\rho}) = \nu \mathbf{KL}(\rho, \bar{\rho}) + F(\rho, \bar{\rho}). \tag{26}$$

This definition agrees with the physical meaning that "Free Energy = Temperature $\times$ Entropy + Energy", and we note that the temperature is proportional to the diffusion coefficient ν. We remark also for two probability densities ρ and $\bar{\rho}$, $F(\rho, \bar{\rho}) \geq 0$ since by looking in Fourier domain $F(\rho, \bar{\rho}) = \int \hat{g}(\xi) |\widehat{\rho - \bar{\rho}}(\xi)|^2 \mathrm{d}\xi \geq 0$ as $\hat{g}(\xi) \geq 0$. Moreover, $F(\rho, \bar{\rho})$ can be regarded as a negative Sobolev norm for $\rho - \bar{\rho}$, which is a metric.

To obtain our main stability estimate in the flavor of entropy dissipation, we first obtain the time evolution of the KL divergence as follows.

Lemma 1 (Time Evolution of the KL divergence)**.** *Given the hypothesis velocity field $f = f(t, x) \in C^1_{t,x}$. Assume that $(\rho_t^f)_{t\in[0,T]}$ and $(\bar{\rho}_t)_{t\in[0,T]}$ are classical solutions to equation 6 and equation 5 respectively. It holds that (recall the definition of δ_t in equation 10)*

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathcal{X}} \rho_t^f \log \frac{\rho_t^f}{\bar{\rho}_t} = -\nu \int_{\mathcal{X}} \rho_t^f |\nabla \log \frac{\rho_t^f}{\bar{\rho}_t}|^2 + \int_{\mathcal{X}} \rho_t^f K * (\rho_t^f - \bar{\rho}_t) \cdot \nabla \log \frac{\rho_t^f}{\bar{\rho}_t} + \int_{\mathcal{X}} \rho_t^f \delta_t \cdot \nabla \log \frac{\rho_t^f}{\bar{\rho}_t},$$

where $\mathcal{X}$ denotes either $\mathbb{R}^d$ or the tours Π^d.

We refer the proof of this lemma and all other lemmas and theorems in this section to the appendix C. We remark that to have the existence of classical solution $(\bar{\rho}_t)_{t\in[0,T]}$, we definitely need the regularity assumptions on $-\nabla V$ and on K. But the linear term $-\nabla V$ will not contribute to the evolution of the relative entropy. See [Jabin and Wang, 2018] for detailed discussions.

Similarly, we have the time evolution of the modulated energy as follows.

Lemma 2 (Time evolution of the modulated energy)**.** *Under the same assumptions as in Lemma 1, given the diffusion coefficient $\nu \geq 0$, it holds that (recall the definition of δ_t in equation 10)*

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} F(\rho_t^f, \bar{\rho}_t) = &- \int_{\mathcal{X}} \rho_t^f |K * (\rho_t^f - \bar{\rho})|^2 - \int_{\mathcal{X}} \rho_t^f \, \delta_t \cdot K * (\rho_t^f - \bar{\rho}_t) \\
&- \frac{1}{2} \int_{\mathcal{X}^2} K(x - y) \cdot \Big(\mathcal{A}[\bar{\rho}_t](x) - \mathcal{A}[\bar{\rho}_t](y) \Big) \mathrm{d}(\rho_t^f - \bar{\rho}_t)^{\otimes 2}(x, y) \\
&+ \nu \int_{\mathcal{X}} \rho_t^f \, K * (\rho_t^f - \bar{\rho}_t) \cdot \nabla \log \frac{\rho_t^f}{\bar{\rho}_t},
\end{aligned}$$

where we recall that $\mathcal{A}[\bar{\rho}_t](x) = -\nabla V(x) + K * \bar{\rho}_t - \nu \nabla \log \bar{\rho}_t$ is defined below equation 5.

By Lemma 1 and careful analysis, in particular by rewriting the Biot-Savart law in the divergence of a bounded matrix-valued function as in equation 19, we obtain the following estimate for the 2D Navier-Stokes case.

Theorem 2 (Stability Estimate of the 2D Navier-Stokes Equation)**.** *Consider the 2D Navier-Stokes equation in the vorticity formulation, i.e. equation 2 with $V = 0$. Notice that when K is the Biot-Savart kernel, $\mathrm{div} K = 0$. Assume that the initial data $\bar{\rho}_0 \in C^\infty(\Pi^d)$ and there exists $c > 1$ such that $\frac{1}{c} \leq \bar{\rho}_0 \leq c$. Assume further the hypothesis velocity field $f(t, x) \in C^1_{t,x}$. Then it holds that*

$$\sup_{t\in[0,T]} \int_{\Pi^d} \rho_t^f \log \frac{\rho_t^f}{\bar{\rho}_t} \mathrm{d}x \leq \frac{1}{\nu} \exp(MT) R(f),$$

where $M = \sup_{t\in[0,T]} M(t; \nu, U, \bar{\rho})$ and $M(t; \nu, U, \bar{\rho}_t) = \frac{2}{\nu} \|U\|^2_{L^\infty} \|\nabla \log \bar{\rho}_t\|^2_{L^\infty} + 4\|U\|_{L^\infty} \left\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \right\|_{L^\infty}$.

We give the complete proof and discussions for possible extension to long time estimate and to improve the dependence on T in the appendix C. This theorem tells us that as long as $R(f)$ is small, the KL divergence between ρ_t^f and $\bar{\rho}_t$ is small and the control is uniform in time $t \in [0, T]$. We state this theorem on torus for simplicity and one may expect similar result on $\mathbb{R}^d$. Also the C^∞ condition can be relaxed for instance to C^2.

To treat the McKean-Vlasov equation 2 with Coulomb interactions, we exploit the time evolution of the modulated free energy $E(\rho_t^f, \bar{\rho}_t)$. Indeed, combining Lemma 1 and Lemma 2, we arrive at the following estimate.

Lemma 3 (Time evolution of the modulated free energy)**.** *Under the same assumptions as in Lemma 1, one has (recall the definition of δ_t in equation 10)*

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\rho_t^f,\bar{\rho}_t) = -\int_{\mathcal{X}}\rho_t^f\Big|K*(\rho_t^f-\bar{\rho}_t)-\nu\nabla\log\frac{\rho_t^f}{\bar{\rho}_t}\Big|^2-\int_{\mathcal{X}}\rho_t^f\,\delta_t\cdot\Big(K*(\rho_t^f-\bar{\rho}_t)-\nu\nabla\log\frac{\rho_t^f}{\bar{\rho}_t}\Big)$$
$$-\frac{1}{2}\int_{\mathcal{X}^2}K(x-y)\cdot\Big(\mathcal{A}[\bar{\rho}_t](x)-\mathcal{A}[\bar{\rho}_t](y)\Big)\mathrm{d}(\rho_t^f-\bar{\rho}_t)^{\otimes 2}(x,y),$$

where $\mathcal{A}[\bar{\rho}_t](x) = -\nabla V(x) + K*\bar{\rho}_t(x) - \nu\nabla\log\bar{\rho}_t(x)$.

Inspired by the mean-field convergence results as in Serfaty [2020] and Bresch et al. [2019b], we finally can control the growth of $E(\rho_t^f,\bar{\rho}_t)$ in the case when $\nu>0$, and $F(\rho_t^f,\bar{\rho}_t)$ in the case when $\nu=0$. Note also that $E(\rho_t^f,\bar{\rho}_t)$ can also control the KL divergence when $\nu>0$.

Theorem 3 (Stability Estimate of McKean-Vlasov Eq. (2) with Coulomb interactions)**.** *Assume that for $t\in[0,T]$, the self-consistent velocity field $\mathcal{A}[\bar{\rho}_t](x)$ is Lipschitz in x and*

$$\sup_{t\in[0,T]}\|\nabla\mathcal{A}[\bar{\rho}_t](\cdot)\|_{L^\infty} = C_1 < \infty.$$

Then there exists $C>0$ such that

$$\sup_{t\in[0,T]}\nu\,\mathbf{KL}(\rho_t^f,\bar{\rho}_t) \le \sup_{t\in[0,T]}E(\rho_t^f,\bar{\rho}_t) \le \exp(CC_1T)R(f).$$

In the deterministic case when $\nu=0$, under the same assumptions, it holds that

$$\sup_{t\in[0,T]}F(\rho_t^f,\bar{\rho}_t) \le \exp(CC_1T)R(f).$$

See the proof and the discussion on the Lipschitz assumptions on $\mathcal{A}[\bar{\rho}_t](\cdot)$ in the appendix C.

4 Related Works

Solving partial differential equations (PDEs) is a key aspect of scientific research, with a wealth of literature in the field [Evans, 2022]. There are general purpose PDE solvers as well as algorithms specifically designed for MVEs. For the interest of this paper, for general purpose PDE solver, we will only consider the instance that can be used to solve the MVE under consideration.

Categorize PDE solvers via solution representation. To better understand the benefits of neural network (NN) based PDE solvers and to compare our approach with others, we categorize the literature based on the representation of the solution to the PDE. These representations can be roughly grouped into four categories:
1. Discretization-based representation: The solution to the PDE is represented as discrete function values at grid points, finite-size cells, or finite-element meshes.
2. Representation as a combination of basis functions: The solution to the PDE is approximated as a sum of basis functions, e.g. Fourier series, Legendre polynomials, or Chebyshev polynomials.
3. Representation using a collection of particles: The solution to the PDE is represented as a collection of particles, each described by its weight, position, and other relevant information.
4. NN-based representation: NNs offer many strategies for representing the solution to the PDE, such as using the NN directly to represent the solution, using normalizing flow or GAN-based parameterization to ensure the non-negativity and conservation of mass of the solution, or using the NN to parameterize the underlying dynamics of the PDE, such as the time-varying velocity field that drives the evolution of the system.

The drawback of the first three strategies is that a sparse representation[3] leads to reduced solution accuracy, while a dense representation results in increased computational and memory cost. NNs, as powerful function approximation tools, are expected to surpass these strategies by being able to handle higher-dimensional, less regular, and more complicated systems [Weinan et al., 2021].

Given a representation strategy of the solution, an effective solver must exploit the underlying properties of the system to find the best candidate solution. Four notable properties that are utilized to design solvers are: (A) the PDE definition or weak formulation of the system, (B) the SDE interpretation of the system, (C) the variational interpretation, particularly the Wasserstein gradient flow interpretation, and (D) the stability property of the system. These properties are combined with the solution representations mentioned earlier to form different methods. For example, the Finite

[3]For example, sparser grid, cell or mesh with less granularity, less basis functions, less particles.

Difference method [Smith et al., 1985], Finite Volume method [Moukalled et al., 2016], and Finite Element method [Johnson, 2012] represent the solution of partial differential equations (PDEs) by discretizing the solution and utilize property (A), at least in their original form. On the other hand, a recent work by Carrillo et al. [2022] solves PDEs admitting a Wasserstein gradient flow structure using the classic JKO scheme [Jordan et al., 1998], which is based on property (C), and the solution is also represented via discretization. The Spectral method [Shen et al., 2011] is a class of methods that exploits property (A) by representing the solution as a combination of basis functions. The Random Vortex Method [Long, 1988] is a highly successful method for solving the vorticity formulation of the 2D Navier-Stokes equation by exploiting property (C) and representing the solution with particles. The Blob method from Carrillo et al. [2019] is another particle-based method for solving PDEs that describe diffusion processes, which also exploits property (C). In the following, we will focus on the methods that uses NN for solution representation.

Comparison with NN-based solvers The Physics-Informed Neural Network (PINN) is a widely used neural network-based solver that leverages property (A) [Raissi et al., 2019, Yang and Perdikaris, 2019]. By expressing a PDE as $\mathcal{A}(\boldsymbol{g}) = 0$ and its time and space boundary conditions as $\mathcal{B}(\boldsymbol{g}) = 0$, where $\boldsymbol{g}$ is a candidate solution and $\mathcal{A}$ and $\mathcal{B}$ are operators acting on $\boldsymbol{g}$, PINN parameterizes $\boldsymbol{g}$ using a neural network $\boldsymbol{g}^\theta$ and optimizes its parameters θ by minimizing the functional $\mathcal{L}^2$ norms $L(\theta) = \|\mathcal{A}(\boldsymbol{g}^\theta)\|^2_{\mathcal{L}^2} + \lambda\|\mathcal{B}(\boldsymbol{g}^\theta)\|^2_{\mathcal{L}^2}$. The hyperparameter λ balances the residuals of $\mathcal{A}$ and $\mathcal{B}$ and must be adjusted for optimal performance. PINN is versatile and can be applied to a wide range of PDEs, but its performance may not be as good as other neural network-based solvers specifically designed for a particular class of PDEs, as it does not take into account other in-depth properties of the system [Krishnapriyan et al., 2021, Wang et al., 2022]. In general, there is no theoretical guarantee on how the loss $L(\theta)$ controls the discrepancy between the candidate solution $\boldsymbol{g}^\theta$ and the ground truth.

A recent work from Zhang et al. [2022] exploits the property (B) to design the Random Deep Vortex Network (RDVN) method for solving the 2D Navier-Stokes equation and achieves SOTA performance for this task. Let $\boldsymbol{u}_t^\theta$ be an estimation of the interaction term $K * \rho_t$ in the SDE 1 and use ρ_t^θ to denote the law of the particle driven by the SDE $\mathrm{d}\mathbf{X}_t = V(\mathbf{X}_t)\mathrm{d}t + \boldsymbol{u}_t^\theta(\mathbf{X}_t)\mathrm{d}t + \sqrt{2\nu}\mathrm{d}\mathbf{B}_t$. The idea of RDVN is to minimize the time average of the $\mathcal{L}^2$ norms $L(\theta) = \int_0^T \|\boldsymbol{u}_t^\theta - K * \rho_t^\theta\|^2_{\mathcal{L}^2}\mathrm{d}t$. Note that in order to simulate the SDE, one needs to discretize the time variable in loss function L. After training θ, ρ_t^θ is output as the estimated solution of the 2D Navier-Stokes equation. However, no convergence guarantee is given for the relationship between L and the discrepancy between ρ_t^θ and the ground truth ρ_t. In our experiment section, we will extend RDVN to solve the MVE with Coulomb interaction as a baseline.

In another line of research, Fan et al. [2022] exploits property (C) to propose the Primal-dual Gradient Flow (PDGF) method for the Fokker-Planck equation. The solution to the Fokker-Planck equation coincides with the Wasserstein gradient flow of the free energy functional. PDGF iteratively creates a sequence of transport maps by approximately solving the minimizing movement scheme, also referred to as the JKO scheme [Jordan et al., 1998]. These learned maps can be used to reconstructed the solution to the Fokker-Planck equation. Unlike its previous work [Mokrov et al., 2021], PDGF uses the variational formulation of the f-divergence, allowing the gradient flow to be directly constructed from empirical distributions. However, the variational formulation that PDGF is based on involves an extra dual variable, making each iteration of PDGF a non-convex non-concave minimization-maximization problem, which can be difficult to solve. While the MVE with Coulomb interaction also has a Wasserstein gradient flow interpretation, unlike the 2D Navier-Stokes equation, it is unclear how PDGF can be generalized to solve this problem. As a result, we did not compare with PDGF in our experiments.

Shen et al. [2022] propose the concept of self-consistency for the Fokker-Planck equation (a specific instance of MVE without the interaction term $K \equiv 0$), which most related to our research (as mentioned in the introduction). Unlike our work where the self-consistency potential is derived via the principle of stability analysis, this previous work constructs the self-consistency potential $R(f)$ for the hypothesis velocity field f by observing that the underlying velocity field f^* is the fixed point of some velocity-consistent transformation $\mathcal{A}$ and they construct $R(f)$ to be a more complicated Sobolev norm of the residual $f - \mathcal{A}(f)$. In their result, they bound the Wasserstein distance between ρ^f and ρ by $R(f)$, which is weaker than our KL type control. The improved KL type control for the Fokker-Planck equation has also been discussed in [Boffi and Vanden-Eijnden, 2022]. A very recent work [Li et al., 2023] extends the self-consistency approach to compute the general Wasserstein gradient flow numerically, without providing further theoretical justification.

5 Experiments

We name the method of minimizing the self-consistency potential 7 as Entropy-dissipation Informed Neural Network (EINN). To show the efficacy and efficiency of the proposed approach, we conduct numerical studies on example problems that admit explicit solutions and compare the results with SOTA NN-based PDE solvers. The included

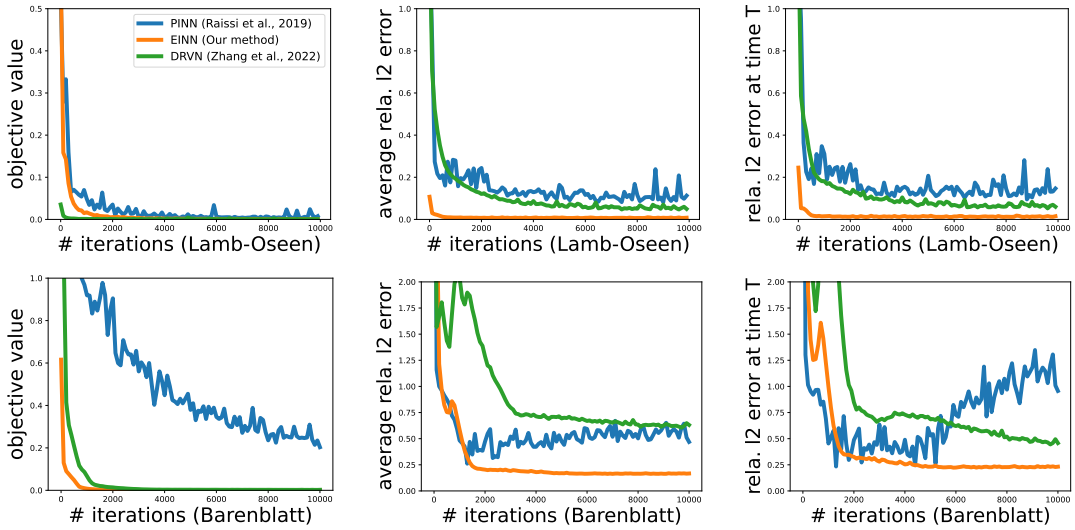

Figure 1: The first row contains results for the 2D Navier-Stokes equation and the second row contains the results for the 3D McKean-Vlasov equation with Coulomb interaction. The first column reports the objective losses, while the second and third columns report the average and last-time-stamp relative ℓ_2 error.

baselines are the Physics-Informed Neural Network (PINN) [Raissi et al., 2019] and the Deep Random Vortex Network (DRVN) [Zhang et al., 2022]. Note that these baselines only considered the 2D Navier-Stokes equation, which we followed in our implementation. We extend them to solve the MVE with the Coulomb interaction for the purpose of comparison, and the detailed implementation are discussed in Appendix A.1.

Equations with an Explicit Solution We consider the following two instances that admit explicit solutions. The first one is for the 2D Navier-Stokes equation while the second one is for the MVE with the Coulomb interaction. We verify these solutions in Appendix A.2.
Lamb-Oseen Vortex (2D Navier-Stokes equation) [Oseen, 1912]: Consider the whole domain case where $\mathcal{X} = \mathbb{R}^2$ and the Biot-Savart interaction kernel defined in equation 4. Let $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. If $\rho_0 = \mathcal{N}(0, \sqrt{2\nu t_0}\boldsymbol{I}_2)$ for some $t_0 \geq 0$, then we have $\rho_t(\boldsymbol{x}) = \mathcal{N}(0, \sqrt{2\nu(t+t_0)}\boldsymbol{I}_2)$.
Barenblatt solutions (McKean-Vlasov equation) [Serfaty and Vázquez, 2014]: Consider the 3D MVE with the Coulomb interaction kernel 3 with the diffusion coefficient set to zero, i.e. $d = 3$ and $\nu = 0$. Let $\mathrm{Uniform}[\mathbb{A}]$ be the uniform distribution over a set $\mathbb{A}$. Consider the whole domain case where $\mathcal{X} = \mathbb{R}^3$. If $\rho_0 = \mathrm{Uniform}[\|\boldsymbol{x}\| \leq (\frac{3}{4\pi}t_0)^{1/3}]$ for some $t_0 \geq 0$, then we have $\rho_t = \mathrm{Uniform}[\|\boldsymbol{x}\| \leq (\frac{3}{4\pi}(t+t_0))^{1/3}]$.

Numerical Results We present the results of our experiments in Figure 1, where the first row contains the result for the Lamb-Oseen vortex (2D Navier-Stokes equation) and the second row contains the result for the Barenblatt model (3D McKean-Vlasov equation). The explicit solutions of these models allow us to assess the quality of the output of the included methods. Specifically, given a hypothesis solution ρ_t^f, the ground truth ρ and the interaction kernel K, define the relative ℓ_2 error at time stamp t as $Q(t) = \int_\Omega \frac{\|K*(\rho_t^f - \rho_t)(\boldsymbol{x})\|}{\|K*\rho_t(\boldsymbol{x})\|} d\boldsymbol{x}$, where Ω is some domain where ρ_t has non-zero density. We are particularly interested in the quality of the convolution term $K * \rho_t^f$ since it has physical meanings. In the Biot-Savart kernel case, it is the velocity of the fluid, while in the Coulomb case it is the Coulomb field. We set Ω to be $[-2, 2]^2$ for the Lamb-Oseen vortex and to $[-0.1, 0.1]^3$ for the Barenblatt model. For both models, we take $\nu = 0.1$, $t_0 = 0.1$, and $T = 1$. The neural network that we use is an MLP with 7 hidden layers, each of which has 20 neurons.

From the first column of Figure 1, we see that the objective loss of all methods have substantially reduced over a training period of 10000 iterations. This excludes the possibility that a baseline has worse performance because the neural network is not well-trained and now the quality of the solution now solely depends on the efficacy of the method. From the second and third columns, we see that the proposed EINN method significantly outperforms the other two methods in terms of the time-average relative ℓ_2 error, i.e. $\frac{1}{T}\int_0^T Q(t)dt$ and the relative ℓ_2 error at the last time stamp $Q(T)$. This shows the advantage of our method.

Conclusion By employing entropy dissipation of the McKean-Vlasov system, we design a self-consistency potential function for a hypothesis velocity field such that it controls the KL divergence between the corresponding hypothesis solution and the ground truth, for any time stamp within the period of interest. Built on this self-consistency potential, we proposed a neural network based MVE solver and derived the detailed computation method of the stochastic gradient, using the classic adjoint method. Through empirical studies on examples of the 2D Navier-Stokes equation and the 3D McKean-Vlasov equation with Coulomb interactions, we show the significant advantage of the proposed method, when compared with two SOTA NN based PDE solvers.

References

N. M. Boffi and E. Vanden-Eijnden. Probability flow solution of the fokker-planck equation. *arXiv preprint arXiv:2206.04642*, 2022.

D. Bresch, P.-E. Jabin, and Z. Wang. On mean-field limits and quantitative estimates with a large class of singular kernels: application to the patlak–keller–segel model. *Comptes Rendus Mathematique*, 357(9):708–720, 2019a.

D. Bresch, P.-E. Jabin, and Z. Wang. Modulated free energy and mean field limit. *Séminaire Laurent Schwartz—EDP et applications*, pages 1–22, 2019b.

S. Cai, Z. Mao, Z. Wang, M. Yin, and G. E. Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738, 2021.

J. A. Carrillo, K. Craig, and F. S. Patacchini. A blob method for diffusion. *Calculus of Variations and Partial Differential Equations*, 58:1–53, 2019.

J. A. Carrillo, K. Craig, L. Wang, and C. Wei. Primal dual methods for wasserstein gradient flows. *Foundations of Computational Mathematics*, pages 1–55, 2022.

S. Cuomo, V. S. Di Cola, F. Giampaolo, G. Rozza, M. Raissi, and F. Piccialli. Scientific machine learning through physics–informed neural networks: where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28. American Mathematical Soc., 2017.

L. C. Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

J. Fan, Q. Zhang, A. Taghvaei, and Y. Chen. Variational wasserstein gradient flow. In *International Conference on Machine Learning*, pages 6185–6215. PMLR, 2022.

F. Golse. On the dynamics of large particle systems in the mean field limit. *Macroscopic and large scale phenomena: coarse graining, mean field limits and ergodicity*, pages 1–144, 2016.

A. Guillin, P. L. Bris, and P. Monmarché. Uniform in time propagation of chaos for the 2d vortex model and other singular stochastic systems. *arXiv preprint arXiv:2108.08675*, 2021.

J. Han, A. Jentzen, et al. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in mathematics and statistics*, 5(4): 349–380, 2017.

J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

P.-E. Jabin and Z. Wang. Quantitative estimates of propagation of chaos for stochastic systems with $\mathrm{W}^{-1,\infty}$ kernels. *Inventiones mathematicae*, 214(1):523–591, 2018.

C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Courier Corporation, 2012.

R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1):1–17, 1998.

G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

A. Krishnapriyan, A. Gholami, S. Zhe, R. Kirby, and M. W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, 34:26548–26560, 2021.

L. Li, S. Hurault, and J. Solomon. Self-consistent velocity matching of probability flows. *arXiv preprint arXiv:2301.13737*, 2023.

Z. Li, N. B. Kovachki, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

D.-G. Long. Convergence of the random vortex method in two dimensions. *Journal of the American Mathematical Society*, 1(4):779–804, 1988.

A. J. Majda, A. L. Bertozzi, and A. Ogawa. Vorticity and incompressible flow. cambridge texts in applied mathematics. *Appl. Mech. Rev.*, 55(4):B77–B78, 2002.

P. Mokrov, A. Korotin, L. Li, A. Genevay, J. M. Solomon, and E. Burnaev. Large-scale wasserstein gradient flows. *Advances in Neural Information Processing Systems*, 34:15243–15256, 2021.

F. Moukalled, L. Mangani, M. Darwish, F. Moukalled, L. Mangani, and M. Darwish. *The finite volume method*. Springer, 2016.

C. Oseen. Uber die wirbelbewegung in einer reibenden flussigkeit. *Ark. Mat. Astro. Fys.*, 7, 1912.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

M. Raissi, A. Yazdani, and G. E. Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

M. Rosenzweig. Mean-field convergence of point vortices to the incompressible euler equation with vorticity in $\mathcal{L}^\infty$. *Archive for Rational Mechanics and Analysis*, 243(3):1361–1431, 2022.

S. Serfaty. Mean field limit for coulomb-type flows. *Duke Math. J.*, 169(15):2887–2935, 2020.

S. Serfaty and J. L. Vázquez. A mean field equation as limit of nonlinear diffusions with fractional laplacian operators. *Calculus of Variations and Partial Differential Equations*, 49:1091–1120, 2014.

J. Shen, T. Tang, and L.-L. Wang. *Spectral methods: algorithms, analysis and applications*, volume 41. Springer Science & Business Media, 2011.

Z. Shen, Z. Wang, S. Kale, A. Ribeiro, A. Karbasi, and H. Hassani. Self-consistency of the fokker planck equation. In P.-L. Loh and M. Raginsky, editors, *Proceedings of Thirty Fifth Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 817–841. PMLR, 02–05 Jul 2022. URL `https://proceedings.mlr.press/v178/shen22a.html`.

G. D. Smith, G. D. Smith, and G. D. S. Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford university press, 1985.

C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

C. Wang, S. Li, D. He, and L. Wang. Is $\mathcal{L}^2$ physics informed loss always suitable for training physics informed neural network? In *Advances in Neural Information Processing Systems*, 2022.

H. Wang, L. Zhang, J. Han, and E. Weinan. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Computer Physics Communications*, 228:178–184, 2018.

E. Weinan, J. Han, and A. Jentzen. Algorithms for solving high dimensional pdes: from nonlinear monte carlo to machine learning. *Nonlinearity*, 35(1):278, 2021.

Y. Yang and P. Perdikaris. Adversarial uncertainty quantification in physics-informed neural networks. *Journal of Computational Physics*, 394:136–152, 2019.

L. Zhang, J. Han, H. Wang, and R. Car. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120(14):143001, 2018.

R. Zhang, P. Hu, Q. Meng, Y. Wang, R. Zhu, B. Chen, Z.-M. Ma, and T.-Y. Liu. Drvn (deep random vortex network): A new physics-informed machine learning method for simulating and inferring incompressible fluid flows. *Physics of Fluids*, 34(10):107112, 2022.

A More Details on the Experiments

A.1 Implementations of Baselines

Objectives of PINN

- For the vorticity equation of the 2D Navier-Stokes equation, let $\boldsymbol{u} : [0, T] \times \mathbb{R}^2 \to \mathbb{R}^2$ be the velocity field (this should not be confused with the velocity field of the continuity equation) such that $\nabla \cdot \boldsymbol{u} = 0$, i.e. $\boldsymbol{u}$ is

divergence-free, and let $\omega = \nabla \times \boldsymbol{u} \in \mathbb{R}$ be the vorticity. We have

$$\frac{\partial \omega}{\partial t} + \nabla \cdot (\omega \boldsymbol{u}) = \nu \Delta \omega, \tag{27}$$

$$\omega = \nabla \times \boldsymbol{u}. \tag{28}$$

We use this form to construct the objective for the PINN method

$$\int_0^T \| \frac{\partial \omega}{\partial t} + \nabla \cdot (\omega \boldsymbol{u}) - \nu \Delta \omega \|^2_{\mathcal{L}(\Omega)^2} + \| \omega - \nabla \times \boldsymbol{u} \|_{\mathcal{L}\mathcal{L}(\Omega)^2} dt, \tag{29}$$

where $\mathcal{L}^2(\Omega)$ denotes the functional $\mathcal{L}^2$ norm on the domain $\Omega = [-2, 2]^2$.

- For the MVE with Coulomb interaction, let g be the Coulomb potential defined in equation 3. We have that $\psi = g * \rho$ is the solution to the Poisson equation $\Delta \psi = -\rho$ and $K * \rho = -\nabla \psi$. We have

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \cdot (-\nabla \psi)) = \nu \Delta \rho \tag{30}$$

$$\Delta \psi = -\rho. \tag{31}$$

Expand the the divergence to obtain

$$\frac{\partial \rho}{\partial t} + \nabla \rho \cdot (-\nabla \psi) + \rho \cdot (-\Delta \psi) = \nu \Delta \rho \tag{32}$$

$$\Delta \psi = -\rho. \tag{33}$$

Now plug in the $\Delta \psi = -\rho$ to arrive at the following equivalent form

$$\frac{\partial \rho}{\partial t} + \nabla \rho \cdot -\nabla \psi + \rho^2 = \nu \Delta \rho \tag{34}$$

$$\Delta \psi = -\rho. \tag{35}$$

We use this form to construct the objective for the PINN method.

$$\int_0^T \| \frac{\partial \rho}{\partial t} + \nabla \rho \cdot -\nabla \psi + \rho^2 - \nu \Delta \rho \|^2_{\mathcal{L}^2(\Omega)} + \| \Delta \psi + \rho \|^2_{\mathcal{L}^2}, \tag{36}$$

where $\mathcal{L}^2(\Omega)$ denotes the functional $\mathcal{L}^2$ norm on the domain $\Omega = [-1, 1]^2$.

DRVN In the original paper [Zhang et al., 2022], only the Biot-Savart kernel is concerned. We can easily extend the DRVN method to the Coulomb case by setting K to be the kernel defined in equation 3.

A.2 Examples with an Explicit Solution

In this section, we verify the explicit solutions discussed in the experiment section.

Lamb-Oseen Vortex on the whole domain $\mathbb{R}^2$. Recall that we consider the 2D Navier-Stokes equation (the MVE with the Biot-Savart interaction kernel (4)). The Lamb-Oseen Vortex model states that, if $\rho_0 = \mathcal{N}(0, \sqrt{2\nu t_0} \boldsymbol{I}_2)$ for some $t_0 \geq 0$, then we have $\rho_t(\boldsymbol{x}) = \mathcal{N}(0, \sqrt{2\nu(t + t_0)} \boldsymbol{I}_2)$.

To verify this, define $\boldsymbol{u}_t(\boldsymbol{x}) = \frac{1}{\sqrt{\nu(t+t_0)}} \boldsymbol{v}(\frac{\boldsymbol{x}}{\sqrt{\nu(t+t_0)}})$, where

$$\boldsymbol{v}(\boldsymbol{x}) = \frac{1}{2\pi} \frac{\boldsymbol{x}^\perp}{\|\boldsymbol{x}\|^2} \left(1 - \exp(-\frac{1}{4} \|\boldsymbol{x}\|^2) \right). \tag{37}$$

One can easily check that $\nabla \cdot \boldsymbol{u}_t \equiv 0$ and hence there exists a function ψ_t such that $\nabla^\perp \psi_t = -\boldsymbol{u}_t$, where $\nabla^\perp$ denotes the perpendicular gradient, defined as $\nabla^\perp = (-\partial_{\boldsymbol{x}_2}, \partial_{\boldsymbol{x}_2})$, and ψ_t is called the stream function in the literature of fluid dynamics. Moreover, one can verify that $\nabla \times \boldsymbol{u}_t = \rho_t$ where $\nabla \times$ denotes the curl of a 2D velocity field, defined as $\nabla \times \boldsymbol{u}(\boldsymbol{x}) = \partial \boldsymbol{u}_2 / \partial \boldsymbol{x}_1 - \partial \boldsymbol{u}_1 / \partial \boldsymbol{x}_2$. Together we have

$$\Delta \psi_t = -\rho_t, \tag{38}$$

i.e., the stream function ψ_t is the solution to the 2D Poisson equation with a source term ρ_t.

Under the boundary condition that $\psi_t(\boldsymbol{x}) \to 0$ for $\|\boldsymbol{x}\| \to \infty$, we can express ψ_t via the unique Green function $G(\boldsymbol{x}) = \frac{1}{2\pi} \ln \|\boldsymbol{x}\|$ as

$$\psi_t(\boldsymbol{x}) = G * \rho_t = \frac{1}{2\pi} \int \ln \|\boldsymbol{x} - \boldsymbol{y}\| \rho_t(\boldsymbol{y}) d\boldsymbol{y}. \tag{39}$$

Consequently, by taking the perpendicular gradient, we obtain

$$\boldsymbol{u}_t = \nabla^\perp \psi_t = \frac{1}{2\pi} \int \frac{(\boldsymbol{x} - \boldsymbol{y})^\perp}{\|\boldsymbol{x} - \boldsymbol{y}\|^2} \rho_t(\boldsymbol{y}) d\boldsymbol{y} = K * \rho_t. \tag{40}$$

Finally, by plugging the expressions of ρ_t and $\boldsymbol{u}_t = K * \rho_t$ in the MVE (2), we verified the Lamb-Oseen vortex.

Barenblatt solutions for the MVE with Coulomb Interaction. Recall that we consider the MVE with the Coulomb interaction kernel (3) for $d = 3$ and set the diffusion coefficient $\nu = 0$, i.e.

$$\frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t \cdot -\nabla \psi_t) = 0 \tag{41}$$

where ψ_t is the solution to the Poisson equation $\Delta \psi_t = -\rho_t$. The Barenblatt solution of the above MVE is stated as follows: If $\rho_0 = \mathrm{Uniform}[\|\boldsymbol{x}\| \le (\frac{3}{4\pi} t_0)^{1/3}]$ for some $t_0 \ge 0$, then we have

$$\rho_t = \mathrm{Uniform}[\|\boldsymbol{x}\| \le (\frac{3}{4\pi}(t + t_0))^{1/3}] \tag{42}$$

We now verify this solution.

Recall that the volume of a three dimensional Euclidean ball with radius R is $\frac{4\pi}{3} R^3$. Hence we can write the density function as $\rho_t(x) = \frac{1}{t+t_0} \chi_{\|\boldsymbol{x}\| \le (\frac{3}{4\pi}(t+t_0))^{1/3}}$, where $\chi_{\mathbb{X}}$ is a function that takes value 1 for $\boldsymbol{x} \in \mathbb{X}$ and takes value 0 for $\boldsymbol{x} \notin \mathbb{X}$. Take

$$\psi_t(\boldsymbol{x}) = \begin{cases} \frac{2(\frac{3}{4\pi}(t+t_0))^{2/3} - \|\boldsymbol{x}\|^2}{6(t+t_0)}, & \|\boldsymbol{x}\| \le (\frac{3}{4\pi}(t+t_0))^{1/3}, \\ \frac{1}{8\pi \|\boldsymbol{x}\|}, & \|\boldsymbol{x}\| > (\frac{3}{4\pi}(t+t_0))^{1/3}. \end{cases} \tag{43}$$

It can be verified that the Poisson equation $\Delta \psi_t = -\rho_t$ holds (note that $\Delta \|\boldsymbol{x}\|^{-1} = 0$, i.e. $\|\boldsymbol{x}\|^{-1}$ is a harmonic function for $d = 3$). Consequently, for a fixed time stamp t and any $\|\boldsymbol{x}\| \le (\frac{3}{4\pi}(t + t_0))^{1/3}$ we have

$$\frac{\partial \rho_t}{\partial t}(\boldsymbol{x}) + \nabla \cdot (\rho_t(\boldsymbol{x}) \cdot -\nabla \phi_t(\boldsymbol{x})) = -\frac{1}{(t+t_0)^2} + \frac{1}{(t+t_0)^2} = 0, \tag{44}$$

which verifies this solution.

B Adjoint Method

Consider the ODE system

$$\begin{aligned} \dot{s}(t) &= \psi(s(t), t, \theta) \\ s(0) &= s_0, \end{aligned}$$

and the objective loss

$$\ell(\theta) = \int_0^T g(s(t), t, \theta) \mathrm{d}t. \tag{45}$$

The following proposition computes the gradient of ℓ w.r.t. θ. We omit the parameters of the functions for succinctness. We note that all the functions in the integrands should be evaluated at the corresponding time stamp t, e.g. $b^\top \frac{\partial h}{\partial \theta} \mathrm{d}t$ abbreviates for $b(t)^\top \frac{\partial}{\partial \theta} h(\xi(t), x(t), t, \theta) \mathrm{d}t$.

Proposition 1.

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \int_0^T a^\top \frac{\partial \psi}{\partial \theta} + \frac{\partial g}{\partial \theta} \mathrm{d}t. \tag{46}$$

where $a(t)$ is solution to the following final value problems

$$\dot{a}^\top + a^\top \frac{\partial \psi}{\partial s} + \frac{\partial g}{\partial s} = 0, a(T) = 0, \tag{47}$$

Proof. Let us define the Lagrange multiplier function (or the adjoint state) $a(t)$ dual to $s(t)$. Moreover, let L be an augmented loss function of the form

$$L = \ell - \int_0^T a^\top (\dot{s} - \psi) \mathrm{d}t. \tag{48}$$

Since we have $\dot{s}(t) = \psi(s(t), t, \theta)$ by construction, the integral term in L is always null and a can be freely assigned while maintaining $\mathrm{d}L/\mathrm{d}\theta = \mathrm{d}\ell/\mathrm{d}\theta$. Using integral by part, we have

$$\int_0^T a^\top \dot{s} \, \mathrm{d}t = a(t)^\top s(t)|_0^T - \int_0^T s^\top \dot{a} \, \mathrm{d}t. \tag{49}$$

We obtain

$$L = -a(t)^\top s(t)|_0^T + \int_0^T \dot{a}^\top s + a^\top \psi + g \, \mathrm{d}t. \tag{50}$$

Now we compute the gradient of L w.r.t. θ as

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \frac{\mathrm{d}L}{\mathrm{d}\theta} = -a(T)^\top \frac{\mathrm{d}x(T)}{\mathrm{d}\theta} + \int_0^T \dot{a}^\top \frac{\mathrm{d}s}{\mathrm{d}\theta} + a^\top \left(\frac{\partial \psi}{\partial \theta} + \frac{\partial \psi}{\partial s} \frac{\mathrm{d}s}{\mathrm{d}\theta} \right) \mathrm{d}t + \int_0^T \frac{\partial g}{\partial s} \frac{\mathrm{d}s}{\mathrm{d}\theta} + \frac{\partial g}{\partial \theta} \mathrm{d}t,$$

which by rearranging terms yields to

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \frac{\mathrm{d}L}{\mathrm{d}\theta} = -a(T)^\top \frac{\mathrm{d}x(T)}{\mathrm{d}\theta} + \int_0^T a^\top \frac{\partial \psi}{\partial \theta} + \frac{\partial g}{\partial \theta} \mathrm{d}t + \int_0^T \left(\dot{a}^\top + a^\top \frac{\partial \psi}{\partial s} + \frac{\partial g}{\partial s} \right) \frac{\mathrm{d}s}{\mathrm{d}\theta} \mathrm{d}t.$$

Now by taking a satisfying the *final* value problems

$$\dot{a}^\top + a^\top \frac{\partial \psi}{\partial s} + \frac{\partial g}{\partial s} = 0, a(T) = 0, \tag{51}$$

we derive the result

$$\frac{\mathrm{d}\ell}{\mathrm{d}\theta} = \int_0^T a^\top \frac{\partial \psi}{\partial \theta} + \frac{\partial g}{\partial \theta} \mathrm{d}t. \tag{52}$$

□

C Detailed Proofs

Proof of Lemma 1. Recall the McKean-Vlasov equation 5 and the continuity equation 9. For simplicity we write that $\rho_t = \rho_t^f$ and omit the integration domain $\mathcal{X}$. Then

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} = \int \partial_t \rho_t \log \frac{\rho_t}{\bar{\rho}_t} + \int \rho_t \partial_t \log \rho_t - \int \rho_t \partial_t \log \bar{\rho}_t \\
&= - \int \mathrm{div}\Big(\rho_t \Big(\Big[- \nabla V(x) + K * \rho_t - \nu \nabla \log \rho_t \Big] + \delta_t \Big) \Big) \log \frac{\rho_t}{\bar{\rho}_t} \\
&+ \int \frac{\rho_t}{\bar{\rho}_t} \mathrm{div}\Big(\bar{\rho}_t \Big(- \nabla V(x) + K * \bar{\rho}_t - \nu \nabla \log \bar{\rho}_t \Big) \Big),
\end{aligned}$$

where we note that $\int \rho_t \partial_t \log \rho_t = \int \partial_t \rho_t = 0$ since the total mass is preserved over time. By integration by parts, one has

$$\frac{\mathrm{d}}{\mathrm{d}t} \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} = I_1 + I_2 + I_3 + \int \rho_t \delta_t \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t},$$

where I_1, I_2, I_3denote the linear part, nonlinear interaction part and the diffusion part separately. More precisely, by integration by parts,

$$\begin{aligned}
I_1 &= \int \mathrm{div}(\rho_t \nabla V(x)) \log \frac{\rho_t}{\bar{\rho}_t} - \int \frac{\rho_t}{\bar{\rho}_t} \mathrm{div}(\bar{\rho}_t \nabla V(x)) \\
&= - \int \rho_t \nabla V(x) \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t} + \int \bar{\rho}_t \nabla \frac{\rho_t}{\bar{\rho}_t} \cdot \nabla V(x) = 0.
\end{aligned}$$

And

$$
\begin{aligned}
I_2 &= -\int \operatorname{div}(\rho_t K * \rho_t) \log \frac{\rho_t}{\bar{\rho}_t} + \int \frac{\rho_t}{\bar{\rho}_t} \operatorname{div}(\bar{\rho}_t K * \bar{\rho}_t) \\
&= \int \rho_t K * \rho_t \nabla \log \frac{\rho_t}{\bar{\rho}_t} - \int \bar{\rho}_t K * \bar{\rho}_t \cdot \nabla \frac{\rho_t}{\bar{\rho}_t} \\
&= \int \rho_t \nabla \log \frac{\rho_t}{\bar{\rho}_t} \cdot K * (\rho_t - \bar{\rho}_t).
\end{aligned}
$$

Given that the kernel K is divergence free, that is $\operatorname{div} K = 0$, one further has

$$
\begin{aligned}
I_2 &= -\int \rho_t \nabla \log \bar{\rho}_t \cdot K * (\rho_t - \bar{\rho}_t) + \int \nabla \rho_t \cdot K * (\rho_t - \bar{\rho}_t) \\
&= -\int \rho_t \nabla \log \bar{\rho}_t \cdot K * (\rho_t - \bar{\rho}_t).
\end{aligned} \tag{53}
$$

Note that this modification will be used in the proof in the 2D Navier-Stokes case. Finally all diffusion terms sum up to I_3 which can be further simplified as

$$
\begin{aligned}
I_3 &= \nu \int \operatorname{div}(\rho_t \nabla \log \rho_t) \log \frac{\rho_t}{\bar{\rho}_t} - \nu \int \frac{\rho_t}{\bar{\rho}_t} \operatorname{div}(\bar{\rho}_t \nabla \log \bar{\rho}_t) \\
&= -\nu \int \rho_t \nabla \log \rho_t \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t} + \nu \int \bar{\rho}_t \nabla \log \bar{\rho}_t \cdot \nabla \frac{\rho_t}{\bar{\rho}_t} \\
&= -\nu \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2.
\end{aligned}
$$

We thus complete the proof of Lemma 1.

□

Proof of Lemma 2. Recall that $K = -\nabla g$. For simplicity we write that $\rho_t = \rho_t^f$. Then

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} F(\rho_t, \bar{\rho}_t) = \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2} \int_{\mathcal{X}^2} g(x-y) \mathrm{d}(\rho_t - \bar{\rho}_t)^{\otimes 2}(x,y) \\
&= \int_{\mathcal{X}} g * (\rho_t - \bar{\rho}_t)(x) \big(\partial_t \rho_t(x) - \partial_t \bar{\rho}_t(x)\big) \mathrm{d}x \\
&= \int g * (\rho_t - \bar{\rho}_t)(x) \operatorname{div}\Big\{ \rho_t \Big([\nabla V(x) - K * \rho_t + \nu \nabla \log \rho_t] - \delta_t \Big) \\
&\qquad\qquad - \bar{\rho}_t \Big(\nabla V(x) - K * \bar{\rho}_t + \nu \log \bar{\rho}_t \Big) \Big\} \\
&= J_1 + J_2 + J_3 + J_4,
\end{aligned}
$$

where J_1, J_2, J_3, J_4 denote the perturbation term, the linear difference term, the nonlinear difference term, and the diffusion term respectively. The perturbation term J_1 reads

$$
J_1 = -\int_{\mathcal{X}} g * (\rho_t - \bar{\rho}_t) \operatorname{div}(\rho_t \delta_t) = -\int_{\mathcal{X}} \rho_t K * (\rho_t - \bar{\rho}_t) \cdot \delta_t.
$$

By integration by parts, the linear difference term can be written as

$$
\begin{aligned}
J_2 &= \int_{\mathcal{X}} g * (\rho_t - \bar{\rho}_t) \operatorname{div}\Big((\rho_t - \bar{\rho}_t) \nabla V \Big) = \int_{\mathcal{X}} K * (\rho_t - \bar{\rho}_t)(\rho_t - \bar{\rho}_t) \nabla V \\
&= \frac{1}{2} \int_{\mathcal{X}^2} K(x-y)(\nabla V(x) - V(y)) \mathrm{d}(\rho_t - \bar{\rho}_t)^{\otimes 2}(x,y),
\end{aligned}
$$

where the last equality is true since $K = -\nabla g$ is an odd function and we do the symmetrization trick, i.e. exchanging the role of x and y to another term and then taking the average.

The nonlinear difference term reads

$$
\begin{aligned}
J_3 &= -\int_{\mathcal{X}} g * (\rho_t - \bar{\rho}_t) \operatorname{div}\Big(\rho_t K * \rho_t - \bar{\rho}_t K * \bar{\rho}_t \Big) \\
&= -\int_{\mathcal{X}} K * (\rho_t - \bar{\rho}_t)(\rho_t K * (\rho_t - \bar{\rho}_t) - \int_{\mathcal{X}} K * (\rho_t - \bar{\rho}_t)(\rho_t - \bar{\rho}_t) K * \bar{\rho}_t \\
&= -\int_{\mathcal{X}} \rho_t |K * (\rho_t - \bar{\rho}_t)|^2 - \frac{1}{2} \int K(x-y)(K * \bar{\rho}_t(x) - K * \bar{\rho}_t(y)) \mathrm{d}(\rho_t - \bar{\rho}_t)^{\otimes 2}(x,y),
\end{aligned}
$$

where again in the last term we do the symmetrization.

The diffusion term reads

$$\begin{aligned}
J_4 &= \nu \int g * (\rho_t - \bar{\rho}_t) \operatorname{div}\Big(\rho_t \nabla \log \rho_t - \bar{\rho}_t \nabla \log \bar{\rho}_t\Big) \\
&= \nu \int K * (\rho_t - \bar{\rho}_t) \rho_t \nabla \log \frac{\rho_t}{\bar{\rho}_t} + \nu \int K * (\rho_t - \bar{\rho}_t)(\rho_t - \bar{\rho}_t) \nabla \log \bar{\rho}_t \\
&= \nu \int_{\mathcal{X}} \rho_t K * (\rho_t - \bar{\rho}_t) \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t} \\
&\quad + \frac{\nu}{2} \int_{\mathcal{X}^2} K(x-y)(\nabla \log \bar{\rho}_t(x) - \nabla \log \bar{\rho}_t(y)) \mathrm{d}(\rho_t - \bar{\rho}_t)^{\otimes 2}.
\end{aligned}$$

To sum it up, we prove the thesis.

□

C.1 Proof of the 2D Navier-Stokes case

Now we proceed to control the growth of the KL divergence $\mathbf{KL}(\rho_t^f | \bar{\rho}_t)$ for the 2D Navier-Stokes case. Since the Biot-Savart law is divergence free, by equation 53 in the proof of Lemma 1, one has

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\Pi^d} \rho_t \log \frac{\rho_t}{\bar{\rho}_t} = -\nu \int_{\Pi^d} \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 - \int_{\Pi^d} \rho_t K * (\rho_t - \bar{\rho}_t) \cdot \nabla \log \bar{\rho}_t + \int_{\Pi^d} \rho_t \delta_t \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t}. \tag{54}$$

Recall that we write the kernel $K = (K_1, \cdots, K_d)$ and its component $K_i = \sum_{j=1}^d \partial_{x_j} U_{ij}(x)$, where $U = (U_{ij})_{1 \le i,j \le d}$ is a matrix-valued potential function for instance can be defined as in equation 19. Consequently

$$-\int \rho_t K * (\rho_t - \bar{\rho}_t) \cdot \nabla \log \bar{\rho}_t = -\sum_{i,j=1}^d \int \rho_t \partial_{x_j} U_{ij} * (\rho_t - \bar{\rho}_t) \partial_{x_i} \log \bar{\rho}_t,$$

which equals to

$$\sum_{i,j=1}^d \int U_{ij} * (\rho_t - \bar{\rho}_t) \partial_{x_j} \big(\frac{\rho_t}{\bar{\rho}_t} \partial_{x_i} \bar{\rho}_t\big) = A + B$$

by integration by parts, where further

$$A = \sum_{i,j=1}^d \int V_{ij} * (\rho_t - \bar{\rho}_t)\, \partial_{x_i} \bar{\rho}_t\, \partial_{x_j} \frac{\rho_t}{\bar{\rho}_t} = \int U * (\rho_t - \bar{\rho}_t) : \nabla \bar{\rho}_t \otimes \nabla \frac{\rho_t}{\bar{\rho}_t},$$

and

$$B = \sum_{i,j=1}^d \int \rho_t U_{ij} * (\rho_t - \bar{\rho}_t) \frac{\partial^2_{x_i x_j} \bar{\rho}_t}{\bar{\rho}_t} = \int \rho_t U * (\rho_t - \bar{\rho}_t) : \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t}.$$

Noticing that $\nabla \frac{\rho_t}{\bar{\rho}_t} = \frac{\rho_t}{\bar{\rho}_t} \nabla \log \frac{\rho_t}{\bar{\rho}_t}$, one estimates A as follows

$$\begin{aligned}
A &= \int \rho_t U * (\rho_t - \bar{\rho}_t) : \nabla \log \bar{\rho}_t \otimes \nabla \log \frac{\rho_t}{\bar{\rho}_t} \\
&\le \frac{\nu}{4} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{1}{\nu} \int \rho_t |(\nabla \log \bar{\rho}_t)^\top U * (\rho - \bar{\rho})|^2 \\
&\le \frac{\nu}{4} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{1}{\nu} \|U\|^2_{L^\infty} \|\nabla \log \bar{\rho}_t\|^2_{L^\infty} \|\rho_t - \bar{\rho}_t\|^2_{L^1},
\end{aligned}$$

and again by Csiszár–Kullback–Pinsker inequality, one has that

$$A \le \frac{\nu}{4} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{2}{\nu} \|U\|^2_{L^\infty} \|\nabla \log \bar{\rho}_t\|^2_{L^\infty} \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t}.$$

Now it only remains to control B. Recall the following famous Gibbs inequality

Lemma 4 (Gibbs inequality)**.** *For any parameter $\eta > 0$, and probability measures $\rho, \bar{\rho} \in \mathcal{P}(\mathcal{X}) \cap L^1(\mathcal{X})$, and ϕ a real-valued function defined on $\mathcal{X}$, one has the following change of reference measure inequality*

$$\int_{\mathcal{X}} \rho(x)\phi(x)\mathrm{d}x \leq \frac{1}{\eta}\Big(\int_{\mathcal{X}} \rho(x) \log \frac{\rho(x)}{\bar{\rho}(x)}\mathrm{d}x + \log \int_{\mathcal{X}} \bar{\rho}(x) \exp(\eta\phi(x))\mathrm{d}x\Big).$$

The proof of this inequality can be found in section 13.1 in [Erdős and Yau, 2017].

To control B, we write that $\phi = U * (\rho_t - \bar{\rho}_t) : \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t}$ and thus $B = \int \rho_t \phi$. We choose a positive parameter $\eta > 0$ such that

$$\frac{1}{\eta} = 2\|U\|_{L^\infty} \Big\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \Big\|_{L^\infty}.$$

Now we apply Lemma 4 to obtain that

$$B = \int \rho_t \phi \leq \frac{1}{\eta} \left(\int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} + \log \int \bar{\rho}_t \exp(\eta\phi) \right).$$

Note that $\eta > 0$ is chosen so small such that

$$\begin{aligned} \eta\|\phi\|_{L^\infty} &\leq \frac{1}{2\|U\|_{L^\infty} \Big\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \Big\|_{L^\infty}} \|U\|_{L^\infty} \|\rho_t - \bar{\rho}_t\|_{L^1} \Big\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \Big\|_{L^\infty} \\ &\leq \frac{1}{2}\|\rho_t - \bar{\rho}_t\|_{L^1} \leq 1, \end{aligned}$$

since for two probability densities it always holds $\|\rho_t - \bar{\rho}_t\|_{L^1} \leq 2$. Consequently, applying the inequality $\exp(x) \leq 1 + x + \frac{e}{2}x^2$ for $|x| \leq 1$, we have

$$\int \bar{\rho}_t \exp(\eta\phi) \leq \int \bar{\rho}_t \left(1 + \eta\phi + \frac{e}{2}\eta^2\phi^2\right) \leq 1 + 0 + \frac{e}{2}\Big(\frac{1}{2}\|\rho_t - \bar{\rho}_t\|_{L^1}\Big)^2 \leq 1 + \frac{e}{4}\mathbf{KL}(\rho_t|\bar{\rho}_t),$$

where

$$\int \bar{\rho}_t \phi = \int U * (\rho_t - \bar{\rho}_t) : \nabla^2 \bar{\rho}_t = \int \sum_{i,j=1}^{d} \partial_{x_i x_j} U * (\rho_t - \bar{\rho}_t)\bar{\rho}_t = \int \mathrm{div} K * (\rho_t - \bar{\rho}_t)\bar{\rho}_t = 0,$$

since $\mathrm{div} K = 0$.

To sum it up, in particular since $\log(1 + x) \leq x$ for $x > 0$, one has

$$B \leq \frac{1}{\eta}\Big(1 + \frac{e}{4}\Big)\mathbf{KL}(\rho_t|\bar{\rho}_t) \leq 4\|U\|_{L^\infty} \Big\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \Big\|_{L^\infty} \mathbf{KL}(\rho_t|\bar{\rho}_t).$$

Combining equation 54, the estimates for A and B, one has

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t} \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} \leq -\frac{3\nu}{4} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + M(t) \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} + \int \rho_t \delta_t \cdot \nabla \log \frac{\rho_t}{\bar{\rho}_t} \\ &\leq -\frac{\nu}{2} \int \rho_t |\nabla \log \frac{\rho_t}{\bar{\rho}_t}|^2 + M(t) \int \rho_t \log \frac{\rho_t}{\bar{\rho}_t} + \frac{1}{\nu} \int \rho_t |\delta_t|^2 \end{aligned} \tag{55}$$

where

$$M(t) = \frac{2}{\nu}\|U\|_{L^\infty}^2 \|\nabla \log \bar{\rho}_t\|_{L^\infty}^2 + 4\|U\|_{L^\infty} \Big\| \frac{\nabla^2 \bar{\rho}_t}{\bar{\rho}_t} \Big\|_{L^\infty} = M(t; \nu, U, \bar{\rho}_t).$$

Since the matrix-valued potential function U is bounded, and under suitable assumptions for the initial data $\bar{\rho}_0$(for instance $\bar{\rho}_0 \in C^3$ and there exists $c > 1$ s.t. $\frac{1}{c} \leq \bar{\rho} \leq c$), one can obtain $\sup_{t\in[0,T]} M(t) \leq M < \infty$. We recall Theorem 2 in [Guillin et al., 2021] as below for completeness.

Theorem 4. *Given the initial data $\bar{\rho}_0 \in C^\infty(\Pi^d)$, such that there exists $c > 1$, $\frac{1}{c} \leq \bar{\rho}_0 \leq c$. Then the vorticity formulation of the 2D Navier-Stokes equation*

$$\partial_t \bar{\rho}_t + \mathrm{div}(\bar{\rho}_t K * \bar{\rho}_t) = \nu\Delta\bar{\rho}_t, \quad \bar{\rho}(0, x) = \bar{\rho}_0(x),$$

has a unique bounded solution $\bar{\rho}(t, x) \in C^\infty([0, \infty) \times \Pi^d)$, and for any $t > 0$, for any $x \in \Pi^d$, it holds that $\frac{1}{c} \leq \bar{\rho}(t, x) \leq c$.

Finally, we simplify Eq. equation 55 to obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int \rho_t \log\frac{\rho_t}{\bar{\rho}_t} \le M\int \rho_t \log\frac{\rho_t}{\bar{\rho}_t} + \frac{1}{\nu}\int \rho_t|\delta_t|^2,$$

where $M = \sup_{t\in[0,T]} M(t;\nu,U,\bar{\rho}_t) < \infty$. By Gronwall inequality, one finally obtains that

$$\sup_{t\in[0,T]}\int_{\Pi^d} \rho_t \log\frac{\rho_t}{\bar{\rho}_t}\mathrm{d}x \le \frac{1}{\nu}\exp(MT)R(\theta).$$

This completes the proof of Theorem 2

Remark 1. *As noted in [Guillin et al., 2021], one can improve the above time-dependent estimate (*$\exp(MT)$*) to uniform-in-time estimate by using Logarithmic Sobolev inequality. Since now we are only care about computing solutions in a fixed time interval* $[0,T]$*, we do not seek to optimize the factor* $\exp(MT)$*. We leave the long time asymptotic analysis as a separate work.*

The McKean-Vlasov PDEs, i.e. equation 2, with bounded interactions $K \in L^\infty$ As mentioned in the main body of this article, it is much easier to obtain the stability estimate for the McKean-Vlasov PDE with bounded interactions.

Theorem 5 (Stability Estimate for McKean-Vlasov PDE with $K \in L^\infty$)**.** *Assume that* $K \in L^\infty$*. One has the estimate that*

$$\sup_{t\in[0,T]} \mathbf{KL}(\rho_t^f|\bar{\rho}_t) \le \frac{1}{\nu}\exp\Big(\frac{2\|K\|_{L^\infty}^2}{\nu}T\Big)R(f),$$

where we recall the self-consitency potential/loss function $R(\theta)$ *reads*

$$R(f) = \int_0^T\int_{\mathcal{X}} |f(t,x) + \nabla V(x) - K*\rho_t^f + \nu\nabla\log\rho_t^\theta|^2\mathrm{d}\rho_t^f(x)\mathrm{d}t.$$

Proof. Here we give the control of the growth of the KL divergence for systems with bounded kernels. Applying Cauchy-Schwarz inequality twice for the entropy dissipation terms in Lemma 1 to obtain

$$\int_{\Pi^d} \rho_t K*(\rho_t-\bar{\rho}_t)\cdot\nabla\log\frac{\rho_t}{\bar{\rho}_t} \le \frac{\nu}{4}\int\rho_t|\nabla\log\frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{1}{\nu}\int\rho_t|K*(\rho_t-\bar{\rho}_t)|^2,$$

and

$$\int_{\Pi^d} \rho_t\delta_t\cdot\nabla\log\frac{\rho_t}{\bar{\rho}_t} \le \frac{\nu}{4}\int\rho_t|\nabla\log\frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{1}{\nu}\int\rho_t|\delta_t|^2.$$

Furthermore,

$$\int\rho_t|K*(\rho_t-\bar{\rho}_t)|^2 \le \|K\|_{L^\infty}^2\|\rho_t-\bar{\rho}_t\|_{L^1}^2 \le 2\|K\|_{L^\infty}^2\int\rho_t\log\frac{\rho_t}{\bar{\rho}_t},$$

where the last inequality is simply the Csiszár–Kullback–Pinsker inequality [Villani et al., 2009]. Combining the above estimates, we obtain that given that $K \in L^\infty$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Pi^d}\rho_t\log\frac{\rho_t}{\bar{\rho}_t} = -\frac{\nu}{2}\int_{\Pi^d}\rho_t|\nabla\log\frac{\rho_t}{\bar{\rho}_t}|^2 + \frac{2\|K\|_{L^\infty}^2}{\nu}\int\rho_t\log\frac{\rho_t}{\bar{\rho}_t} + \frac{1}{\nu}\int\rho_t|\delta_t|^2.$$

Currently, we are not interested in the long time behavior, so we first ignore the negative term above to obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Pi^d}\rho_t\log\frac{\rho_t}{\bar{\rho}_t} \le \frac{2\|K\|_{L^\infty}^2}{\nu}\int\rho_t\log\frac{\rho_t}{\bar{\rho}_t} + \frac{1}{\nu}\int\rho_t|\delta_t|^2.$$

By Gronwall inequality, we obtain that

$$\int_{\Pi^d}\rho_t\log\frac{\rho_t}{\bar{\rho}_t} \le \frac{1}{\nu}\exp\Big(\frac{2\|K\|_{L^\infty}^2}{\nu}t\Big)\int_0^t\int\rho_s|\delta_s|^2\mathrm{d}x\mathrm{d}s.$$

□

C.2 The McKean-Vlasov equation with Coulomb interactions

Proof of Theorem 3. We first prove the case when $\nu > 0$. Applying Cauchy-Schwarz inequality to the right-hand side of $\frac{\mathrm{d}}{\mathrm{d}t}E(\rho_t^f, \bar{\rho}_t)$ in Lemma 3, one has

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}E(\rho_t^f, \bar{\rho}_t) \leq \frac{1}{2}\int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 \mathrm{d}x \\ - \frac{1}{2}\int_{\mathcal{X}^2} K(x-y)\cdot\Big(\mathcal{A}[\bar{\rho}_t](x) - \mathcal{A}[\bar{\rho}_t](y)\Big)\mathrm{d}(\rho_t^f - \bar{\rho}_t)^{\otimes 2}(x,y).\end{aligned}$$

By Lemma 5.2 in Bresch et al. [2019b], as long as the ground truth "velocity field" $\mathcal{A}[\bar{\rho}_t]$ is Lipschitz, i.e. $\mathcal{A}[\bar{\rho}] \in W^{1,\infty}$, or equivalently $\nabla^2 V \in W^{1,\infty}, \nabla^2 \log \bar{\rho}_t \in L^\infty, K * \bar{\rho}_t \in W^{1,\infty}$, using the particular structure introduced by the Coulomb interactions (note that $-\Delta g = \delta_0$ and $K = -\nabla g$), we have the estimate

$$\begin{aligned}&- \frac{1}{2}\int_{\mathcal{X}^2} K(x-y)\cdot\Big(\mathcal{A}[\bar{\rho}_t](x) - \mathcal{A}[\bar{\rho}_t](y)\Big)\mathrm{d}(\rho_t^\theta - \bar{\rho}_t)^{\otimes 2}(x,y) \\ &\leq C\|\nabla\mathcal{A}[\bar{\rho}_t]\|_{L^\infty} F(\bar{\rho}_t^f, \bar{\rho}_t).\end{aligned}$$

This estimate can be obtained either by Fourier method [Bresch et al., 2019b] or by the stress-energy tensor approach as in [Serfaty, 2020]. We emphasize that those assumptions made on $(\bar{\rho}_t)_{t\in[0,T]}$ can be obtained by propagating similar conditions on the initial data $\bar{\rho}_0$. This estimate actually holds for more general choices of g or K. See more examples including Riesz kernels in [Bresch et al., 2019b]. Moreover, the Lipschitz regularity of $\mathcal{A}[\bar{\rho}_t]$ can also be relaxed a bit. See for instance in [Rosenzweig, 2022].

Combining previous two estimates, one has

$$\frac{\mathrm{d}}{\mathrm{d}t}E(\rho_t^f, \bar{\rho}_t) \leq \frac{1}{2}\int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 \mathrm{d}x + CC_1 F(\rho_t^f, \bar{\rho}_t) \leq \frac{1}{2}\int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 \mathrm{d}x + CC_1 E(\rho_t^f, \bar{\rho}_t).$$

Then applying Gronwall inequality concludes the proof of the case when $\nu > 0$.

Now we prove the deterministic case when $\nu = 0$. Now the relative entropy or KL divergence does not play a role since there is no Laplacian term in equation 2. Lemma 2 now reads

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}F(\rho_t^f, \bar{\rho}_t) = -\int_{\mathcal{X}} \rho_t^f |K * (\rho_t^f - \bar{\rho})|^2 - \int_{\mathcal{X}} \rho_t^f \, \delta_t \cdot K * (\rho_t^f - \bar{\rho}_t) \\ - \frac{1}{2}\int_{\mathcal{X}^2} K(x-y)\cdot\Big(\mathcal{A}[\bar{\rho}_t](x) - \mathcal{A}[\bar{\rho}_t](y)\Big)\mathrm{d}(\rho_t^f - \bar{\rho}_t)^{\otimes 2}(x,y).\end{aligned}$$

Applying Cauchy-Schwarz to the 2nd term in the right-hand side above, we obtain that

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\rho_t^f, \bar{\rho}_t) \leq \frac{1}{2}\int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 - \frac{1}{2}\int_{\mathcal{X}^2} K(x-y)\cdot\Big(\mathcal{A}[\bar{\rho}_t](x) - \mathcal{A}[\bar{\rho}_t](y)\Big)\mathrm{d}(\rho_t^f - \bar{\rho}_t)^{\otimes 2}(x,y).$$

Again assuming that the "velocity field" $\mathcal{A}[\bar{\rho}_t](\cdot)$ is Lipschitz will give us

$$\frac{\mathrm{d}}{\mathrm{d}t}F(\rho_t^f, \bar{\rho}_t) \leq \frac{1}{2}\int_{\mathcal{X}} \rho_t^f \, |\delta_t|^2 + CC_1 F(\rho_t^f, \bar{\rho}_t).$$

Applying Gronwall inequality again conclude all the proof.

□