On Propagation of Chaos for the Fisher-Rao Gradient Flow in Entropic Mean-Field Optimization

Petra Lazić
Faculty of Mathematics and Physics, University of Ljubljana and University of Zagreb

Linshan Liu
Heriot-Watt University and Maxwell Institute for Mathematical Sciences

Mateusz B. Majka
Heriot-Watt University and Maxwell Institute for Mathematical Sciences

Abstract

We consider a class of optimization problems on the space of probability measures motivated by the mean-field approach to studying neural networks. Such problems can be solved by constructing continuous-time gradient flows that converge to the minimizer of the energy function under consideration, and then implementing discrete-time algorithms that approximate the flow. In this work, we focus on the Fisher-Rao gradient flow and we construct an interacting particle system that approximates the flow as its mean-field limit. We discuss the connection between the energy function, the gradient flow and the particle system and explain different approaches to smoothing out the energy function with an appropriate kernel in a way that allows for the particle system to be well-defined. We provide a rigorous proof of the existence and uniqueness of thus obtained kernelized flows, as well as a propagation of chaos result that provides a theoretical justification for using the corresponding kernelized particle systems as approximation algorithms in entropic mean-field optimization.

1 INTRODUCTION

We consider the following optimization problem on the space of probability measures $\mathcal{P}(\mathcal{X})$ on $\mathcal{X} \subset \mathbb{R}^d$

$$\min_{m \in \mathcal{P}(\mathcal{X})} V^\sigma(m), \quad V^\sigma(m) := F(m) + \sigma \, \mathrm{KL}(m|\pi), \quad (1)$$

where $F : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ is a (possibly non-convex) functional bounded from below, $\sigma > 0$ is a regularization parameter, $\pi \in \mathcal{P}(\mathcal{X})$ is a fixed reference measure and KL denotes the relative entropy (the KL-divergence). While some general results in Section 2 will be stated for domains $\mathcal{X} \subset \mathbb{R}^d$ which do not have to be compact, for the crucial examples studied in Section 3 we will additionally require $\mathcal{X}$ to be compact. In recent years, there has been considerable interest in problems of this type, motivated by the mean-field approach to the problem of training neural networks (see Mei et al. (2018) or (Hu et al., 2021, Section 3) and the references therein), as well as in the context of reinforcement learning, in policy optimization for entropy-regularized Markov Decision Processes with neural network approximation in the mean-field regime (Leahy et al., 2022; Lascu and Majka, 2025).

In order to solve (1), one typically aims to construct a gradient flow $(\mu_t)_{t \geq 0}$ on $\mathcal{P}(\mathcal{X})$ that converges to the minimizer $m^{*,\sigma}$ of (1) when $t \to \infty$. The most commonly used example is the Wasserstein gradient flow

$$\partial_t \mu_t = \nabla \cdot \left(\mu_t \nabla \frac{\delta V^\sigma}{\delta \mu}(\mu_t, \cdot) \right) \quad (2)$$

defined via the flat derivative (first variation) $\frac{\delta V^\sigma}{\delta \mu}$ of the energy function V^σ (see Definition D.1). The conditions guaranteeing the convergence of (2) to $m^{*,\sigma}$ have been studied by numerous authors in various settings, including Ambrosio et al. (2008); Hu et al. (2021); Nitanda et al. (2022); Chizat (2022); Leahy et al. (2022) and many others.

However, from the point of view of applications, an equally important question is how to approximate gradient flows such as (2) by a practically implementable algorithm. One possible approach leads through the Jordan-Kinderlehrer-Otto (JKO) schemes (see Jordan et al. (1998) for the original paper or Salim et al. (2020); Lascu et al. (2024) for more recent expositions). Another, which we are going to focus on in the present paper, utilizes an interpretation of (2) as

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

the mean-field limit of an interacting particle system. In the latter approach, one typically aims to prove a propagation of chaos result, i.e., a result that shows that as the number of particles approaches infinity, particles become independent and they all follow the same mean-field dynamics. This can be then used as a theoretical justification that an appropriately constructed interacting particle system may be used to produce (after a discretisation) an algorithm that approximates the minimizer of (1) when the number of particles and the number of iterations are both sufficiently large.

Propagation of chaos for the Wasserstein gradient flow has been studied in detail in various settings (utilizing the interpretation of (2) as the Fokker-Planck equation for the mean-field Langevin SDE). A (far from complete) list of references includes Chen et al. (2025); Monmarché et al. (2024); Durmus et al. (2020); Carrillo et al. (2003); Malrieu (2001); Delarue and Tse (2025); Lacker and Le Flem (2023); Suzuki et al. (2023); Nitanda (2024); Nitanda et al. (2025); Gu and Kim (2025). A related active strain of research involves the propagation of chaos for kinetic models Monmarché (2017); Guillin and Monmarché (2021); Schuh (2024); Chen et al. (2024).

In the present work, we focus on a different gradient flow, the so-called Fisher-Rao gradient flow given by

$$\partial_t \mu_t = -\mu_t \frac{\delta V^\sigma}{\delta \mu}(\mu_t, \cdot). \tag{3}$$

The interest in studying this flow in the context of optimization problems (1) is motivated by the fact that in some settings, its convergence to the minimizer could be easier to verify than for the Wasserstein flow Liu et al. (2023); Kerimkulov et al. (2025); Lascu et al. (2024). There has been a considerable literature studying the fundamental properties of Fisher-Rao gradient flows such as well-posedness in various settings, see e.g. Carrillo et al. (2024); Zhu and Mielke (2024) and the references therein, also in combination with the Wasserstein flow as the Wasserstein-Fisher-Rao gradient flow (also known as Hellinger-Kantorovich) Liero et al. (2018); Gallouët and Monsaingeon (2017); Lu et al. (2019); Rotskoff et al. (2019).

However, unlike for the Wasserstein gradient flow (2), in the context of the Fisher-Rao flow (3) much less is known about particle approximations (with a few exceptions that will be discussed in detail in Section 2.4). In particular, the main goal of the present paper is to fill a gap in the literature by providing a rigorous proof for a particle approximation for the Fisher-Rao gradient flow (3) corresponding to a class of minimization problems of the form (1).

1.1 Contributions

We summarize the main contributions of our paper:

- In Section 2, we propose a general framework for studying particle approximations of Fisher-Rao gradient flows. This part extends the results from Cavil et al. (2017); Lu et al. (2019); Rotskoff et al. (2019); Domingo-Enrich et al. (2020) (see Remark 2.8 and Section 2.4 for details).
- In Section 3, we discuss different approaches to smoothing out Fisher-Rao flows (which is crucial for practical implementation) and we show that all the discussed methods fall within our framework from Section 2. This part expands on Lu et al. (2019); Pampel et al. (2023); Lu et al. (2023); Carrillo et al. (2019).
- In Section 4, we propose a practical algorithm for approximating Fisher-Rao flows, utilizing a method from Section 3.

2 MAIN RESULTS

In this section, we work on a general (not necessarily compact) space $\mathcal{X} \subset \mathbb{R}^d$ and we focus on the flow

$$\partial_t \mu_t = -\mu_t a(\mu_t, \cdot) \tag{4}$$

where the flat derivative $\frac{\delta V^\sigma}{\delta \mu}$ on the right hand side of (3) has been replaced with a function $a : \mathcal{P}(\mathcal{X}) \times \mathcal{X} \to \mathbb{R}$. In Section 3, we will explain the rationale behind approximating the flat derivative $\frac{\delta V^\sigma}{\delta \mu}$ of the energy function V^σ in (1) with its kernelized counterpart. Different choices of kernelizations will lead to different choices of a and hence in this section we keep our notation general, in order to produce a broadly applicable theoretical framework. For convenience, we will refer to flows (4) as Fisher-Rao flows, although it should be understood that they are "true" Fisher-Rao flows only for some choices of a (see Section 3.2 for more details).

We remark that we always choose a in a way that ensures that μ_t remains a probability distribution. This is achieved by requiring that for all $\mu \in \mathcal{P}(\mathcal{X})$,

$$\int_{\mathcal{X}} a(\mu, x)\mu(dx) = 0, \tag{5}$$

which then immediately implies that $\partial_t \int_{\mathcal{X}} \mu_t(dx) = -\int_{\mathcal{X}} a(\mu_t, x)\mu_t(dx) = 0$, and hence the flow preserves the mass of the initial measure. It is easy to check that all examples of a studied in Section 3 satisfy property (5).

2.1 Preliminaries

Before we proceed, let us introduce some necessary notation. For $p \in [1, \infty)$, and for any normed vector space $(E, \|\cdot\|)$, we define the set of probability measures with finite p-th moment $\mathcal{P}_p(E)$ as

$$\mathcal{P}_p(E) := \left\{ \mu \in \mathcal{P}(E) : \int_E \|x\|^p \mu(dx) < \infty \right\}.$$

We equip $\mathcal{P}_p(E)$ with the p-Wasserstein distance

$$\mathcal{W}_p(m, m') := \left(\inf_{\pi \in \Pi(m,m')} \int_E \|x - y\|^p \, \pi(dx, dy) \right)^{1/p},$$

where $\Pi(m, m') \subset \mathcal{P}(E \times E)$ is the set of all couplings of m and m'.

For any $T > 0$, we consider the path space $\mathcal{C}([0, T]; E)$ with the supremum norm

$$\|x - y\|_T := \sup_{0 \le s \le T} \|x_s - y_s\|,$$

and write $\mathcal{P}_p(\mathcal{C}([0, T]; E))$ for the corresponding p-moment space. The induced p-Wasserstein distance on path measures is

$$\mathcal{W}_{p,T}(m, m') := \left(\inf_{\pi \in \Pi(m,m')} \int \|x - y\|_T^p \, \pi(dx, dy) \right)^{1/p}.$$

We require a to satisfy the following assumption:

Assumption 1. Function a is bounded and Lipschitz, i.e., there exist constants M_a, $L_a > 0$ such that for all $x, y \in \mathcal{X}$ and $m, m' \in \mathcal{P}_2(\mathcal{X})$,

$$|a(m, x)| \le M_a, \tag{6}$$

and

$$|a(m, x) - a(m', y)| \le L_a \big(|x - y| + \mathcal{W}_2(m, m') \big). \tag{7}$$

Following the ideas from Liero et al. (2018) (see also Domingo-Enrich et al. (2020)), we work with the notions of lifts and projections of measures. This will allow us to use measures defined on the extended space $\mathcal{X} \times \mathbb{R}_+$, with the second component representing a local weight.

Definition 2.1 (Lifted measure and projection)**.** Let $\mu \in \mathcal{P}_1(\mathcal{X})$ and $\nu \in \mathcal{P}_1(\mathcal{X} \times \mathbb{R}_+)$. We say that ν is a lifted measure of μ, and conversely that μ is the projection of ν, if for all $\varphi \in \mathcal{C}_c^\infty(\mathcal{X})$,

$$\int_{\mathcal{X}} \varphi(x) \, d\mu(x) = \int_{\mathcal{X} \times \mathbb{R}_+} w \, \varphi(x) \, d\nu(x, w), \tag{8}$$

where $\mathcal{C}_c^\infty(\mathcal{X})$ denotes the space of smooth, compactly supported functions on $\mathcal{X}$. In this case we define the projection operator

$$\mathtt{h} : \mathcal{P}_1(\mathcal{X} \times \mathbb{R}_+) \to \mathcal{P}_1(\mathcal{X}),$$

so that $\mu = \mathtt{h}\nu$ whenever (8) holds.

Note that the requirement for the measures in Definition 2.1 to have finite first moments ensures that the integral on the right hand side of (8) is finite.

2.2 Existence of mean-field dynamics corresponding to Fisher-Rao flow

Building on the notion of lifted and projected measures defined above, we now establish a rigorous connection between the Fisher–Rao gradient flow and a corresponding mean-field (single particle) dynamics. The key idea is to lift the flow from the probability space $\mathcal{P}_1(\mathcal{X})$ to the extended space $\mathcal{P}_1(\mathcal{X} \times \mathbb{R}_+)$, where the additional coordinate w represents a local mass weight. This lifted formulation reveals that the original Fisher–Rao flow can be interpreted as the projection of an evolution equation on the extended space. Based on this representation, we construct a mean-field dynamic whose law evolves according to the lifted flow, and show that, under suitable regularity assumptions, this dynamic admits a unique solution on any finite time horizon. We begin by stating the precise correspondence between the lifted and projected flows.

Definition 2.2. Let $\nu_0 \in \mathcal{P}_1(\mathcal{X} \times \mathbb{R}_+)$. We say that $(\nu_t)_{t \ge 0} \subset \mathcal{P}_1(\mathcal{X} \times \mathbb{R}_+)$ is a weak solution to the lifted flow

$$\partial_t \nu_t(x, w) = \frac{\partial}{\partial w} \big(\nu_t(x, w) \, w \, a(\mathtt{h}\nu_t, x) \big), \tag{9}$$

with initial condition ν_0, if for any $\psi \in \mathcal{C}_c^\infty(\mathcal{X} \times \mathbb{R}_+)$ and any $t > 0$ we have

$$\begin{aligned} &\frac{d}{dt} \int_{\mathcal{X} \times \mathbb{R}_+} \psi(x, w) \, d\nu_t(x, w) \\ &= - \int_{\mathcal{X} \times \mathbb{R}_+} w \, a(\mathtt{h}\nu_t, x) \, \frac{\partial}{\partial w} \psi(x, w) \, d\nu_t(x, w). \end{aligned} \tag{10}$$

Definition 2.3. Let $\mu_0 \in \mathcal{P}_1(\mathcal{X})$. We say that $(\mu_t)_{t \ge 0} \subset \mathcal{P}_1(\mathcal{X})$ is a weak solution to the Fisher-Rao gradient flow

$$\partial_t \mu_t = -\mu_t \, a(\mu_t, \cdot), \tag{11}$$

with initial condition μ_0, if for any $\psi \in \mathcal{C}_c^\infty(\mathcal{X})$ and any $t > 0$ we have

$$\frac{d}{dt} \int_{\mathcal{X}} \psi(x) \, d\mu_t(x) = - \int_{\mathcal{X}} \psi(x) \, a(\mu_t, x) \, d\mu_t(x). \tag{12}$$

Remark 2.4. Note that by a standard approximation argument, if $(\nu_t)_{t\geq 0}$ is a weak solution to the lifted flow in the sense of Definition 2.2, then (10) holds also for functions of the form

$$\psi(x,w) = w\,\phi(x), \quad \phi \in \mathcal{C}_c^\infty(\mathcal{X}).$$

This will be important for arguments where we switch between the lifted and projected flows, especially in the proof of the following lemma.

Lemma 2.5. *Let $(\nu_t)_{t\geq 0} \subset \mathcal{P}_1(\mathcal{X}\times\mathbb{R}_+)$ be a weak solution to the lifted flow* (9) *in the sense of Definition 2.2, with initial condition $\nu_0 \in \mathcal{P}_1(\mathcal{X}\times\mathbb{R}_+)$. Then the projected flow $(\mu_t)_{t\geq 0} := (\hbar\nu_t)_{t\geq 0} \subset \mathcal{P}_1(\mathcal{X})$ is a weak solution to the Fisher–Rao flow* (11) *in the sense of Definition 2.3, with initial condition $\mu_0 := \hbar\nu_0$.*

With this relation at hand, we can construct a mean-field particle dynamic whose law evolves according to the lifted flow. The precise statement is given in the following theorem.

Theorem 2.6. *Let $\nu_0 \in \mathcal{P}_1(\mathcal{X}\times\mathbb{R}_+)$. Consider the mean-field system*

$$\begin{aligned} dX_t &= 0, \\ dw_t &= -w_t\, a\big(\hbar\nu_t, X_t\big)\, dt, \\ (X_0, w_0) &\sim \nu_0, \end{aligned} \tag{13}$$

where $\nu_t := \mathrm{Law}(X_t, w_t)$ denotes the joint law of (X_t, w_t). If system (13) *admits a solution, then the curve $(\nu_t)_{t\geq 0}$ is a weak solution to the lifted Fisher–Rao flow* (9) *in the sense of Definition 2.2, and hence $(\mu_t)_{t\geq 0} := (\hbar\nu_t)_{t\geq 0}$ is a weak solution to the Fisher–Rao flow* (11) *in the sense of Definition 2.3.*

Remark 2.7. System (13) can be interpreted as a particle X_t with a corresponding weight w_t. Note that the spatial position X_t of the particle is sampled once from the initial probability distribution and remains fixed over time, while only the associated weight w_t evolves according to the dynamic in (13). This reflects the geometry of the Fisher–Rao metric, which governs mass change without spatial transport.

Remark 2.8. Consider the setting corresponding to our primary object of interest, i.e., when the function a in (4) is given by the flat derivative of V^σ defined in (1). Then, for measures μ such that V^σ has a flat derivative at μ,

$$a(\mu,\cdot) = \frac{\delta F}{\delta\mu}(\mu,\cdot) + \sigma \log\frac{\mu}{\pi} \tag{14}$$

(up to an additive constant - the details will be discussed in Section 3). If we assume the necessary differentiability, we can consider the Wasserstein-Fisher-Rao gradient flow given by

$$\partial_t \mu_t(x) = \nabla\cdot(\mu_t \nabla a(\mu_t, x)) - \mu_t(x) a(\mu_t, x). \tag{15}$$

Then, due to the interpretation of the Wasserstein gradient flow as the mean-field Langevin SDE (Hu et al., 2021), we would expect to obtain the following corresponding mean-field (single particle) system

$$\begin{aligned} dX_t &= -\nabla\left(\frac{\delta F}{\delta\mu}(\mathtt{h}\nu_t, X_t) - \sigma\log\pi(X_t)\right) dt \\ &\quad + \sqrt{2\sigma}\, dW_t, \\ dw_t &= -w_t\, a(\mathtt{h}\nu_t, X_t)\, dt, \\ \nu_t &:= \mathrm{Law}(X_t, w_t), \end{aligned} \tag{16}$$

where both the location of the particle and the weight evolve in time. There are, however, several technical challenges with studying system (16). If (16) is considered on a compact space $\mathcal{X} \subset \mathbb{R}^d$, this creates an issue with the Wasserstein part of the flow (due to the diffusive behaviour of X_t) and one needs to make sure that the particle stays on the right space. On the other hand, if $\mathcal{X} = \mathbb{R}^d$, then the Fisher-Rao part becomes problematic since the function a defined in (14) is unbounded (due to the flat derivative of the relative entropy being unbounded). While working on the present paper, we were unable to overcome these technical challenges which is why we focus on the system (13) corresponding to the "pure" Fisher-Rao gradient flow (and in Section 3 when we discuss specific choices of a related to (14), we work on a compact $\mathcal{X}$ to ensure that a stays bounded). This is consistent with a recent paper (Domingo-Enrich et al., 2020), which studied (in the context of two-player game theory) Wasserstein-Fisher-Rao flows corresponding to energy functions (1) without entropy-regularization (i.e., with $\sigma = 0$), which leads to a system of the form (16) where X_t moves only according to a deterministic gradient descent; as well as Wasserstein flows corresponding to (1) with entropy regularization but without the Fisher-Rao part. Similarly to Domingo-Enrich et al. (2020), we are unable to cover Wasserstein-Fisher-Rao (WFR) flows corresponding to (1) with entropy-regularization, but unlike Domingo-Enrich et al. (2020), we study Fisher-Rao flows for (1) with entropy-regularization, which were not covered there. We will attempt to treat the challenging case of WFR flows in a future work.

We now address the well-posedness of mean-field particle dynamics.

Theorem 2.9. *Suppose that a satisfies Assumption 1. Let the initial law $\nu_0 \in \mathcal{P}_2(\mathcal{X}\times\mathbb{R}_+)$. Then, for any $T > 0$, there exists a unique random dynamic $(X_t, w_t)_{t\in[0,T]}$ that solves system* (13)*. In particular,*

$$\mathrm{Law}\left((X_t, w_t)_{t\in[0,T]}\right) \in \mathcal{P}_2\big(\mathcal{C}([0,T]; \mathcal{X}\times\mathbb{R}_+)\big).$$

Moreover, the solution is pathwise unique: if two solutions share the same initial condition almost surely, then they coincide for all $t \in [0,T]$ almost surely.

2.3 Interacting Particle Systems and Propagation of Chaos

We investigate the approximation of the Fisher–Rao gradient flow by interacting particle systems. Our objective is to establish propagation of chaos: as the number of particles $N \to \infty$, the empirical measure of the interacting system converges to the law of i.i.d. copies of the mean-field solution, with convergence measured in the 2-Wasserstein distance.

Before we discuss the interacting particle system that is our main object of interest, in order to make the notation precise we first introduce a system $(X^{i,N}, w_t^{i,N})$ of N i.i.d. copies of the mean-field dynamic (13), which will be used as an auxiliary tool in our approximation estimates. For each $i \in [\![1, N]\!]$, we define the dynamic of $(X^{i,N}, w_t^{i,N})$ by

$$\begin{aligned} &X^{i,N} \sim \mu_0^{i,N}, \\ &w_0^{i,N} \geq 0, \quad \text{with } \sum_{i=1}^N w_0^{i,N} = N, \\ &dw_t^{i,N} = -w_t^{i,N} a(\mathtt{h}\nu_t^i, X^{i,N})\, dt, \end{aligned} \tag{17}$$

where $\nu_t^i := \mathrm{Law}(X^{i,N}, w_t^{i,N})$ is the marginal law of the i-th particle. We denote the empirical measure of thus obtained *non-interacting* system by

$$\nu_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{(X^{i,N}, w_t^{i,N})}. \tag{18}$$

We will use this notation throughout our proofs in Appendix B.

Interacting particle system. We are now ready to introduce the interacting particle system, where the interaction is governed by weighted empirical distributions. For each $i \in [\![1, N]\!]$, we define $(\tilde{X}^{i,N}, \tilde{w}_t^{i,N})$ by

$$\begin{aligned} &\tilde{X}^{i,N} \sim \tilde{\mu}_0^{i,N}, \\ &\tilde{w}_0^{i,N} \geq 0, \quad \text{with } \sum_{i=1}^N \tilde{w}_0^{i,N} = N, \\ &d\tilde{w}_t^{i,N} = -\tilde{w}_t^{i,N}\, a\left(\tilde{\mu}_t^N, \tilde{X}^{i,N}\right)\, dt, \end{aligned} \tag{19}$$

where $\tilde{\mu}_t^N := \frac{1}{N}\sum_{i=1}^N \tilde{w}_t^{i,N} \delta_{\tilde{X}^{i,N}}$ is the weighted empirical distribution. Note that the empirical measure on the extended space is

$$\tilde{\nu}_t^N := \frac{1}{N} \sum_{i=1}^N \delta_{(\tilde{X}^{i,N}, \tilde{w}_t^{i,N})}, \quad \text{so that} \quad \tilde{\mu}_t^N = \mathtt{h}\tilde{\nu}_t^N. \tag{20}$$

We now state the main result on the propagation of chaos.

Theorem 2.10. *Suppose that the function a satisfies Assumption 1. Fix a finite time horizon $T > 0$ and let $\nu \in \mathcal{P}_2(\mathcal{C}([0,T]; \mathcal{X} \times \mathbb{R}_+))$ be the law of the single particle $(X, w_t) = (X, w_t)_{t\in[0,T]}$ defined by the mean-field dynamic (13). Choose particles $(X^{i,N}, w_t^{i,N})$ as i.i.d. copies of (X, w_t) as defined in (17) and let $(\tilde{X}^{i,N}, \tilde{w}_t^{i,N})$ be the interacting particle system defined by (19). Suppose further that the initial conditions satisfy*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\left|X^{i,N} - \tilde{X}^{i,N}\right|^2 + \left|w_0^i - \tilde{w}_0^{i,N}\right|^2\right] = 0. \tag{21}$$

Let $\tilde{\nu}^N \in \mathcal{P}_2(\mathcal{C}([0,T]; \mathcal{X} \times \mathbb{R}_+))$ be the empirical measure of the interacting particle system defined by (20). Then,

$$\lim_{N\to\infty} \mathbb{E}\left[\mathcal{W}_{2,T}(\tilde{\nu}^N, \nu)\right] = 0.$$

We remark that condition (21) is automatically satisfied when the initializations for both systems considered in Theorem 2.10 are identical.

Corollary 2.11. *Under the assumptions of Theorem 2.10, the projected empirical distribution $\tilde{\mu}_t^N := \mathtt{h}\tilde{\nu}_t^N$ converges in the 2-Wasserstein distance, uniformly on $[0,T]$, to the mean-field law $\mu_t := \mathtt{h}\nu_t$. That is, for any finite time horizon $T > 0$, we have*

$$\lim_{N\to\infty} \sup_{t\in[0,T]} \mathbb{E}[\mathcal{W}_2\left(\mu_t, \tilde{\mu}_t^N\right)] = 0.$$

We remark that Theorem 2.10 and Corollary 2.11 are non-quantitative, and obtaining convergence rates would require further work, which is beyond the scope of the present paper.

2.4 Discussion of other propagation of chaos results in the literature

The framework in Cavil et al. (2017) establishes propagation of chaos results for particle approximations of the following class of non-conservative nonlinear PDEs:

$$\begin{aligned} \partial_t v = &\sum_{i,j=1}^d \partial_{ij}^2 \left((\Phi\Phi^\top)_{i,j}(t,x,v)v\right) \\ &- \nabla \cdot (g(t,x,v)v) + \Lambda(t,x,v)v, \quad v(0,dx) = v_0(dx). \end{aligned} \tag{22}$$

In this equation, the first two terms correspond to the Wasserstein component of WFR flows, while the last term plays the role of a Fisher–Rao-type reaction. However, unlike the WFR flow (15), where both the transport and reaction components are derived from a single energy functional, here the different terms are

specified independently and may correspond to unrelated energies.

More importantly, the reaction term $\Lambda(t,x,v)v$ in Cavil et al. (2017) depends locally on the solution v: that is, Λ is evaluated using the pointwise value $v(t,x)$, without reference to the global structure of the distribution. This contrasts with our setting, where, as we will explain in Section 3, the reaction term involves a nonlocal dependence on the entire probability measure. Hence, PDE (22) does not in general conserve total mass: the solution $v(t,\cdot)$ is not necessarily a probability density for all t. In contrast, the nonlocal normalization we include ensures mass preservation at all times. Moreover, the regularity assumptions also differ, cf. Assumption 1 in Cavil et al. (2017) to our Assumption 1.

Another key distinction is that we work on a compact state space $\mathcal{X} \subset \mathbb{R}^d$, whereas the equations in Cavil et al. (2017) are defined on the whole space $\mathbb{R}^d$. Extending the results from Cavil et al. (2017) to compact domains would require working with appropriately defined boundary conditions for the PDE (22).

As we discussed in Remark 2.8, another related paper (Domingo-Enrich et al., 2020) studied propagation of chaos for Wasserstein-Fisher-Rao flows without entropy regularization, and for Wassesrstein flows corresponding to the entropy-regularized energy (1) (but without the Fisher-Rao part) and hence their results are not applicable to our setting.

Finally, there has been some work on the propagation of chaos for interacting particle systems defined with a killing-replication mechanism with exponential clocks (rather than with evolving weights - cf. also the discussion in Section 4) in Rotskoff et al. (2019); Lu et al. (2019). However, we were unable to make the proofs from those works fully rigorous in our setting (1), due to the unboundedness of the flat derivative of the relative entropy.

3 ENTROPIC MEAN-FIELD PROBLEMS

In this section, we assume $\mathcal{X}$ to be a compact subset of $\mathbb{R}^d$, and we focus on the energy function

$$V^\sigma(m) = F(m) + \sigma\,\mathrm{KL}(m|\pi)\,, \text{ for all } m \in \mathcal{P}(\mathcal{X}). \tag{23}$$

Note that the choice of a compact $\mathcal{X}$ removes the technical problem with flat differentiability of the KL-divergence, which on unbounded domains would have to be rigorously justified Liu et al. (2023); Aubin-Frankowski et al. (2022). Note also that we do not require convexity of $\mathcal{X}$ since the pure Fisher-Rao flow does not change the support of the initial measure, i.e., it evolves only the mass without any transport in the state space.

The key difficulty in dealing with Fisher-Rao flows (3) corresponding to the energy function (23) lies in the construction of the corresponding interacting particle system. In our setting, one can show that the flat derivative of V^σ is given by

$$\frac{\delta V^\sigma}{\delta \mu}(\mu_t, x) = \frac{\delta F}{\delta \mu}(\mu_t, x) + \sigma \log \frac{\mu_t(x)}{\pi(x)} - \sigma\,\mathrm{KL}(\mu_t|\pi). \tag{24}$$

Note that the term $-\sigma\,\mathrm{KL}(\mu_t|\pi)$ is necessary to ensure that the equation (3) is conservative, i.e., that all measures μ_t are indeed probability measures. However, this makes (24) well-defined only if the measures μ_t are absolutely continuous with respect to π, which creates problems with defining a particle system approximating (3). Indeed, one cannot evaluate the right hand side of (3) at the empirical measure of a corresponding particle system, which makes it necessary to define an auxiliary flow, where the problematic terms are replaced by their counterparts involving convolutions with appropriately defined kernels. This is the rationale behind the introduction of the so-called "kernelized" flows, where instead of the expression in (24), we work with its kernelized version.

3.1 Setting for entropic mean-field optimization

Throughout Section 3, we impose the following standing assumptions on the energy functional F, the mollifier kernel K_ε, and the reference measure π.

Assumption 2. We assume the energy functional $F : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ satisfies the following properties:

(i) F is lower semi-continuous with respect to weak convergence in $\mathcal{P}_2(\mathcal{X})$.

(ii) F is bounded from below: there exists $F_{\min} \in \mathbb{R}$ such that

$$F(m) \geq F_{\min} \quad \text{for all } m \in \mathcal{P}(\mathcal{X}).$$

(iii) Flat derivative is bounded: there exists a constant $C_F > 0$ such that for all $\mu \in \mathcal{P}(\mathcal{X})$ and $x \in \mathcal{X}$,

$$\left|\frac{\delta F}{\delta \mu}(\mu, x)\right| \leq C_F.$$

(iv) Flat derivative is Lipschitz: there exists a constant $L_F > 0$ such that for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$,

$$\left|\frac{\delta F}{\delta \mu}(\mu, x) - \frac{\delta F}{\delta \mu}(\nu, y)\right| \leq L_F\left(\mathcal{W}_2(\mu,\nu) + |x-y|\right).$$

Note that these assumptions on F are standard in the mean-field optimization literature and, in particular, they are satisfied in mean-field models of neural networks studied in papers such as Hu et al. (2021); Chen et al. (2023), as well as in mean-field models in policy optimization in reinforcement learning (Leahy et al., 2022; Lascu and Majka, 2025).

Assumption 3. Let $\xi\colon \mathbb{R}^d \to \mathbb{R}_+$ be a smooth, Lipschitz (with constant L_ξ), radial probability density function with full support $\mathbb{R}^d$ and finite second moment. For $\varepsilon > 0$, define the rescaled function on the compact space $\mathcal{X}$ by

$$\xi_\varepsilon(x) := \frac{1}{C_{\varepsilon,d}}\, \varepsilon^{-d}\, \xi\Big(\frac{x}{\varepsilon}\Big)\,,$$

where $C_{\varepsilon,d}$ is the normalization constant ensuring that

$$\int_{\mathcal{X}} \xi_\varepsilon(x)\, dx = 1.$$

In particular, ξ_ε is a probability density function on $\mathcal{X}$. The mollifier kernel K_ε on $\mathcal{X}$ is then defined as

$$K_\varepsilon(x) := \big(\xi_\varepsilon * \xi_\varepsilon\big)(x).$$

Assumption 4. We assume the reference density $\pi(x)$ satisfies the following properties:

(i) $\pi(x) = e^{-U(x)}$ for some continuous potential function $U : \mathcal{X} \to \mathbb{R}$.

(ii) π is globally Lipschitz on $\mathcal{X}$: there exists $L_\pi > 0$ such that

$$|\pi(x) - \pi(y)| \le L_\pi\, |x - y| \quad \text{for all } x, y \in \mathcal{X}.$$

Note that under Assumption 4, since $\pi \propto e^{-U}$ and $\mathcal{X}$ is compact, there exist constants $0 < \pi_{\min} < \pi_{\max} < \infty$ such that

$$\pi_{\min} \le \pi(x) \le \pi_{\max}, \qquad \text{for all } x \in \mathcal{X}.$$

3.2 Kernelization Strategies

We summarize four kernelization strategies for approximating Fisher–Rao flows corresponding to energy functions (1), based on related strategies that have appeared in the literature in recent years. We will then show in Proposition 3.1 that for each of these approaches, the resulting function a satisfies our Assumption 1 and therefore all our results from Section 2 are applicable. For a general perspective on kernelization, see also Zhu and Mielke (2024). Recall our notation for the flow $\partial_t \mu_t = -\mu_t a(\mu_t, \cdot)$ and consider the following choices of a that approximate the flat derivative $\frac{\delta V^\sigma}{\delta \mu}$ given in (24).

Smoothing only the evolving measure (Lu et al., 2019). Here the kernel is applied to μ_t both inside the logarithm and in the KL divergence. The resulting dynamics read

$$\begin{aligned}\partial_t \mu_t = -\,\mu_t \Bigg(&\frac{\delta F}{\delta \mu}(\mu_t, \cdot) + \sigma\, \log \frac{K_\varepsilon * \mu_t}{\pi} \\ &- \sigma\, \mathrm{KL}(K_\varepsilon * \mu_t \mid \pi)\Bigg).\end{aligned} \tag{25}$$

Smoothing both the evolving and the target measures (Pampel et al., 2023). In this variant, both μ_t and π are mollified by K_ε, which yields

$$\begin{aligned}\partial_t \mu_t = -\,\mu_t \Bigg(&\frac{\delta F}{\delta \mu}(\mu_t, \cdot) + \sigma\, \log \frac{K_\varepsilon * \mu_t}{K_\varepsilon * \pi} \\ &- \sigma\, \mathrm{KL}(K_\varepsilon * \mu_t \mid K_\varepsilon * \pi)\Bigg).\end{aligned} \tag{26}$$

Kernelizing the energy via Lu et al. (2023). Here the kernel is applied already at the level of the energy function, rather than just at the level of the flow; one studies the "true" Fisher–Rao gradient flow corresponding to the modified energy function

$$V_\varepsilon^\sigma(m) = F(m) \;+\; \sigma \int \log \frac{K_\varepsilon * m}{\pi}\; m(dx).$$

The induced dynamics takes the form

$$\begin{aligned}\partial_t \mu_t = -\,\mu_t \sigma \Bigg(&\frac{1}{\sigma} \frac{\delta F}{\delta \mu}(\mu_t, \cdot) + \log \frac{K_\varepsilon * \mu_t}{\pi} \\ &+ K_\varepsilon * \Big(\frac{\mu_t}{K_\varepsilon * \mu_t}\Big) - \int \log \frac{K_\varepsilon * \mu_t}{\pi}\; \mu_t(dx) - 1\Bigg).\end{aligned} \tag{27}$$

Kernelizing the energy via Carrillo et al. (2019). Another choice of a kernelized energy function replaces the KL term by $\mathrm{KL}(K_\varepsilon * m \mid \pi)$, leading to the energy

$$V_\varepsilon^\sigma(m) = F(m) \;+\; \sigma\, \mathrm{KL}(K_\varepsilon * m \mid \pi),$$

whose Fisher–Rao gradient flow is

$$\begin{aligned}\partial_t \mu_t = -\,\mu_t \Bigg(&\frac{\delta F}{\delta \mu}(\mu_t, \cdot) + \sigma\, K_\varepsilon * \log \frac{K_\varepsilon * \mu_t}{\pi} \\ &- \sigma\, \mathrm{KL}(K_\varepsilon * \mu_t \mid \pi)\Bigg).\end{aligned} \tag{28}$$

Both (27) and (28) preserve the structure of Fisher–Rao flows, i.e, they are genuine Fisher–Rao gradient flows of the corresponding kernelized energy functions. A notable drawback of (28), however, is the presence of nested convolutions and integrals—e.g., the term $K_\varepsilon * \log(K_\varepsilon * \mu_t/\pi)$—which prevents a full discretization into finite particle sums even when μ_t is an empirical measure, which in practice entails higher computational cost (see the discussion in (Carrillo et al., 2019, Section 1)).

For all kernelizations defined above, we have the following result.

Proposition 3.1. *Suppose Assumptions 2, 3 and 4 hold. Consider the gradient flow $\partial_t \mu_t = -\mu_t a(\mu_t, \cdot)$ defined via* (25), (26), (27) *or* (28)*. Then a satisfies Assumption 1.*

The proof can be found in Appendix E.

3.3 Regularization by χ^2-divergence

Even though the relative entropy is the most popular choice of regularizer in mean-field optimization problems, other choices are possible, and our framework is indeed applicable to energy functions more general than (1). To illustrate this, we discuss briefly energy functions regularized by the χ^2-divergence (with an appropriate kernelization analogous to kernelization (28) for the entropy-regularized energy). Recall that given a fixed reference measure $\pi \in \mathcal{P}(\mathcal{X})$, for any measure $\mu \in \mathcal{P}(\mathcal{X})$ absolutely continuous with respect to π, the χ^2-divergence is defined by $\chi^2(m|\pi) = \int (\frac{dm}{d\pi} - 1)^2 d\pi$. We consider the functional $V_\varepsilon^\sigma(m) := F(m) + \sigma\chi^2(K_\varepsilon * m \,|\, \pi)$. Its Fisher–Rao gradient flow is $\partial_t \mu_t = -\mu_t a(\mu_t, \cdot)$, where

$$\begin{aligned} a(m,x) := \frac{\delta F}{\delta \mu}(m,x) + 2\sigma K_\varepsilon * \left(\frac{K_\varepsilon * m}{\pi}\right)(x) \\ - 2\sigma \int K_\varepsilon * \left(\frac{K_\varepsilon * m}{\pi}\right)(z)\, m(z)\, dz. \end{aligned} \tag{29}$$

We have the following result.

Proposition 3.2. *Suppose Assumptions 2, 3 and 4 hold. Then the function a defined by* (29) *satisfies Assumption 1.*

The proof can be found in Appendix E.5.

3.4 Weak-* Convergence of Minimizers under Kernelization

In the context of mean-field optimization problems (1), an important question related to kernelizations of the corresponding gradient flows is whether those kernelized gradient flows are associated with energy functions whose minimizers are close to the minimizers of the original energy function in (1). In this subsection, we partially answer this question for the kernelized Fisher–Rao gradient flow associated with the kernelization in (27), namely the case where the flow corresponds to the energy function

$$V_\varepsilon^\sigma(m) := F(m) + \sigma \int_{\mathcal{X}} \log \frac{K_\varepsilon * m}{\pi}\, dm(x).$$

We study the limiting behaviour of minimizers as $\varepsilon \to 0$, and prove that any sequence of minimizers of V_ε^σ admits a subsequence converging in the weak-* topology to a minimizer of the original problem (1).

Theorem 3.3 (Weak-* convergence of minimizers)**.** *Suppose Assumptions 2, 3 and 4 hold. For each $\varepsilon > 0$, let $\mu_\varepsilon \in \mathcal{P}_2(\mathcal{X})$ be a minimizer of V_ε^σ. Then there exists a subsequence $(\varepsilon_k)_{k\geq 1}$ such that $\varepsilon_k \to 0$ as $k \to \infty$, and a measure $\mu \in \mathcal{P}_2(\mathcal{X})$ such that*

$$\mu_{\varepsilon_k} \overset{*}{\rightharpoonup} \mu \quad \text{as } k \to \infty,$$

and μ is a minimizer of V^σ given in (1)*.*

We would like to remark that a version of Theorem 3.3 can also be obtained, under some additional assumptions, on $\mathcal{X} = \mathbb{R}^d$. For the sake of completeness, we discuss the details in Appendix C, even though in order to apply the remaining results in this paper to the energy function (23), we require compactness of $\mathcal{X}$.

4 ALGORITHM CONSTRUCTION

The practical usage of the discussed gradient flows depends on developing implementable algorithms for approximating the corresponding interacting particle systems (19). In our algorithm, we draw the position of the particles $\tilde{X}^{i,N}$ at the beginning from the initial distributions $\tilde{\mu}_0^{i,N}$, and afterwards they remain constant. Then we apply a time discretization of the equations for the weights $\tilde{w}^{i,N}$ by an Euler scheme, and hence we obtain a numerical scheme which updates the weights of particles in each step. In the algorithm discussed in this section, the dynamics for the weights follows the equation (25), for a smooth kernel K_ε.

Combining the theoretical results of the present paper with results guaranteeing convergence of the Fisher-Rao gradient flow to the minimizer $m^{\sigma,*}$ of (1) as $t \to \infty$ (see e.g. Liu et al. (2023)), suggests that for a large number of particles, a large number of iterations, and a small ε, the output of our algorithm should provide a reasonable approximation of $m^{\sigma,*}$. However, a full quantitative analysis of the resulting approximation error remains a challenging open problem for future research. Note that the full analysis would need to take into account the following four errors: the error between the continuous flow and the target (due

to running the algorithm for finite time), the error between the "correct" flow and the kernelized flow (due to the use of the kernel K_ε), the error between the continuous particle system and the continuous mean-field flow (due to using a finite number of particles) and the discretisation error for the particle system (due to a positive time-step).

Algorithm: Fisher–Rao gradient descent

Input: particles $X^i \sim \mu_0^i$ with uniform weights $w_0^i = 1$, for $i = 1, \dots, N$
Δt = time interval, J = number of iterations

Steps: update weights

for $j = 1 : J$ **do**

$\tilde{\mu}_{j-1}^N = \frac{1}{N}\sum_{l=1}^N w_{j-1}^l \delta_{X^l}$

for $i = 1 : N$ **do**

$\tilde{V_j}^i := \frac{\delta F}{\delta \mu}(\tilde{\mu}_{j-1}^N, X^i)$
$+\sigma \log\left(\frac{1}{N}\sum_{l=1}^N w_{j-1}^l K_\varepsilon(X^i - X^l)\right)$
$-\sigma \log \pi(X^i) - \sigma \frac{1}{N}\sum_{k=1}^N w_{j-1}^k \times$
$\left(\log\left(\frac{1}{N}\sum_{l=1}^N w_{j-1}^l K_\varepsilon(X^k - X^l)\right) - \log \pi(X^k)\right)$

$\hat{w}_j^i = w_{j-1}^i \exp\left(-\tilde{V}_j^i \, \Delta t\right)$

$w_j^i = N\hat{w}_j^i / \sum_{l=1}^N \hat{w}_j^l$

Output: the weighted empirical distribution of the particles approximates $m^{\sigma,*}$

Note that we need the coefficient N in the update for w_j^i since the weights are supposed to add up to N (cf. (19)).

We remark that in our setting it is also possible to construct an algorithm corresponding to the Wasserstein-Fisher-Rao gradient flow (16). This algorithm would have an additional step within the outer loop, corresponding to the diffusion movement of the particles due to the Wasserstein part of the flow (i.e., this step would correspond to the discretisation of the SDE in (16)). From a practical point of view, such an algorithm is expected to perform better than the algorithm corresponding to the "pure" Fisher-Rao flow, due to the additional exploration of the state space provided by the diffusion movement (the main drawback of the pure Fisher-Rao flow is that it does not expand the support of the initial distribution of the particles - see also Remark 4.1). However, since our theoretical results cover only the case of the pure Fisher-Rao flow (cf. Remark 2.8), we formulate the algorithm without the diffusion part.

Finally, note that due to the Feynman-Kac formula, there is a possible interpretation of the Fisher-Rao flow (4) as a Kolmogorov equation for a stochastic process with killing (see (Applebaum, 2009, Section 6.7.2) or (Karatzas and Shreve, 1991, Theorem 5.7.6 and Exercise 5.7.10)). This leads to an alternative idea for constructing a particle system approximating (4), which instead of rebalancing weights, uses killing and replication mechanism with appropriately defined exponential clocks. Such algorithms were used in Lu et al. (2019); Pampel et al. (2023); Rotskoff et al. (2019), however, we were unable to provide a rigorous proof for propagation of chaos for such particle systems in our framework (1), which is why in this paper we are working with system (19).

Remark 4.1. As the final remark on the practical applicability of the Fisher-Rao algorithm presented in this section, we would like to stress that the main issue with the pure Fisher-Rao flow is that it is very sensitive to initialization, which is reflected in the theoretical results in papers such as Lu et al. (2019, 2023); Liu et al. (2023), which state that, in order for the FR flow to converge (exponentially) to the target μ^*, the initialization μ_0 has to satisfy a "warm start" condition, i.e., there has to exist a constant $C > 0$ such that for any $x \in \mathcal{X}$,

$$\frac{d\mu_0}{d\mu^*}(x) \geq C.$$

In other words, μ_0 and μ^* have to be sufficiently similar to each other and in particular they need to have matching supports. Since in practice it may be difficult to choose initialization in a way that guarantees this "warm start" condition, especially in very high-dimensional settings, this may limit the applicability of the pure FR flow.

However, a natural idea for constructing a practically feasible algorithm that utilizes the pure FR flow, would be to initially run a different flow to provide an initial exploration of the state space and then to switch to the pure FR flow. For instance, one could first run a particle system approximating the pure Wasserstein flow (from an arbitrary initialization), for a certain amount of time $t_0 > 0$ that guarantees that there exists $C > 0$ such that for any $x \in \mathcal{X}$,

$$\frac{d\mu_{t_0}}{d\mu^*}(x) \geq C$$

and then "switch off" the Wasserstein flow and run a particle approximation of the pure FR flow. According to the theory from the papers cited above, a flow like this would converge exponentially to the target (and the corresponding algorithm would be cheaper to run than an algorithm that uses both Wasserstein and FR flows all the time). A fully rigorous analysis of the corresponding algorithm would require propagation of chaos results for both pure Wasserstein flows (which has been covered extensively in the literature, cf. the discussion in the introduction) and the pure FR flows (which we provide in the present paper). A full study of such algorithms (and in particular of the question of how to optimally choose t_0) will be the topic of our future work.

Akcnowledgement

PL acknowledges funding from the Slovenian Research and Innovation Agency (ARIS) under programme No. P1-0448 and Croatian Science Foundation grant no. 2277.

References

Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2 edition.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition.

Aubin-Frankowski, P.-C., Korba, A., and Léger, F. (2022). Mirror descent with relative smoothness in measure spaces, with application to Sinkhorn and EM. In *Advances in Neural Information Processing Systems*, volume 35, pages 17263–17275. Curran Associates, Inc.

Carrillo, J. A., Chen, Y., Zhengyu Huang, D., Huang, J., and Wei, D. (2024). Fisher-Rao Gradient Flow: Geodesic Convexity and Functional Inequalities. *arXiv e-prints*, page arXiv:2407.15693.

Carrillo, J. A., Craig, K., and Patacchini, F. S. (2019). A blob method for diffusion. *Calculus of Variations and Partial Differential Equations*, 58(2):53.

Carrillo, J. A., McCann, R. J., and Villani, C. (2003). Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoam.*, 19(3):971–1018.

Carrillo, J. A., Patacchini, F. S., Sternberg, P., and Wolansky, G. (2016). Convergence of a particle method for diffusive gradient flows in one dimension. *SIAM Journal on Mathematical Analysis*, 48(6):3708–3741.

Cavil, A. L., Oudjane, N., and Russo, F. (2017). Particle system algorithm and chaos propagation related to non-conservative mckean type stochastic differential equations. *Stochastics and Partial Differential Equations: Analysis and Computations*, 5(1):1–37.

Chen, F., Lin, Y., Ren, Z., and Wang, S. (2024). Uniform-in-time propagation of chaos for kinetic mean field langevin dynamics. *Electronic Journal of Probability*, 29(none).

Chen, F., Ren, Z., and Wang, S. (2023). Entropic fictitious play for mean field optimization problem. *J. Mach. Learn. Res.*, 24:Paper No. [211], 36.

Chen, F., Ren, Z., and Wang, S. (2025). Uniform-in-time propagation of chaos for mean field Langevin dynamics. *Ann. Inst. Henri Poincaré Probab. Stat.*, 61(4):2357–2404.

Chizat, L. (2022). Mean-field langevin dynamics: Exponential convergence and annealing. *Transactions on Machine Learning Research*.

Delarue, F. and Tse, A. (2025). Uniform in time weak propagation of chaos on the torus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 61(2):1021–1074.

Domingo-Enrich, C., Jelassi, S., Mensch, A., Rotskoff, G., and Bruna, J. (2020). A mean-field analysis of two-player zero-sum games. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20215–20226. Curran Associates, Inc.

Durmus, A., Eberle, A., Guillin, A., and Zimmer, R. (2020). An elementary approach to uniform in time propagation of chaos. *Proc. Am. Math. Soc.*, 148(12):5387–5398.

Feinberg, E. A., Kasyanov, P. O., and Zadoianchuk, N. V. (2014). Fatou's lemma for weakly converging probabilities. *Theory of Probability & Its Applications*, 58(4):683–689.

Gallouët, T. O. and Monsaingeon, L. (2017). A jko splitting scheme for kantorovich–fisher–rao gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1100–1130.

Gu, A. and Kim, J. (2025). Mirror Mean-Field Langevin Dynamics. *arXiv e-prints*, page arXiv:2505.02621.

Guillin, A. and Monmarché, P. (2021). Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes. *J. Stat. Phys.*, 185(2):20. Id/No 15.

Hu, K., Ren, Z., Šiška, D., and Szpruch, L. (2021). Mean-field langevin dynamics and energy landscape of neural networks. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(4):2043–2065.

Jordan, R., Kinderlehrer, D., and Otto, F. (1998). The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

Kerimkulov, B., Leahy, J., Šiška, D., Szpruch, Ł., and Zhang, Y. (2025). A fisher–rao gradient flow for entropy-regularised markov decision processes in

polish spaces. *Foundations of Computational Mathematics*.

Lacker, D. (2018). Mean field games and interacting particle systems.

Lacker, D. and Le Flem, L. (2023). Sharp uniform-in-time propagation of chaos. *Probability Theory and Related Fields*, pages 1–38.

Lascu, R.-A. and Majka, M. B. (2025). Non-convex entropic mean-field optimization via Best Response flow. *Advances in Neural Information Processing Systems (NeurIPS 2025)*.

Lascu, R.-A., Majka, M. B., and Szpruch, L. (2024). A fisher-rao gradient flow for entropic mean-field min-max games. *Transactions on Machine Learning Research*.

Lascu, R.-A., Majka, M. B., Šiška, D., and Szpruch, Ł. (2024). Linear convergence of proximal descent schemes on the Wasserstein space. *arXiv e-prints*, page arXiv:2411.15067.

Leahy, J.-M., Kerimkulov, B., Šiška, D., and Szpruch, Ł. (2022). Convergence of policy gradient for entropy regularized MDPs with neural network approximation in the mean-field regime. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 12222–12252. PMLR.

Liero, M., Mielke, A., and Savaré, G. (2018). Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117.

Liu, L., Majka, M. B., and Szpruch, L. (2023). Polyak–Łojasiewicz inequality on the space of measures and convergence of mean-field birth-death processes. *Appl Math Optim 87, 48 (2023)*.

Lu, Y., Lu, J., and Nolen, J. (2019). Accelerating langevin sampling with birth-death. arXiv e-prints. arXiv:1905.09863.

Lu, Y., Slepčev, D., and Wang, L. (2023). Birth–death dynamics for sampling: global convergence, approximations and their asymptotics. *Nonlinearity*, 36(11):5731.

Malrieu, F. (2001). Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Processes Appl.*, 95(1):109–132.

Mei, S., Montanari, A., and Nguyen, P.-M. (2018). A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences*, 115(33).

Monmarché, P. (2017). Long-time behaviour and propagation of chaos for mean field kinetic particles. *Stochastic Processes Appl.*, 127(6):1721–1737.

Monmarché, P., Ren, Z., and Wang, S. (2024). Time-uniform log-Sobolev inequalities and applications to propagation of chaos. *Electron. J. Probab.*, 29:Paper No. 154, 38.

Nitanda, A. (2024). Improved Particle Approximation Error for Mean Field Neural Networks. *Advances in Neural Information Processing Systems (NeurIPS 2024)*.

Nitanda, A., Lee, A., Tan Xing Kai, D., Sakaguchi, M., and Suzuki, T. (2025). Propagation of Chaos for Mean-Field Langevin Dynamics and its Application to Model Ensemble. *Forty-second International Conference on Machine Learning (ICML2025)*.

Nitanda, A., Wu, D., and Suzuki, T. (2022). Convex analysis of the mean field langevin dynamics. In Camps-Valls, G., Ruiz, F. J. R., and Valera, I., editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 9741–9757. PMLR.

Pampel, B., Holbach, S., Hartung, L., and Valsson, O. (2023). Sampling rare event energy landscapes via birth-death augmented dynamics. *Physical Review E*, 107(2).

Rotskoff, G., Jelassi, S., Bruna, J., and Vanden-Eijnden, E. (2019). Neuron birth-death dynamics accelerates gradient descent and converges asymptotically. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5508–5517. PMLR.

Salim, A., Korba, A., and Luise, G. (2020). The Wasserstein proximal gradient algorithm. In *Advances in Neural Information Processing Systems*, volume 33, pages 12356–12366. Curran Associates, Inc.

Schuh, K. (2024). Global contractivity for Langevin dynamics with distribution-dependent forces and uniform in time propagation of chaos. *Ann. Inst. Henri Poincaré Probab. Stat.*, 60(2):753–789.

Suzuki, T., Nitanda, A., and Wu, D. (2023). Uniform-in-time propagation of chaos for the mean field gradient Langevin dynamics. *The Eleventh International Conference on Learning Representations (ICLR2023)*.

Zhu, J.-J. and Mielke, A. (2024). Kernel Approximation of Fisher-Rao Gradient Flows. *arXiv e-prints*, page arXiv:2410.20622.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Table of Contents

1 INTRODUCTION **1**

1.1 Contributions 2

2 MAIN RESULTS **2**

2.1 Preliminaries 3

2.2 Existence of mean-field dynamics corresponding to Fisher-Rao flow 3

2.3 Interacting Particle Systems and Propagation of Chaos 5

2.4 Discussion of other propagation of chaos results in the literature 5

3 ENTROPIC MEAN-FIELD PROBLEMS **6**

3.1 Setting for entropic mean-field optimization 6

3.2 Kernelization Strategies 7

3.3 Regularization by χ^2-divergence 8

3.4 Weak-* Convergence of Minimizers under Kernelization 8

4 ALGORITHM CONSTRUCTION **8**

A WELL-POSEDNESS OF FISHER–RAO FLOWS **14**

B PROPAGATION OF CHAOS FOR THE INTERACTING PARTICLE SYSTEM **17**

C WEAK-* CONVERGENCE OF KERNELIZED MINIMIZERS **20**

C.1 Existence of minimizers - Compact case 20

C.2 Existence of minimizers - Non Compact case 21

C.3 Convergence of minimizers (for both cases) 24

D SUPPLEMENTARY DEFINITIONS AND TECHNICAL RESULTS **25**

D.1 Flat Derivative 25

D.2 Technical Estimates 25

D.3 Construction of the Kernel on a compact space 26

E VERIFICATION OF THE BOUNDNESS AND LIPSCHITZ CONDITION **28**

E.1 For Flow (25) 31

E.2 For Flow (26) 32

E.3 For Flow (27) 34

E.4 For Flow (28) 36

E.5 Boundness and Lipschitz Condition for Kernelized Fisher-Rao Gradient Flow Regularized by Chi Square Divergence 37

Appendix: On Propagation of Chaos for the Fisher-Rao Gradient Flow in Entropic Mean Field Optimization

A WELL-POSEDNESS OF FISHER–RAO FLOWS

Proof of Lemma 2.5. We aim to show that if ν_t solves (9), then the projected density $\mu_t(x) = \int_0^\infty w\nu_t(x,w)\,dw$ satisfies (11) in the weak sense, namely we want to show for all test function $\varphi \in \mathcal{C}_c^\infty(\mathcal{X})$

$$\begin{aligned} &\partial_t \int_{\mathcal{X}} \varphi(x)\mu_t(x)dx = -\int_{\mathcal{X}} \varphi(x)\cdot a(\mu_t,x)\cdot \mu_t(x)dx \\ &\mu_t(x) := \int_0^\infty w\nu_t(x,w)dw. \end{aligned}.$$

Since we assumed that ν_t solves (9) weakly, we know that for all $\psi \in \mathcal{C}_c^\infty(\mathcal{X}\times\mathbb{R}_+)$ we have

$$\partial_t \int_{\mathcal{X}\times\mathbb{R}_+} \psi(x,w)\nu_t(x,w)dxdw = -\int_{\mathcal{X}\times\mathbb{R}_+} \frac{\partial\psi}{\partial w}(x,w)\cdot\nu_t(x,w)\cdot w\cdot a(\mu_t,x)\,dxdw.$$

Thanks to Remark (2.4), we choose ψ in the form of $\psi(x,w) = w\varphi(x)$ for a $\varphi \in \mathcal{C}_c^\infty(\mathcal{X})$. Hence the above equation becomes

$$\partial_t \int_{\mathcal{X}\times\mathbb{R}_+} w\varphi(x)\nu_t(x,w)dxdw = -\int_{\mathcal{X}\times\mathbb{R}_+} \varphi(x)\cdot\nu_t(x,w)\cdot w\cdot a(\mu_t,x)\,dxdw.$$

Note that the LHS could be rewrite as:

$$LHS = \partial_t \int_{\mathcal{X}} \varphi(x)\int_0^\infty w\nu_t(x,w)dwdx = \partial_t \int_{\mathcal{X}} \varphi(x)\mu_t(x)dx$$

Moreover, the RHS could be rewrite as:

$$RHS = -\int_{\mathcal{X}} \varphi(x)\cdot a(\mu_t,x)\left(\int_0^\infty w\nu_t(x,w)dw\right)dx = -\int_{\mathcal{X}} \varphi(x)\cdot a(\mu_t,x)\cdot\mu_t(x)dx.$$

Putting it together, we have shown that

$$\partial_t \int_{\mathcal{X}} \varphi(x)\mu_t(x)dx = -\int_{\mathcal{X}} \varphi(x)\cdot a(\mu_t,x)\cdot\mu_t(x)dx,$$

This completes the proof. □

Proof of Theorem 2.6. Let us take a test function $f \in \mathcal{C}_c^\infty(\mathcal{X}\times\mathbb{R}_+)$. By chain rule, we have

$$\begin{aligned} f(X,w_t) &= f(X,w_0) + \int_0^t \frac{\partial f}{\partial w}(X,w_s)dw_s \\ &= f(X,w_0) - \int_0^t \frac{\partial f}{\partial w}(X,w_s)w_s a(\mathtt{h}\,\mathrm{Law}(X,w_s),X)ds \end{aligned}$$

Now, if we take expectations on both sides, we get

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathbb{R}_+} f(x,w)\,\mathrm{Law}(X,w_t)(x,w)dxdw \\
&= \int_{\mathcal{X}\times\mathbb{R}_+} f(x,w)\,\mathrm{Law}(X,w_0)(x,w)dxdw \\
&- \int_0^t \int_{\mathcal{X}\times\mathbb{R}_+} \frac{\partial f}{\partial w}(x,w) w a(x, \mathtt{h}\,\mathrm{Law}(X,w_s))\,\mathrm{Law}(X,w_s)(x,w)dxdwds
\end{aligned}$$

For the second trem, by using integration by part, we get the following equation

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathbb{R}_+} f(x,w)\,\mathrm{Law}(X,w_t)(x,w)dxdw \\
&= \int_{\mathcal{X}\times\mathbb{R}_+} f(x,w)\,\mathrm{Law}(X,w_0)(x,w)dxdw \\
&+ \int_0^t \int_{\mathcal{X}\times\mathbb{R}_+} f(x,w)\frac{\partial}{\partial w}\left(\mathrm{Law}(X,w_s)(x,w) w a(\mathtt{h}\,\mathrm{Law}(X,w_s),x)\right) dxdwds
\end{aligned}$$

where the last equality shows $\mathrm{Law}(X,w_s)$ is the weak solution in the sense

$$\begin{cases} \partial_t \nu_t = \frac{\partial}{\partial w}(\nu_t(x,w) w a(\mathtt{h}\nu_t, x)) \\ \nu_t \rightharpoonup \nu_0 \end{cases} \tag{30}$$

□

Proof of Theorem 2.9. We abbreviate $\mathcal{P}_2(\mathcal{C}([0,T];\mathcal{X}\times\mathbb{R}_+))$ by $\mathcal{P}_2^T$ and define the mapping

$$\begin{aligned}
\Psi : \mathcal{P}_2^T &\to \mathcal{P}_2^T, \\
\nu &\mapsto \mathrm{Law}(X^\nu, w^\nu),
\end{aligned}$$

where, for a fixed $\nu \in \mathcal{P}_2^T$,

$$\begin{aligned}
&(X^\nu, w_0^\nu) \sim \nu_0, \\
&dw_t^\nu = -w_t^\nu\, a\big(\mathtt{h}\nu_t, X^\nu\big)\, dt, \quad t \le T.
\end{aligned} \tag{31}$$

Here ν is treated as an external input, independent of the process itself, so (31) is not a McKean–Vlasov equation but an ODE in (X^ν, w_t^ν). By the classical existence and uniqueness theory for ODEs with Lipschitz drift, the system admits a unique strong solution whenever the coefficients are globally Lipschitz in (x,w). In our setting, it suffices to check that

$$(x,w) \mapsto w\, a(\mathtt{h}\nu_t, x)$$

is globally Lipschitz. Indeed, for any $(x,w),(x',w') \in \mathcal{X}\times\mathbb{R}_+$,

$$\begin{aligned}
|w\, a(\mathtt{h}\,\nu_t, x) - w'\, a(\mathtt{h}\,\nu_t, x')| &\le |w-w'|\,|a(\mathtt{h}\,\nu_t, x)| + |w'|\,|a(\mathtt{h}\,\nu_t, x) - a(\mathtt{h}\,\nu_t, x')| \\
&\le M_a\,|w-w'| + e^{M_a T} L_a\,|x-x'|,
\end{aligned}$$

where M_a and L_a denote the uniform bound and Lipschitz constant of a, respectively. Moreover, the boundedness of a implies, via Proposition D.2, that w_t^ν is uniformly bounded in t, which ensures that the drift in (31) is indeed globally Lipschitz.

Furthermore,

$$\mathbb{E}\left[\|(X^\nu, w^\nu)\|_T^2\right] < \infty,$$

so for all $\nu \in \mathcal{P}_2^T$, the mapping $\Psi(\nu)$ remains in $\mathcal{P}_2^T$.

Now let $\nu, \nu' \in \mathcal{P}_2^T$ be such that $\nu_0 = \nu_0'$. We associate to them the processes (X^ν, w^ν) and $(X^{\nu'}, w^{\nu'})$, which share identical initial conditions almost surely:

$$(X_0^\nu, w_0^\nu) = (X_0^{\nu'}, w_0^{\nu'}) \quad \text{a.s.}$$

Let $\Psi(\nu)$ and $\Psi(\nu')$ denote the probability measures generated by (31). We then have:

$$
\begin{aligned}
&\mathbb{E}\left[\|w^\nu - w^{\nu'}\|_T^2\right] \le T\mathbb{E}\left[\int_0^T \left|w_s^\nu a(\mathtt{h}\nu_s, X^\nu) - w_s^{\nu'} a(\mathtt{h}\nu'_s, X^{\nu'})\right|^2 ds\right] \\
&\le 2T\mathbb{E}\left[\int_0^T \left|w_s^\nu a(\mathtt{h}\nu_s, X^\nu) - w_s^{\nu} a(\mathtt{h}\nu'_s, X^{\nu'})\right|^2 ds\right] \\
&+ 2T\mathbb{E}\left[\int_0^T \left|w_s^\nu a(\mathtt{h}\nu'_s, X^{\nu'}) - w_s^{\nu'} a(\mathtt{h}\nu'_s, X^{\nu'})\right|^2 ds\right] \\
&\le 2TL_a^2 e^{2M_aT}\mathbb{E}\left[\int_0^T \mathcal{W}_2^2\left(\mathtt{h}\nu_s, \mathtt{h}\nu'_s\right) + \left|X^\nu - X^{\nu'}\right|^2 ds\right] + 2TM_a^2\mathbb{E}\left[\int_0^T \left|w_s^\nu - w_s^{\nu'}\right|^2 ds\right] \\
&\le 2TL_a^2 e^{2M_aT}\mathbb{E}\left[\int_0^T \mathcal{W}_2^2\left(\mathtt{h}\nu_s, \mathtt{h}\nu'_s\right) + \left|X^\nu - X^{\nu'}\right|^2 ds\right] + 2TM_a^2\mathbb{E}\left[\int_0^T \left\|w^\nu - w^{\nu'}\right\|_s^2 ds\right]
\end{aligned}
$$

Here the first inequality follows from the Cauchy–Schwarz inequality, the second from the elementary bound $(a+b)^2 \le 2a^2 + 2b^2$, and the third from the Lipschitz continuity of a together with the boundedness of w^ν (Proposition D.2) and the uniform bound $|a| \le M_a$. Finally, the last inequality uses the fact that $|f(s)| \le \|f\|_s$ for any $s \le T$.

Note that we have

$$
\mathbb{E}\left[\left|X_0^\nu - X_0^{\nu'}\right|^2\right] = 0.
$$

Combining them together, we have

$$
\begin{aligned}
&\mathbb{E}\left[\left\|X^\nu - X^{\nu'}\right\|_T^2 + \|w^\nu - w^{\nu'}\|_T^2\right] = \mathbb{E}\left[\|w^\nu - w^{\nu'}\|_T^2\right] \\
&\le 2TL_a^2 e^{2M_aT}\int_0^T \mathcal{W}_2^2\left(\mathtt{h}\nu_s, \mathtt{h}\nu'_s\right) ds + 2TM_a^2\mathbb{E}\left[\int_0^T \left\|w^\nu - w^{\nu'}\right\|_s^2 ds\right].
\end{aligned}
$$

By applying Grönwall's lemma together with Proposition D.2, and then using $\mathcal{W}_2(\nu_s, \nu'_s) \le \mathcal{W}_{2,s}(\nu, \nu')$ for all $s \in [0, T]$, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\|w^\nu - w^{\nu'}\|_T^2\right] &\le 2TL_a^2 e^{4M_aT + 2T^2M_a^2}\int_0^T \mathcal{W}_2^2(\nu_s, \nu'_s) ds \\
&\le 2TL_a^2 e^{4M_aT + 2T^2M_a^2}\int_0^T \mathcal{W}_{2,s}^2(\nu, \nu') ds.
\end{aligned}
$$

We denote $C := 2TL_a^2 e^{4M_aT + 2T^2M_a^2}$. We have

$$
\mathbb{E}\left[\|X^\nu - X^{\nu'}\|_T^2 + \|w^\nu - w^{\nu'}\|_T^2\right] \le C\int_0^t \mathcal{W}_{2,s}^2(\nu, \nu') ds. \tag{32}
$$

Moreover, since the joint laws of $((X^\nu, w^\nu), (X^{\nu'}, w^{\nu'}))$ is one of the coupling between $\Psi(\nu)$ and $\Psi(\nu')$, that is, it belongs to the set $\Pi(\Psi(\nu), \Psi(\nu'))$. We have

$$
\begin{aligned}
\mathcal{W}_{2,T}^2\left(\Psi(\nu), \Psi(\nu')\right) &= \inf_{\pi \in \Pi((\Psi(\nu), \Psi(\nu')))} \mathbb{E}_{((X^\nu, w^\nu), (X^{\nu'}, w^{\nu'})) \sim \pi}\left[\|X^\nu - X^{\nu'}\|_T^2 + \|w^\nu - w^{\nu'}\|_T^2\right] \\
&\le \mathbb{E}\left[\|X^\nu - X^{\nu'}\|_T^2 + \|w^\nu - w^{\nu'}\|_T^2\right] \\
&\le C\int_0^T \mathcal{W}_{2,s}^2(\nu, \nu') ds
\end{aligned}
$$

We conclude that

$$\mathcal{W}_{2,T}^2\left(\Psi(\nu),\Psi(\nu')\right) \leq C\int_0^T \mathcal{W}_{2,s}^2(\nu,\nu')ds. \tag{33}$$

Hence for all $\nu \in \mathcal{P}_{2,T}$ we have

$$\begin{aligned}
\mathcal{W}_{2,T}^2\left(\Psi^{k+1}(\nu),\Psi^k(\nu)\right) &\leq C\int_0^T \mathcal{W}_{2,s}^2\left(\Psi^k(\nu),\Psi^{k-1}(\nu')\right)ds \\
&\leq C^k\int_0^T\int_0^s\int_0^{s_1}\cdots\int_0^{s_{k-2}} \mathcal{W}_{2,s_{k-1}}^2(\Psi(\nu),\nu')ds_{k-1}\cdots ds_1 ds \\
&\leq \frac{C^kT^k}{k!}\mathcal{W}_{2,T}^2(\Psi(\nu),\nu).
\end{aligned}$$

As $(\mathcal{P}_{2,T},\mathcal{W}_{2,T})$ is complete and $\Psi^k(\nu)$ is a Cauchy sequence, the sequence $\Psi^k(\nu)$ converges to the unique fixed point of Ψ. □

B PROPAGATION OF CHAOS FOR THE INTERACTING PARTICLE SYSTEM

Remark B.1. For the particle system under consideration, the total weight

$$\sum_{i=1}^N w_t^{i,N}$$

is conserved in time. Indeed, using the property of a that

$$\int a(m,x)\,m(dx) = 0 \quad \text{for any probability measure } m,$$

we compute

$$\begin{aligned}
\partial_t\sum_{i=1}^N w_t^{i,N} &= \sum_{i=1}^N w_t^{i,N}a(\mathtt{h}\nu_t^N, X_t^{i,N}) \\
&= \int a(\mathtt{h}\nu_t^N,x)\sum_{i=1}^N w_t^{i,N}\delta_{X_t^{i,N}}(x)\,dx \\
&= \left(\sum_{i=1}^N w_t^{i,N}\right)\int a(\mathtt{h}\nu_t^N,x)\,\mathtt{h}\nu_t^N(dx) = 0,
\end{aligned}$$

which implies $\sum_{i=1}^N w_t^{i,N} = \sum_{i=1}^N w_0^{i,N}$ for all $t \geq 0$. The same property holds for the tilde system $\tilde{\nu}_t^N$.

In this work, we choose $\sum_{i=1}^N w_0^{i,N} = N$. This convention ensures that, with the empirical measure

$$\nu_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{(w_t^{i,N},X_t^{i,N})},$$

its projection

$$\mathtt{h}\nu_t^N = \frac{1}{N}\sum_{i=1}^N w_t^{i,N}\,\delta_{X_t^{i,N}}$$

is a probability measure. The same reasoning applies to $\tilde{\nu}_t^N$ and $\mathtt{h}\tilde{\nu}_t^N$.

We now state an important remark on the convergence result.

Remark B.2. Let $\mathcal{X}$ be a separable metric space and $(X_i)_{i\geq 1}$ be i.i.d. $\mathcal{X}$-valued random variables with law μ. Let

$$\mu^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}$$

be the empirical measure. If $p \geq 1$ and $\mu \in \mathcal{P}^p(\mathcal{X})$, then

$$\mathcal{W}_p(\mu^N, \mu) \to 0 \quad \text{almost surely as } N \to \infty,$$

and moreover,

$$\mathbb{E}\big[\mathcal{W}_p^p(\mu^N, \mu)\big] \to 0.$$

A proof of this result can be found in (Lacker, 2018, Corollary 2.14).

Proof of Theorem 2.10. Let us rewrite the particle systems (17) and (19) in integral form. For the independent system we have

$$\begin{cases} X_t^i = X_0^i, \\ w_t^i = w_0^i - \int_0^t w_s^i \, a\big(\mathtt{h}\nu_s, X^i\big)\, ds, \end{cases}$$

and for the interacting system

$$\begin{cases} \tilde{X}_t^{i,N} = \tilde{X}_0^{i,N}, \\ \tilde{w}_t^{i,N} = \tilde{w}_0^{i,N} - \int_0^t \tilde{w}_s^{i,N} \, a\big(\mathtt{h}\tilde{\nu}_s^N, \tilde{X}^{i,N}\big)\, ds. \end{cases}$$

We have

$$\begin{aligned}
&\mathbb{E}\left[\|w^i - \tilde{w}^{i,N}\|_T^2\right] \leq \mathbb{E}\left[\left|w_0^i - \tilde{w}_0^{i,N}\right|\right] + T\mathbb{E}\left[\int_0^T \left|w_s^i a(\mathtt{h}\nu_s^i, X^i) - \tilde{w}_s^{i,N} a(\mathtt{h}\tilde{\nu}_s^N, \tilde{X}^{i,N})\right|^2 ds\right] \\
&\leq \mathbb{E}\left[\left|w_0^i - \tilde{w}_0^{i,N}\right|\right] \\
&+ 2T\mathbb{E}\left[\int_0^T \left|w_s^i a(\mathtt{h}\nu_s^i, X^i) - w_s^i a(\mathtt{h}\tilde{\nu}_s^N, \tilde{X}^{i,N})\right|^2 + \left|w_s^i a(\mathtt{h}\tilde{\nu}_s^N, \tilde{X}^{i,N}) - \tilde{w}_s^{i,N} a(\tilde{\mu}_s^N, \tilde{X}^{i,N})\right|^2 ds\right] \\
&\leq \mathbb{E}\left[\left|w_0^i - \tilde{w}_0^{i,N}\right|\right] \\
&+ 2TL_a^2 e^{2M_a T}\mathbb{E}\left[\int_0^T \mathcal{W}_2^2\left(\mathtt{h}\nu_s^i, \mathtt{h}\tilde{\nu}_s^N\right) + \left|X^i - \tilde{X}^{i,N}\right|^2 ds\right] + 2TM_a^2\mathbb{E}\left[\int_0^T \left|w_s^i - \tilde{w}_s^{i,N}\right|^2 ds\right] \\
&\leq \mathbb{E}\left[\left|w_0^i - \tilde{w}_0^{i,N}\right|\right] + 2TL_a^2 e^{2M_a T}\mathbb{E}\left[\int_0^T \mathcal{W}_2^2\left(\mathtt{h}\nu_s, \mathtt{h}\tilde{\nu}_s^N\right) ds\right] \\
&+ (2TM_a^2 + 2TL_a^2 e^{2M_a T})\mathbb{E}\left[\int_0^T \left\|X^i - \tilde{X}^{i,N}\right\|_s^2 + \left\|w^i - \tilde{w}^{i,N}\right\|_s^2 ds\right]
\end{aligned}$$

where the first inequality follows from integrating the square of the drift term over $[0, T]$ and applying Cauchy–Schwarz inequality. The second inequality uses the standard bound $(a+b)^2 \leq 2a^2 + 2b^2$. The third inequality follows from the boundedness of a by M_a, its Lipschitz continuity with constant L_a, and the uniform bound $|w_s^i| \leq e^{M_a T}$ from Proposition D.2. Finally, the last inequality is obtained by combining like terms and applying the bound $|w_s| \leq \|w\|_s$.

Moreover, since X^i and $\tilde{X}^{i,N}$ are constant in time, their sup–norm difference reduces to the pointwise difference, i.e.,

$$\mathbb{E}\left[\left\|X^i - \tilde{X}^{i,N}\right\|_T^2\right] = \mathbb{E}\left[\left|X^i - \tilde{X}^{i,N}\right|^2\right].$$

Combining them together, we have

$$\begin{aligned}
&\mathbb{E}\left[\left\|X^i-\tilde{X}^{i,N}\right\|_T^2+\|w^i-\tilde{w}^{i,N}\|_T^2\right]\\
&\leq 2TL_a^2e^{2M_aT}\mathbb{E}\left[\int_0^T\mathcal{W}_2^2\left(\mathtt{h}\nu_s,\mathtt{h}\tilde{\nu}_s^N\right)ds\right]+\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]\\
&+\left(2TM_a^2+2TL_a^2e^{2M_aT}\right)\mathbb{E}\left[\int_0^T\left\|X^i-\tilde{X}^{i,N}\right\|_s^2+\left\|w^i-\tilde{w}^{i,N}\right\|_s^2ds\right]
\end{aligned}$$

Let us denote $C_1=2TL_a^2e^{2M_aT}$ and $C_2=2TM_a^2+2TL_a^2e^{2M_aT}$. Remark that both C_1 and C_2 are independent with N the number of particles. By applying Grönwall's lemma and using the stability estimates from Propositions D.2 and D.3, we obtain:

$$\begin{aligned}
&\mathbb{E}\left[\left\|X^i-\tilde{X}^{i,N}\right\|_T^2+\|w^i-\tilde{w}^{i,N}\|_T^2\right]\\
&\leq e^{C_2T}\left(C_1\mathbb{E}\left[\int_0^T\mathcal{W}_2^2(\mathtt{h}\nu_s,\mathtt{h}\tilde{\nu}_s^N)ds\right]+\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]\right)\\
&\leq e^{C_2T}\left(C_1e^{2M_aT}\mathbb{E}\left[\int_0^T\mathcal{W}_2^2(\nu_s,\tilde{\nu}_s^N)ds\right]+\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]\right)\\
&\leq e^{C_2T}\left(C_1e^{2M_aT}\mathbb{E}\left[\int_0^T\mathcal{W}_{2,s}^2(\nu,\tilde{\nu}^N)ds\right]+\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]\right).
\end{aligned}\tag{34}$$

We also have

$$\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu^N,\tilde{\nu}^N\right)\right]\leq\frac{1}{N}\mathbb{E}\left[\sum_{i=1}^N\left\|X^i-\tilde{X}^{i,N}\right\|_T^2+\left\|w^i-\tilde{w}^{i,N}\right\|_T^2\right],\tag{35}$$

which follows by choosing the specific coupling $\frac{1}{N}\sum_{i=1}^N\delta_{\left((w^i,X^i),(\tilde{w}^{i,N},\tilde{X}^{i,N})\right)}$. By using the triangle inequality, (35) and (34) in order, we obtain:

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\tilde{\nu}^N\right)\right]\leq 2\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\nu^N\right)\right]+2\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu^N,\tilde{\nu}^N\right)\right]\\
&\leq 2\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\nu^N\right)\right]+\frac{2}{N}\mathbb{E}\left[\sum_{i=1}^N\left\|X^i-\tilde{X}^{i,N}\right\|_T^2+\left\|w^i-\tilde{w}^{i,N}\right\|_T^2\right]\\
&\leq 2\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\nu^N\right)\right]\\
&+2e^{C_2T}C_1e^{2M_aT}\mathbb{E}\left[\int_0^T\mathcal{W}_{2,s}^2(\nu,\tilde{\nu}^N)ds\right]+\frac{2e^{C_2T}}{N}\sum_{i=1}^N\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]
\end{aligned}$$

By applying Gronwall's lemma, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\tilde{\nu}^N\right)\right]\\
&\leq 2e^{2e^{C_2T}C_1e^{2M_aT}T}\left(\mathbb{E}\left[\mathcal{W}_{2,T}^2\left(\nu,\nu^N\right)\right]+\frac{e^{C_2T}}{N}\sum_{i=1}^N\mathbb{E}\left[\left|X^i-\tilde{X}^{i,N}\right|^2+\left|w_0^i-\tilde{w}_0^{i,N}\right|^2\right]\right)
\end{aligned}$$

which yields the result thanks to Remark B.2 and our assumption on the initial condition 21. □

Proof of Corollary 2.11. By Propositions D.2 and D.3, we have for all $t\in[0,T]$,

$$\mathcal{W}_2^2\left(\mu_t,\tilde{\mu}_t^N\right)\leq e^{2M_a}\mathcal{W}_2^2\left(\nu_t,\tilde{\nu}_t^N\right)\leq e^{2M_a}\mathcal{W}_{2,T}^2\left(\nu,\tilde{\nu}^N\right).$$

Since the right-hand side does not depend on t, taking the supremum over $t \in [0,T]$ yields

$$\sup_{t\in[0,T]} \mathcal{W}_2^2\left(\mu_t, \tilde{\mu}_t^N\right) \leq e^{2M_a} \mathcal{W}_{2,T}^2\left(\nu, \tilde{\nu}^N\right).$$

Applying Theorem 2.10 then gives the desired result. □

C WEAK-* CONVERGENCE OF KERNELIZED MINIMIZERS

In this section, we investigate the variational properties of the kernelized energy

$$V_\varepsilon^\sigma(m) = F(m) + \sigma \int \log\left(\frac{K_\varepsilon * m}{\pi}\right) dm,$$

with particular emphasis on the existence of minimizers and their behaviour as the regularization parameter $\varepsilon \to 0$.

We will prove Theorem 3.3 under Assumptions 2, 3 and 4 for compact $\mathcal{X} \subset \mathbb{R}^d$, and also formulate an additional result for $\mathcal{X} = \mathbb{R}^d$.

More precisely, for $\mathcal{X} = \mathbb{R}^d$, we need the following growth condition on the potential U (recall that $\pi \propto e^{-U}$).

Assumption 5. We assume that there exist constants $C_0, A_0 \in \mathbb{R}$ and $C_1, A_1 > 0$ such that

$$C_0 + C_1|x|^2 \leq U(x) \leq A_0 + A_1|x|^2, \qquad \forall x \in \mathbb{R}^d.$$

Then we have the following additional result.

Theorem C.1 (Weak-* convergence of minimizers - Non Compact case)**.** *Suppose $\mathcal{X} = \mathbb{R}^d$, let Assumptions 2, 3, 4 and 5 hold, and suppose the kernel ξ is Gaussian. Then there exists a minimizer of V_ε^σ in $\mathcal{P}_2(\mathbb{R}^d)$ for all $\varepsilon > 0$. Moreover, for all $\varepsilon > 0$, if $\mu_\varepsilon \in \mathcal{P}_2(\mathbb{R}^d)$ is a minimizer of V_ε^σ then there exists a subsequence $(\varepsilon_k)_{k\geq 1}$ such that $\varepsilon_k \to 0$ as $k \to \infty$ such that*

$$\mu_{\varepsilon_k} \overset{*}{\rightharpoonup} \mu \quad \text{as } k \to \infty,$$

where μ is a minimizer of V^σ.

We will split the proof into the compact case (Theorem 3.3) and the non-compact case (Theorem C.1). We will follow Carrillo et al. (2019).

C.1 Existence of minimizers - Compact case

Proposition C.2. *Suppose Assumptions 2, 3 and 4 hold. Then for all $\varepsilon > 0$, the energy functional*

$$V_\varepsilon^\sigma(m) := F(m) + \sigma \int_{\mathcal{X}} \log\left(\frac{K_\varepsilon * m}{\pi}\right) dm$$

is lower semi-continuous with respect to weak convergence in $\mathcal{P}(\mathcal{X})$.

Proof. Let $(m_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ be a sequence of probability measures converging weakly to $m \in \mathcal{P}(\mathcal{X})$. We aim to show

$$\liminf_{n\to\infty} \left(F(m_n) + \sigma \int \log\left(\frac{K_\varepsilon * m_n}{\pi}\right) dm_n\right) \geq F(m) + \sigma \int \log\left(\frac{K_\varepsilon * m}{\pi}\right) dm.$$

Since F is lower semi-continuous and π is continuous and bounded from below by assumption we know that $m \mapsto F(m) - \sigma \int \log \pi dm$ is lower semi-continuous. Hence it suffices to establish

$$\liminf_{n\to\infty} \int \log(K_\varepsilon * m_n)\, dm_n \geq \int \log(K_\varepsilon * m)\, dm.$$

Note that for fixed $\varepsilon > 0$, the function $x \mapsto \log(K_\varepsilon * m_n(x))$ is continuous and bounded from below by $\log K_{\min,\varepsilon}$, where $K_{\min,\varepsilon} := \inf_{x\in\mathcal{X}} K_\varepsilon(x)$. Therefore, we can apply a version of the generalized Fatou's lemma (see, e.g., Feinberg et al. (2014)), which yields:

$$\liminf_{n\to\infty} \int \log\left(\frac{K_\varepsilon * m_n(x)}{K_{\min,\varepsilon}}\right) dm_n(x) \geq \int \liminf_{n\to\infty, x'\to x} \log\left(\frac{K_\varepsilon * m_n(x')}{K_{\min,\varepsilon}}\right) dm(x). \tag{36}$$

Moreover, for each $x \in \mathcal{X}$, the convolution $K_\varepsilon * m_n(x')$ converges pointwise to $K_\varepsilon * m(x)$ as $n \to \infty$ and $x' \to x$, due to weak convergence and the continuity of the kernel. Therefore,

$$\liminf_{n\to\infty, x'\to x} \log\left(\frac{K_\varepsilon * m_n(x')}{K_{\min,\varepsilon}}\right) = \log\left(\frac{K_\varepsilon * m(x)}{K_{\min,\varepsilon}}\right). \tag{37}$$

Combining (36) and (37), we conclude that

$$\liminf_{n\to\infty} \int \log(K_\varepsilon * m_n)\, dm_n \geq \int \log(K_\varepsilon * m)\, dm,$$

which proves the lower semicontinuity of V_ε^σ. □

Proof of Theorem 3.3 (Step 1: Existence of minimizers). For a compact domain $\mathcal{X}$, the argument is straightforward. The energy V_ε^σ is bounded from below. Moreover, any sequence of probability measures (in particular, any minimizing sequence) is automatically tight as a direct consequence of compactness. Indeed, for any $p \geq 1$,

$$\int_{\mathcal{X}} |x|^p\, dm(x) \leq \sup_{x\in\mathcal{X}} |x|^p.$$

Together with the lower semicontinuity established in Proposition C.3, this implies the existence of a weakly-$*$ convergent subsequence whose limit is a minimizer of V_ε^σ. □

C.2 Existence of minimizers - Non Compact case

Proposition C.3. *Suppose Assumptions 2, 3, 4 and 5 hold and the kernel ξ is Gaussian. Then for all $\varepsilon > 0$, the energy functional*

$$V_\varepsilon^\sigma(m) := F(m) + \sigma \int_{\mathbb{R}^d} \log\left(\frac{K_\varepsilon * m}{\pi}\right) dm$$

is lower semi-continuous with respect to weak convergence in $\mathcal{P}_2(\mathbb{R}^d)$.

Proof. Let $(m_n)_{n\in\mathbb{N}} \subset \mathcal{P}_2(\mathbb{R}^d)$ be a sequence of probability measures converging weakly to some $m \in \mathcal{P}_2(\mathbb{R}^d)$. We aim to show

$$\liminf_{n\to\infty} \left(F(m_n) + \sigma \int \log\left(\frac{K_\varepsilon * m_n}{\pi}\right) dm_n\right) \geq F(m) + \sigma \int \log\left(\frac{K_\varepsilon * m}{\pi}\right) dm.$$

Since F is lower semi-continuous and $\log \pi$ is integrable with respect to any $m \in \mathcal{P}_2(\mathbb{R}^d)$ due to the quadratic growth of U, it suffices to establish

$$\liminf_{n\to\infty} \int \log\left(K_\varepsilon * m_n\right) dm_n \geq \int \log\left(K_\varepsilon * m\right) dm.$$

Then we want to apply the generalized Fatou's Lemma (which holds for non negative functions). However, unlike in the compact case, when we work on $\mathbb{R}^d$ we do not have a lower bound for the kernel, hence we need a different approach in order to get a non negative function. From Carrillo et al. (2019)[Proposition 3.9] equation (54) we know that if K is Gaussian, then there exists $x_0 \in \mathbb{R}^d$ and $C_0, C_1 \in \mathbb{R}$ such that for n sufficient large we have

$$\log(K_\varepsilon * m_n)(x) \geq C_0|x - x_0|^2 + C_1.$$

If we denote the LHS by $f_n(x)$ and RHS by $q(x)$, we can use them to apply generalized Fatou and we have

$$\liminf_{n\to\infty} \int (f_n(x) - q(x)) dm_n(x) \geq \int \liminf_{n\to\infty, x'\to x} (f_n(x') - q(x'))\, dm(x). \tag{38}$$

Since

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \big(-q(x)\big)\, dm_n(x) = \int_{\mathbb{R}^d} \big(-q(x)\big)\, dm(x) = \int_{\mathbb{R}^d} \liminf_{n\to\infty x'\to x} \big(-q(x')\big)\, dm(x),$$

we can cancel the q terms on both sides. This yields the same result as following.

We conclude that

$$\liminf_{n\to\infty} \int \log(K_\varepsilon * m_n)\, dm_n \geq \int \log(K_\varepsilon * m)\, dm,$$

which proves the lower semicontinuity of V_ε^σ. □

Lemma C.4. *(Carrillo et al., 2016, Lemma 4.1) Suppose ρ is a probability density on $\mathbb{R}^d$ with finite second moment $M_2(\rho) := \int_{\mathbb{R}^d} |x|^2 \rho(x) dx$. Then for all $\delta > 0$, we have*

$$\int_{\mathbb{R}^d} \log \rho(x)\rho(x) dx \geq -\left(\frac{2\pi}{\delta}\right)^{d/2} - \delta M_2(\rho),$$

Proof. We split the integration domain:

$$\begin{aligned}
\int_{\mathbb{R}^d} \log \rho(x)\rho(x) dx &= \int_{\{\log \rho(x)\leq 0\}} \log \rho(x)\rho(x) dx + \int_{\{\log \rho(x)>0\}} \log \rho(x)\rho(x) dx \\
&\geq \int_{\{\log \rho(x)\leq 0\}} \log \rho(x)\rho(x) dx \\
&= -\int_{\{\rho(x)\leq 1\}} |\log \rho(x)|\, \rho(x) dx \\
&= -\int_{\{\rho(x)\leq e^{-\delta|x|^2}\}} |\log \rho(x)|\, \rho(x) dx - \int_{\{e^{-\delta|x|^2}\leq \rho(x)\leq 1\}} |\log \rho(x)|\, \rho(x) dx \\
&\geq -\delta \int_{\{\rho(x)\leq e^{-\delta|x|^2}\}} \sqrt{\rho(x)} dx - \int_{\{e^{-\delta|x|^2}\leq \rho(x)\leq 1\}} |x|^2 \rho(x) dx.
\end{aligned}$$

where we get the last inequality by using the fact that $x|\log(x)| \leq \sqrt{x}$ for $x \in (0, 1]$. For the first term, since $\rho(x) \leq 1$, we estimate:

$$\int_{\{\rho(x)\leq e^{-\delta|x|^2}\}} \sqrt{\rho(x)} dx \leq \int_{\mathbb{R}^d} e^{-\delta|x|^2/2}\, dx = \left(\frac{2\pi}{\delta}\right)^{d/2}.$$

For the second term, note that

$$\int_{\{e^{-\delta|x|^2}\leq \rho(x)\leq 1\}} |x|^2 \rho(x) dx \leq M_2(\rho).$$

Combining the estimates:

$$\int_{\mathbb{R}^d} \log \rho(x)\rho(x) dx \geq -\delta M_2(\rho) - \left(\frac{2\pi}{\delta}\right)^{d/2}.$$

□

Proposition C.5. *Suppose Assumptions 2, 3 and 4 hold. Then we have for all $\delta > 0$*

$$\frac{1}{\sigma} V_\varepsilon^\sigma(m) \geq \frac{1}{\sigma} F_{\min} + C_0 + C_1 M_2(m) - (2\pi/\delta)^{d/2} - 2\delta \left(M_2(m) + \varepsilon^2 M_2(\xi)\right) \tag{39}$$

Proof. We begin by rewriting the regularised energy as the sum of three terms:

$$V_\varepsilon^\sigma(m) = F(m) + \sigma \int \log K_\varepsilon * m(x) dm(x) - \sigma \int \log \pi(x) dm(x).$$

We estimate each term separately. First, thanks to Assumption 2(ii)

$$F(m) \geq F_{\min}.$$

Next, for the third term, note that since $\pi(x) = e^{-U(x)}$, we have

$$-\int \log \pi(x) dm(x) = \int U(x) dm(x) \geq \int C_0 + C_1|x|^2 dm(x) = (C_0 + C_1 M_2(m))$$

For the second term, we have for all $\delta > 0$

$$\begin{aligned}
\int \log(K_\varepsilon * m)(x) dm(x) &= \int \log(\xi_\varepsilon * (\xi_\varepsilon * m))(x) dm(x) \\
&= \int \log\left(\int \xi_\varepsilon(y)(\xi_\varepsilon * m)(x-y) dy\right) dm(x) \\
&\geq \int \left(\int \xi_\varepsilon(y) \log(\xi_\varepsilon * m(x-y)) dy\right) dm(x) \\
&= \int \xi_\varepsilon * \log(\xi_\varepsilon * m)(x) dm(x) \\
&= \int \log(\xi_\varepsilon * m)(x) \xi_\varepsilon * m(x) dx \\
&\geq -(2\pi/\delta)^{d/2} - \delta M_2(\xi_\varepsilon * m).
\end{aligned}$$

where the first inequality is from Jensen's inequality and the last inequality is from Lemma C.4. To control the second moment of the convolution, observe that

$$M_2(\xi_\varepsilon * m) = \int |x|^2 \xi_\varepsilon * m(x) dx = \int |x|^2 \int \xi_\varepsilon(x-y) dm(y) dx = \iint |y+z|^2 \xi_\varepsilon(z) dz dm(y)$$

where in the last step we applied change of variable $z = x - y$. Then by using the inequality $|x+y|^2 \leq 2|x|^2 + 2|y|^2$ we have

$$\begin{aligned}
M_2(\xi_\varepsilon * m) &\leq 2 \iint |z|^2 \xi_\varepsilon(z) dz dm(y) + 2 \iint |y|^2 \xi_\varepsilon(z) dz dm(y) \\
&= 2 \int |z|^2 \xi_\varepsilon(z) dz + 2 \int |y|^2 dm(y) \\
&\leq 2\varepsilon^2 M_2(\xi) + 2M_2(m).
\end{aligned}$$

where we used the remark D.5 in the last step.

Combining all the above estimates, we conclude that $V_\varepsilon^\sigma(m)$ is bounded from above by a constant depending on σ, $F_{\min}$, δ, $M_2(\xi)$, and $M_2(m)$. □

Proof of Theorem C.1 (Step 1: Existence of minimizers). For the non-compact domain $\mathbb{R}^d$, Proposition C.5 plays a twofold role. First, it ensures that the energy V_ε^σ is bounded from below for all $m \in \mathcal{P}_2(\mathbb{R}^d)$, so that minimizing sequences are well defined. Second, for any minimizing sequence, choosing $\delta = C_1/4$ makes the coefficient of $M_2(m)$ positive, yielding the estimate

$$\frac{1}{\sigma} V_\varepsilon^\sigma(m) \geq \frac{1}{\sigma} F_{\min} + C_0 + \frac{C_1}{2} M_2(m) - (4\pi/C_1)^{d/2} - C_1 \varepsilon^2 M_2(\xi).$$

As a consequence, any minimizing sequence $(m_n)_n$ has uniformly bounded second moments. By Prokhorov's theorem and the lower semicontinuity of V_ε^σ, there exists a weakly-$*$ convergent subsequence whose limit is a minimizer. □

C.3 Convergence of minimizers (for both cases)

Lemma C.6. *Let $\mathcal{X} \subset \mathbb{R}^d$ be compact and ξ satisfy Assumption 3, or $\mathcal{X} = \mathbb{R}^d$ and ξ be a Gaussian kernel. Let $(\mu_\varepsilon)_{\varepsilon>0}$ be a sequence in $\mathcal{P}(\mathcal{X})$ such that $\mu_\varepsilon \overset{*}{\rightharpoonup} \mu$ as $\varepsilon \to 0$ for some $\mu \in \mathcal{P}(\mathcal{X})$. Then*

$$\xi_\varepsilon * \mu_\varepsilon \overset{*}{\rightharpoonup} \mu,$$

where the convergence $\overset{*}{\rightharpoonup}$ is understood in the bounded-Lipschitz sense, i.e. tested against all bounded Lipschitz functions $f : \mathcal{X} \to \mathbb{R}$.

Proof. Suppose f be a bounded Lipschitz function. We have

$$\left| \int f d(\xi_\varepsilon * \mu_\varepsilon) - \int f d\mu \right| \leq \left| \int f d(\xi_\varepsilon * \mu_\varepsilon) - \int f d\mu_\varepsilon \right| + \left| \int f d\mu_\varepsilon - \int f d\mu \right|.$$

The second term converges to 0, and for the first term we have

$$\begin{aligned}
\left| \int f d(\xi_\varepsilon * \mu_\varepsilon) - \int f d\mu_\varepsilon \right| &= \left| \int \int (f(x) - f(y)) \xi_\varepsilon(x-y) dy d\mu_\varepsilon(x) \right| \\
&\leq \|\nabla f\|_{L^\infty} \left| \int \int |x-y| \xi_\varepsilon(x-y) dy d\mu_\varepsilon(x) \right| \\
&= \|\nabla f\|_{L^\infty} \int \int |z| \, \xi_\varepsilon(z) \, dz d\mu_\varepsilon(x) \\
&= \|\nabla f\|_{L^\infty} M_1(\xi_\varepsilon).
\end{aligned}$$

In the last inequality, the term $M_1(\xi_\varepsilon)$ can be bounded by $\frac{\varepsilon}{C_{\varepsilon,d}} M_1(\xi)$ or $\varepsilon M_1(\xi)$ depending on the the space $\mathcal{X}$ that we work with. Thus the first term also converges to 0. □

Theorem C.7. *Suppose F is lower semi-continuous. Our energy V_ε^σ Γ-converges to V^σ in the following sense: For $(\mu_\varepsilon)_\varepsilon \subset \mathcal{P}(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{X})$,*

- *if $\mu_\varepsilon \overset{*}{\rightharpoonup} \mu$, we have*
$$\liminf_{\varepsilon \to 0} V_\varepsilon^\sigma(\mu_\varepsilon) \geq V^\sigma(\mu)$$
- $$\limsup_{\varepsilon \to 0} V_\varepsilon^\sigma(\mu) \leq V^\sigma(\mu)$$

Proof. Since for all μ

$$\int \log(K_\varepsilon * \mu) d\mu \geq \int \xi_\varepsilon * \log(\xi_\varepsilon * \mu) d\mu = \int \log(\xi_\varepsilon * \mu) d\xi_\varepsilon * \mu,$$

we have

$$\liminf_{\varepsilon \to 0} V_\varepsilon^\sigma(\mu_\varepsilon) \geq \liminf_{\varepsilon \to 0} \left(F(\mu_\varepsilon) + \sigma \, \mathrm{KL}(\xi_\varepsilon * \mu_\varepsilon | \pi) \right) \geq V^\sigma(\mu)$$

where the last inequality used l.s.c of F and KL as well as the weak convergence of $\xi_\varepsilon * \mu_\varepsilon$ that we got from Lemma C.6. Hence the first item is proven. We have the second item because

$$V^\sigma(\mu) - V_\varepsilon^\sigma(\mu) = \sigma \, \mathrm{KL}(\mu | K_\varepsilon * \mu) \geq 0.$$

□

Proof of Theorem 3.3 and Theorem C.1 (Step 2: Convergence). For any $\varepsilon > 0$, since μ_ε is a minimizer of V_ε^σ, we have for all ν

$$V_\varepsilon^\sigma(\mu_\varepsilon) \leq V_\varepsilon^\sigma(\nu).$$

Consequently,

$$\liminf_{\varepsilon\to 0} V_\varepsilon^\sigma(\mu_\varepsilon) \le \limsup_{\varepsilon\to 0} V_\varepsilon^\sigma(\nu) \le V^\sigma(\nu),$$

where the last inequality follows from Theorem C.7. Since there exists ν such that $V^\sigma(\nu) < \infty$, the sequence $(V_\varepsilon^\sigma(\mu_\varepsilon))_{\varepsilon>0}$ is uniformly bounded from above.

Moreover, by the quadratic growth estimate established above, there exist constants $C_0 \in \mathbb{R}$ and $C_1 > 0$, independent of ε, such that

$$V_\varepsilon^\sigma(\mu_\varepsilon) \ge F_{\min} + C_0 + \frac{C_1}{2} M_2(\mu_\varepsilon) - (4\pi/C_1)^{d/2} - C_1\varepsilon^2\, M_2(\xi).$$

Since the left-hand side is uniformly bounded from above and the last term is uniformly bounded as $\varepsilon \to 0$, it follows that $(M_2(\mu_\varepsilon))_{\varepsilon>0}$ is uniformly bounded.

If $\mathcal{X}$ is compact, tightness of $(\mu_\varepsilon)_{\varepsilon>0}$ is immediate. If we work with $\mathbb{R}^d$, the uniform second-moment bound implies tightness. Therefore, up to a subsequence, $\mu_\varepsilon \rightharpoonup \mu$ weakly for some $\mu \in \mathcal{P}_2(\mathcal{X})$.

By the Γ-convergence, we conclude that for all ν

$$V^\sigma(\mu) \le \liminf_{\varepsilon\to 0} V_\varepsilon^\sigma(\mu_\varepsilon) \le V^\sigma(\nu),$$

which shows that μ is a minimizer of V^σ. □

D SUPPLEMENTARY DEFINITIONS AND TECHNICAL RESULTS

D.1 Flat Derivative

Definition D.1. Fix $q \ge 0$ and let $\mathcal{P}_q(\mathcal{X})$ be the space of probability measures on $\mathcal{X}$ with finite q-th moments. A functional $F : \mathcal{P}_q(\mathcal{X}) \to \mathbb{R}$, is said to admit a first order linear derivative (or a flat derivative), if there exists a functional $\frac{\delta F}{\delta m} : \mathcal{P}_q(\mathcal{X}) \times \mathcal{X} \to \mathbb{R}$, such that

1. For all $a \in \mathcal{X}$, $\mathcal{P}_q(\mathcal{X}) \ni m \mapsto \frac{\delta F}{\delta m}(m, a)$ is continuous.
2. For any $\nu \in \mathcal{P}_q(\mathcal{X})$, there exists $C > 0$ such that for all $a \in \mathcal{X}$ we have
$$\left|\frac{\delta F}{\delta m}(\nu, a)\right| \le C(1 + |a|^q)\,.$$
3. For all m, $m' \in \mathcal{P}_q(\mathcal{X})$,
$$F(m') - F(m) = \int_0^1 \int_{\mathcal{X}} \frac{\delta F}{\delta m}(m + \lambda(m' - m), a)\,(m' - m)\,(da)\,d\lambda. \tag{40}$$

The functional $\frac{\delta F}{\delta m}$ is then called the linear (functional) derivative of F on $\mathcal{P}_q(\mathcal{X})$.

D.2 Technical Estimates

Proposition D.2 (Uniform Bounds and Wasserstein Stability of the Projection)**.** *Suppose Assumption 6 holds, and let $T > 0$ be a finite time horizon. Then the following statements hold for all $t \in [0, T]$:*

(i) *Let w_t be a solution to the particle dynamics* (13)*. Then w_t remains uniformly bounded as*
$$e^{-M_a T} \le w_t \le e^{M_a T}. \tag{41}$$

(ii) *Let $M > 0$ and let $\nu, \nu' \in \mathcal{P}_2\left(\mathcal{C}([0, T]; \mathcal{X} \times [0, M])\right)$ be two probability measures on the path space. Then their projected measures satisfy the following Wasserstein stability estimate:*
$$\mathcal{W}_2^2\left(\hbar\nu_t, \hbar\nu_t'\right) \le M^2\, \mathcal{W}_2^2\left(\nu_t, \nu_t'\right), \quad \text{for all } t \in [0, T].$$
In particular, if ν and ν' are the laws of solutions to (13)*, then due to* (41) *we obtain*
$$\mathcal{W}_2^2\left(\hbar\nu_t, \hbar\nu_t'\right) \le e^{2M_a T}\, \mathcal{W}_2^2\left(\nu_t, \nu_t'\right), \quad \text{for all } t \in [0, T]. \tag{42}$$

Proof. The first item is proven directly by the boundedness assumption. We have

$$w_t = e^{-\int_0^t a(\mathtt{h}\nu_t, X)ds}.$$

Hence

$$e^{-M_a T} = e^{\int_0^T -M_a dt} \leq w_t \leq e^{\int_0^T M_a dt} = e^{M_a T}.$$

For the second item, let denote π_t be any coupling for ν_t and ν'_t. We denote that $\tilde{\mathtt{h}}\pi_t$ is the coupling for $\mathtt{h}\nu_t$ and $\mathtt{h}\nu'_t$. It is easy to check that $\tilde{\mathtt{h}}\pi_t$ is the projection (different to (8), this projection is for product space) of π_t in the sense that for any test function f we have

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{X}} f(x_t, x'_t)\tilde{\mathtt{h}}\pi_t(dx_t, dx'_t) \\
&= \int_{(\mathcal{X}\times[0,M])\times(\mathcal{X}\times[0,M])} w_t w'_t f(x_t, x'_t)\pi_t(dx_t, dw_t, dx'_t, dw'_t).
\end{aligned}$$

Hence we have

$$\begin{aligned}
&\int_{\mathcal{X}\times\mathcal{X}} |x_t - x'_t|^2 \tilde{\mathtt{h}}\pi_t(dx_t, dx'_t) \\
&= \int_{(\mathcal{X}\times[0,M])\times(\mathcal{X}\times[0,M])} w_t w'_t |x_t - x'_t|^2 \pi_t(dx_t, dw_t, dx'_t, dw'_t) \\
&\leq M^2 \int_{(\mathcal{X}\times[0,M])\times(\mathcal{X}\times[0,M])} |x_t - x'_t|^2 \pi_t(dx_t, dw_t, dx'_t, dw'_t) \\
&\leq M^2 \int_{(\mathcal{X}\times[0,M])\times(\mathcal{X}\times[0,M])} \left(|x_t - x'_t|^2 + |w_t - w'_t|^2\right) \pi_t(dx_t, dw_t, dx'_t, dw'_t)
\end{aligned}$$

which yields the result if we take infimum on both sides. □

Proposition D.3. *Let μ, $\mu' \in \mathcal{P}_2(\mathcal{C}([0,T];\mathcal{X}))$. Then, for any $t \in [0,T]$,*

$$\mathcal{W}_2(\mu_t, \mu'_t) \leq \mathcal{W}_{2,t}(\mu, \mu').$$

Proof. Let Π be a coupling of μ and μ' and for a fixed $t \in [0,T]$ consider a projection map $\Gamma_t : \mathcal{C}([0,T];\mathcal{X}) \to \mathcal{X}$ given as $\Gamma_t(\gamma) := \gamma_t$. Let π_t be the push-forward of Π under Γ_t. Then

$$\int_{\mathcal{X}\times\mathcal{X}} |x-y|^2 \pi_t(dx, dy) = \int_{\mathcal{C}([0,T];\mathcal{X})\times\mathcal{C}([0,T];\mathcal{X})} |\gamma_t - \gamma'_t|^2 \Pi(d\gamma, d\gamma').$$

Note that π_t constructed this way is a coupling of $\mu_t = \Gamma_t(\mu)$ and $\mu'_t = \Gamma_t(\mu')$ and hence, using $|\gamma_t - \gamma'_t| \leq \sup_{s\in[0,t]} |\gamma_s - \gamma'_s|$ and then taking the infimum over Π on the right-hand side, we get

$$\mathcal{W}_2^2(\mu_t, \mu'_t) \leq \int_{\mathcal{X}\times\mathcal{X}} |x-y|^2 \pi_t(dx, dy) \leq \mathcal{W}_{2,t}^2(\mu, \mu').$$

□

D.3 Construction of the Kernel on a compact space

Our example is based on the standard Gaussian kernel truncated and renormalized on a compact space. Let ξ denote the standard Gaussian density on $\mathbb{R}^d$:

$$\xi(z) := \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|z|^2}{2}\right).$$

We then define the mapping

$$\xi_\varepsilon : \mathcal{X} \to \mathbb{R}_+, \quad \xi_\varepsilon(x) := \frac{1}{C_{\varepsilon,d,L}} \cdot \frac{1}{\varepsilon^d} \cdot \xi\left(\frac{x}{\varepsilon}\right),$$

where $C_{\varepsilon,d,L}$ is the normalization constant ensuring $\int_{\mathcal{X}} \xi_\varepsilon(x)\,dx = 1$. We define

$$C_{\varepsilon,d,L} := \int_{\mathcal{X}} \frac{1}{\varepsilon^d} \xi\left(\frac{x}{\varepsilon}\right) dx$$

and we have

$$C_{\varepsilon,d} = \left(2\Phi\left(\frac{L}{\varepsilon}\right) - 1\right)^d,$$

where Φ denotes the cumulative distribution function of the standard normal distribution. As $\varepsilon \to 0$, we observe that $C_{\varepsilon,d,L} \nearrow 1$. Thus, if we restrict ε to a compact interval $(0, \varepsilon_{\max}]$, then

$$C_{\varepsilon,d,L} \in \left[\left(2\Phi\left(\frac{L}{\varepsilon_{\max}}\right) - 1\right)^d, 1\right). \tag{43}$$

We now examine the properties of K_ε.

1. **Normalization.** By construction, ξ_ε is a probability density function on $\mathcal{X}$. Consequently, its convolution is also a probability density function on $\mathcal{X}$.

2. **Upper and lower bounds.** Since ξ is smooth and strictly positive, so is ξ_ε. Moreover, since

$$\xi_\varepsilon(x) = \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^d} \cdot \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|x|^2}{2\varepsilon^2}\right),$$

we obtain the following pointwise bounds for $x \in \mathcal{X}$:

$$\sup_{x\in\mathcal{X}} \xi_\varepsilon(x) \le \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^d} \cdot \frac{1}{(2\pi)^{d/2}},$$
$$\inf_{x\in\mathcal{X}} \xi_\varepsilon(x) \ge \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^d} \cdot \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{dL^2}{2\varepsilon^2}\right).$$

Since $K_\varepsilon = \xi_\varepsilon * \xi_\varepsilon$, the same type of bounds hold for K_ε. In particular, we may introduce the constants

$$K_{\max,\varepsilon} := \sup_{x\in\mathcal{X}} K_\varepsilon(x), \qquad K_{\min,\varepsilon} := \inf_{x\in\mathcal{X}} K_\varepsilon(x),$$

which are strictly positive and finite. Note that $K_{\max,\varepsilon}$ and $K_{\min,\varepsilon}$ depend on (ε, d, L), but for simplicity of notation we suppress this dependence.

3. **Lipschitz continuity.** Since ξ is smooth, we have

$$\begin{aligned}
|\nabla \xi_\varepsilon(x)| &= \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^d} \cdot \frac{1}{(2\pi)^{d/2}} \left|\nabla \exp\left(-\frac{|x|^2}{2\varepsilon^2}\right)\right| \\
&= \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^d} \cdot \frac{1}{(2\pi)^{d/2}} \frac{|x|}{\varepsilon^2} \exp\left(-\frac{|x|^2}{2\varepsilon^2}\right) \\
&\le \frac{1}{C_{\varepsilon,d}} \cdot \frac{1}{\varepsilon^{d+1}} \cdot \frac{1}{(2\pi)^{d/2}}\, e^{-1/2},
\end{aligned}$$

where the last inequality follows from the fact that $re^{-r^2/2} \le e^{-1/2}$ for all $r \ge 0$.

Therefore, ξ_ε is globally Lipschitz with constant of order $O(\varepsilon^{-(d+1)})$. Moreover, since

$$\nabla K_\varepsilon = (\nabla \xi_\varepsilon) * \xi_\varepsilon,$$

we obtain

$$\|\nabla K_\varepsilon\|_\infty \le \|\nabla \xi_\varepsilon\|_\infty \, \|\xi_\varepsilon\|_{L^1} = \|\nabla \xi_\varepsilon\|_\infty,$$

which shows that K_ε inherits the same Lipschitz constant. Hence K_ε is Lipschitz continuous on $\mathcal{X}$ with constant of order $O(\varepsilon^{-(d+1)})$.

Remark D.4. The kernel K_ε defined above is a smooth function on $\mathcal{X}$, bounded from below and above and globally Lipschitz. More precisely, there exist positive constants

$$0 < K_{\min,\varepsilon} \le K_{\max,\varepsilon} < \infty, \quad L_{K_\varepsilon} < \infty,$$

such that

$$K_{\min,\varepsilon} \le K_\varepsilon(x) \le K_{\max,\varepsilon} \tag{44}$$

and

$$|K_\varepsilon(x) - K_\varepsilon(y)| \le L_{K_\varepsilon}\, |x-y|, \quad \forall\, x, y \in \mathcal{X}.$$

The values of $K_{\min,\varepsilon}$, $K_{\max,\varepsilon}$, and L_{K_ε} depend on the dimension d, the diameter of $\mathcal{X}$, the parameter ε, and the choice of the base function ξ.

Remark D.5. For $p \ge 0$ and an integrable kernel $\xi : \mathbb{R}^d \to [0,\infty)$. Suppose $M_p(\xi) := \int_{\mathbb{R}^d} |u|^p\, \xi(u)\, du < \infty$. Let ξ_ε be as in Assumption 3. Then the p-th moment of ξ_ε satisfies

$$\begin{aligned} M_p(\xi_\varepsilon) &:= \int_{\mathcal{X}} |x|^p\, \xi_\varepsilon(x)\, dx \\ &= \frac{\varepsilon^p}{C_{\varepsilon,d}} \int_{[-L/\varepsilon, L/\varepsilon]^d} |u|^p\, \xi(u)\, du \\ &\le \frac{\varepsilon^p}{C_{\varepsilon,d}}\, M_p(\xi). \end{aligned}$$

E VERIFICATION OF THE BOUNDNESS AND LIPSCHITZ CONDITION

The main results above—existence and uniqueness (Theorem 2.9) and propagation of chaos (Theorem 2.10)—require that the drift $a(m,x)$ be uniformly bounded and Lipschitz continuous in both the spatial variable x and the measure variable m. We prove that these properties hold for a broad class of kernel-based drifts a. The resulting bounds are explicit and depend only on the smoothing kernel, the reference density, and the diameter of $\mathcal{X}$.

Proposition E.1. *Let $a(m,x)$ be given by any of the kernelized forms* (25)–(28)*. Suppose Assumptions 2, 3, and 4 hold. Then $a(m,x)$ satisfies the boundedness and Lipschitz conditions* (6) *and* (7)*.*

The Lipschitz continuity and boundedness conditions can also be extended to a Fisher–Rao gradient flow for an energy functional with the kernelized chi-square divergence as a regularizer.

Proposition E.2. *Consider the energy functional*

$$F(m) + \sigma \int \left(\frac{K_\varepsilon * m}{\pi} - 1 \right)^2 \pi(x)\, dx.$$

Its corresponding Fisher–Rao gradient flow takes the form (4)*, where*

$$a(m,x) = \frac{\delta F}{\delta \mu}(m,x) + 2\sigma\, K_\varepsilon * \left(\frac{K_\varepsilon * m}{\pi} \right)(x) - 2\sigma \int K_\varepsilon * \left(\frac{K_\varepsilon * m}{\pi} \right)(y)\, m(dy).$$

Suppose that Assumptions 2, 3, and 4 hold. Then $a(m,x)$ satisfies both the boundedness condition (6) *and the Lipschitz continuity condition* (7)*.*

Proposition E.3. *For the boundedness condition, we have the following estimates (uniform in $\mu \in \mathcal{P}(\mathcal{X})$ and $x \in \mathcal{X}$):*

(1) $\left| \frac{\delta F}{\delta \mu}(\mu, x) \right| \le C_F$.

(2) $\left| \log \frac{K_\varepsilon * \mu(x)}{\pi(x)} \right| \le \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}} \right)$.

(3) $\mathrm{KL}(K_\varepsilon * \mu | \pi) \le \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}$.

(4) $\left|\log \frac{K_\varepsilon * \mu(x)}{K_\varepsilon * \pi(x)}\right| \le \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}\right).$

(5) $\mathrm{KL}(K_\varepsilon * \mu | K_\varepsilon * \pi) \le \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}.$

(6) $K_\varepsilon * \left(\frac{\mu}{K_\varepsilon * \mu}\right)(x) \le \frac{K_{\max,\varepsilon}}{K_{\min,\varepsilon}}.$

(7) $\frac{1}{(K_\varepsilon * \mu)(x)} \le \frac{1}{K_{\min,\varepsilon}}.$

Proof. E.3(1) follows from Assumption 2. For E.3(2), E.3(3), E.3(4), and E.3(5), we use

$$\frac{K_{\min,\varepsilon}}{\pi_{\max}} \le \frac{K_\varepsilon * \mu(x)}{\pi(x)} \le \frac{K_{\max,\varepsilon}}{\pi_{\min}}, \quad \forall x \in \mathcal{X},$$

which follows from Assumptions 3 and 4. For E.3(6), we compute

$$\begin{aligned} K_\varepsilon * \left(\frac{\mu}{K_\varepsilon * \mu}\right)(x) &= \int K_\varepsilon(x-y) \left(\frac{\mu}{K_\varepsilon * \mu}\right)(y)\, dy \\ &= \int \frac{K_\varepsilon(x-y)}{K_\varepsilon * \mu(y)}\, \mu(y)\, dy \\ &\le \frac{K_{\max,\varepsilon}}{K_{\min,\varepsilon}}. \end{aligned}$$

Finally, E.3(7) is immediate from $(K_\varepsilon * \mu)(x) \ge K_{\min,\varepsilon}$. □

Proposition E.4. *For all $\mu, \nu \in \mathcal{P}_2(\mathcal{X})$ and $x, y \in \mathcal{X}$, the following quantitative bounds hold:*

(1)
$$\left|\frac{\delta F}{\delta \mu}(\mu, x) - \frac{\delta F}{\delta \mu}(\nu, y)\right| \le L_F\big(|x-y| + \mathcal{W}_2(\mu,\nu)\big).$$

(2)
$$|\log K_\varepsilon * \mu(x) - \log K_\varepsilon * \nu(x)| \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\, \mathcal{W}_1(\mu,\nu).$$

(3)
$$|\log K_\varepsilon * \nu(x) - \log K_\varepsilon * \nu(y)| \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\, |x-y|.$$

(4)
$$|\log K_\varepsilon * \pi(x) - \log K_\varepsilon * \pi(y)| \le \frac{L_\pi}{\pi_{\min}}\, |x-y|.$$

(5)
$$\left|\log \frac{K_\varepsilon * \nu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(y)}{\pi(y)}\right| \le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) |x-y|.$$

(6)
$$\left|K_\varepsilon * \left(\log \frac{K_\varepsilon * \nu}{\pi}\right)(x) - K_\varepsilon * \left(\log \frac{K_\varepsilon * \nu}{\pi}\right)(y)\right| \le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) |x-y|.$$

(7)
$$\left|K_\varepsilon * \left(\log \frac{K_\varepsilon * \nu}{K_\varepsilon * \pi}\right)(x) - K_\varepsilon * \left(\log \frac{K_\varepsilon * \nu}{K_\varepsilon * \pi}\right)(y)\right| \le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) |x-y|.$$

(8)
$$|\mathrm{KL}(K_\varepsilon * \mu \mid \pi) - \mathrm{KL}(K_\varepsilon * \nu \mid \pi)| \le \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) \mathcal{W}_1(\mu,\nu).$$

(9)

$$\left|\frac{1}{(K_\varepsilon * \mu)(x)} - \frac{1}{(K_\varepsilon * \mu)(y)}\right| \leq \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\,|x-y|.$$

Proof. For (E.4(1)), we use Assumption 2.

For (E.4(2)),

$$\begin{aligned}
|\log(K_\varepsilon * \mu)(x) - \log(K_\varepsilon * \nu)(x)| &\leq \frac{1}{K_{\min,\varepsilon}}\,|(K_\varepsilon * \mu)(x) - (K_\varepsilon * \nu)(x)| \\
&= \frac{1}{K_{\min,\varepsilon}}\left|\int K_\varepsilon(x-z)\,(\mu-\nu)(dz)\right| \\
&\leq \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\,\mathcal{W}_1(\mu,\nu),
\end{aligned}$$

For (E.4(3)), use that log is $1/K_{\min,\varepsilon}$-Lipschitz on $[K_{\min,\varepsilon},\infty)$ and that $x \mapsto (K_\varepsilon * \nu)(x)$ is L_{K_ε}-Lipschitz. Hence

$$\begin{aligned}
|\log K_\varepsilon * \nu(x) - \log K_\varepsilon * \nu(y)| &\leq \frac{1}{K_{\min,\varepsilon}}\,|K_\varepsilon * \nu(x) - K_\varepsilon * \nu(y)| \\
&= \frac{1}{K_{\min,\varepsilon}}\left|\int \big(K_\varepsilon(z-x) - K_\varepsilon(z-y)\big)\nu(dz)\right| \\
&\leq \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\,|x-y|.
\end{aligned}$$

For (E.4(4)), again we use that log is $1/\pi_{\min}$-Lipschitz on $[\pi_{\min},\infty)$ and that π is L_π-Lipschitz. Hence

$$\begin{aligned}
|\log K_\varepsilon * \pi(x) - \log K_\varepsilon * \pi(y)| &\leq \frac{1}{\pi_{\min}}\,|K_\varepsilon * \pi(x) - K_\varepsilon * \pi(y)| \\
&\leq \frac{1}{\pi_{\min}}\int K_\varepsilon(z)\,|\pi(x-z) - \pi(y-z)|\,dz \\
&\leq \frac{L_\pi}{\pi_{\min}}\,|x-y|.
\end{aligned}$$

For (E.4(5)), again we use the Lipschitz condition of log and we have

$$\begin{aligned}
\left|\log\frac{K_\varepsilon * \nu(x)}{\pi(x)} - \log\frac{K_\varepsilon * \nu(y)}{\pi(y)}\right| &\leq |\log K_\varepsilon * \nu(x) - \log K_\varepsilon * \nu(y)| + |\log\pi(x) - \log\pi(y)| \\
&\leq \frac{1}{K_{\min,\varepsilon}}\,|K_\varepsilon * \nu(x) - K_\varepsilon * \nu(y)| + \frac{1}{\pi_{\min}}\,|\pi(x)-\pi(y)| \\
&= \frac{1}{K_{\min,\varepsilon}}\left|\int \big(K_\varepsilon(z-x) - K_\varepsilon(z-y)\big)\nu(dz)\right| + \frac{1}{\pi_{\min}}\,|\pi(x)-\pi(y)| \\
&\leq \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right)|x-y|.
\end{aligned} \tag{45}$$

For (E.4(6)), we use (E.4(5)) and the fact that $\mathrm{Lip}(K_\varepsilon * f) \leq \mathrm{Lip}(f)$.

For (E.4(7)), we use (E.4(3)), (E.4(4)), and $\mathrm{Lip}(K_\varepsilon * f) \leq \mathrm{Lip}(f)$.

For (E.4(8)), we have

$$
\begin{aligned}
|\mathrm{KL}(K_\varepsilon * \mu \mid \pi) - \mathrm{KL}(K_\varepsilon * \nu \mid \pi)| &\le \left| \int \left(\log \frac{K_\varepsilon * \mu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(x)}{\pi(x)} \right) (K_\varepsilon * \mu)(x)\, dx \right| \\
&\quad + \left| \int \big((K_\varepsilon * \mu)(x) - (K_\varepsilon * \nu)(x) \big) \log \frac{K_\varepsilon * \nu(x)}{\pi(x)}\, dx \right| \\
&\le \int |\log K_\varepsilon * \mu(x) - \log K_\varepsilon * \nu(x)|\, (K_\varepsilon * \mu)(x)\, dx \\
&\quad + \left| \int (\mu(x) - \nu(x))\, K_\varepsilon * \left(\log \frac{K_\varepsilon * \nu(x)}{\pi(x)} \right) dx \right| \\
&\le \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}} \right) \mathcal{W}_1(\mu, \nu),
\end{aligned}
$$

where in the last line we used (E.4(2)) and (E.4(4)).

For (E.4(9)),

$$
\begin{aligned}
\left| \frac{1}{(K_\varepsilon * \mu)(x)} - \frac{1}{(K_\varepsilon * \mu)(y)} \right| &= \left| \frac{(K_\varepsilon * \mu)(x) - (K_\varepsilon * \mu)(y)}{(K_\varepsilon * \mu)(x)(K_\varepsilon * \mu)(y)} \right| \\
&\le \frac{1}{K_{\min,\varepsilon}^2}\, |(K_\varepsilon * \mu)(x) - (K_\varepsilon * \mu)(y)| \\
&\le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\, |x - y|.
\end{aligned}
$$

□

E.1 For Flow (25)

Define the coefficient associated with the kernelized flow (25) by

$$
a(\mu, x) := \frac{\delta F}{\delta \mu}(\mu, x) + \sigma \Big((K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{\pi})(x) - \mathrm{KL}(K_\varepsilon * \mu | \pi) \Big).
$$

By Proposition E.1, a satisfies both a boundedness estimate and a Lipschitz estimate. In particular, under Assumptions 2, 3, and 4, there exist constants $C_\varepsilon^1 > 0$ and $L_\varepsilon^1 > 0$ (given explicitly in Proposition E.1) such that, for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$,

$$
|a(\mu, x)| \le C_\varepsilon^1, \qquad |a(\mu, x) - a(\nu, y)| \le L_\varepsilon^1 \big(\mathcal{W}_2(\mu, \nu) + |x - y| \big).
$$

Proof. Thanks to (E.3(1)) and (E.3(2)) we have

$$
\left| \log \frac{K_\varepsilon * \mu(x)}{\pi(x)} \right| \le \max \left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}} \right),
$$

and

$$
\mathrm{KL}\left(K_\varepsilon * \mu | \pi \right) \le \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}. \tag{46}
$$

Therefore, the desired boundedness condition holds with the constant

$$
C_\varepsilon^1 := C_F + \sigma \max \left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}} \right) + \sigma \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}.
$$

We now turn to the proof of the Lipschitz condition. By the triangle inequality,

$$
|a(\mu, x) - a(\nu, y)| \le (i) + \sigma(ii) + \sigma(iii),
$$

where

$$(i) := \left|\frac{\delta F}{\delta \mu}(\mu, x) - \frac{\delta F}{\delta \mu}(\nu, y)\right|, \quad (ii) := \left|\log \frac{K_\varepsilon * \mu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(y)}{\pi(y)}\right|,$$

$$(iii) := \left|\mathrm{KL}(K_\varepsilon * \mu \mid \pi) - \mathrm{KL}(K_\varepsilon * \nu \mid \pi)\right|.$$

Thanks to (E.4(1)), we have

$$(i) \leq L_F \left(\mathcal{W}_2(\mu, \nu) + |x - y|\right). \tag{47}$$

For the term (ii), by triangle inequality we have

$$(ii) \leq \left|\log \frac{K_\varepsilon * \mu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(x)}{\pi(x)}\right| + \left|\log \frac{K_\varepsilon * \nu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(y)}{\pi(y)}\right|.$$

We bound these two term separately by (E.4(2)) and (E.4(4)) and we get:

$$(ii) \leq \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} \mathcal{W}_1(\mu, \nu) + \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) |x - y|. \tag{48}$$

Finally, for the terms (iii), thanks to (E.4(5)) we have

$$(iii) \leq \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) \mathcal{W}_1(\mu, \nu) \tag{49}$$

Combining the bounds from (47), (48), and (49), and using the fact that $\mathcal{W}_1(\mu, \nu) \leq \mathcal{W}_2(\mu, \nu)$, we obtain the desired Lipschitz estimate with the stated constant.

$$L_\varepsilon^1 := L_F + \sigma \left(\frac{3L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{2L_\pi}{\pi_{\min}}\right).$$

□

E.2 For Flow (26)

Define the coefficient associated with the kernelized flow (26) by

$$a(\mu, x) := \frac{\delta F}{\delta \mu}(\mu, x) + \sigma\Big(\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{K_\varepsilon * \pi}\big)(x) - \mathrm{KL}(K_\varepsilon * \mu | K_\varepsilon * \pi)\Big).$$

By Proposition E.1, a satisfies both a boundedness estimate and a Lipschitz estimate. In particular, under Assumptions 2, 3, and 4, there exist constants $C_\varepsilon^2 > 0$ and $L_\varepsilon^2 > 0$ such that, for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$,

$$|a(\mu, x)| \leq C_\varepsilon^2, \qquad |a(\mu, x) - a(\nu, y)| \leq L_\varepsilon^2\big(\mathcal{W}_2(\mu, \nu) + |x - y|\big).$$

Proof. Thanks to (E.3(3)) and (E.3(4)) we have

$$\left|\log \frac{K_\varepsilon * \mu(x)}{K_\varepsilon * \pi(x)}\right| \leq \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}\right),$$

and

$$\mathrm{KL}\left(K_\varepsilon * \mu | K_\varepsilon * \pi\right) \leq \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}. \tag{50}$$

Therefore, the desired boundedness condition holds with the constant

$$C_\varepsilon^2 := C_F + \sigma \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}\right) + \sigma \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}.$$

We now turn to the proof of the Lipschitz condition. By the triangle inequality,

$$|a(\mu,x)-a(\nu,y)| \le (i)+\sigma(ii)+\sigma(iii),$$

where

$$\begin{aligned}(i) &:= \left|\frac{\delta F}{\delta\mu}(\mu,x)-\frac{\delta F}{\delta\mu}(\nu,y)\right|,\\ (ii) &:= \left|\log\frac{K_\varepsilon*\mu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(y)}{K_\varepsilon*\pi(y)}\right|,\\ (iii) &:= \left|\mathrm{KL}(K_\varepsilon*\mu \mid K_\varepsilon*\pi)-\mathrm{KL}(K_\varepsilon*\nu \mid K_\varepsilon*\pi)\right|.\end{aligned}$$

We will show separately that each part is Lipschitz. Thanks to (E.4(1)), we have

$$(i) \le L_F\left(\mathcal{W}_2(\mu,\nu)+|x-y|\right). \tag{51}$$

For the term (ii), by triangle inequality we have

$$(ii) \le \left|\log\frac{K_\varepsilon*\mu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}\right|+\left|\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(y)}{K_\varepsilon*\pi(y)}\right|$$

For the first term on RHS, thanks to (E.4(2))we have

$$\left|\log\frac{K_\varepsilon*\mu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}\right| = |\log K_\varepsilon*\mu(x)-\log K_\varepsilon*\nu(x)| \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\mathcal{W}_1(\mu,\nu).$$

For the second term on RHS, we have

$$\begin{aligned}&\left|\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(y)}{K_\varepsilon*\pi(y)}\right|\\ &\le |\log K_\varepsilon*\nu(x)-\log K_\varepsilon*\nu(y)|+|\log K_\varepsilon*\pi(x)-\log K_\varepsilon*\pi(y)|\\ &\le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}+\frac{L_\pi}{\pi_{\min}}\right)|x-y|\end{aligned}$$

where the second inequality follows from (E.4(6)) and (E.4(7)). Hence

$$(ii) \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\mathcal{W}_1(\mu,\nu)+\left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}+\frac{L_\pi}{\pi_{\min}}\right)|x-y|. \tag{52}$$

Regarding the term (iii), we have

$$\begin{aligned}(iii) &:= |\mathrm{KL}(K_\varepsilon*\mu|K_\varepsilon*\pi)-\mathrm{KL}(K_\varepsilon*\nu|K_\varepsilon*\pi)|\\ &\le \left|\int\left(\log\frac{K_\varepsilon*\mu(x)}{K_\varepsilon*\pi(x)}-\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}\right)K_\varepsilon*\mu(x)dx\right|\\ &+\left|\int(K_\varepsilon*\mu(x)-K_\varepsilon*\nu(x))\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}dx\right|\\ &\le \int|\log K_\varepsilon*\mu(x)-\log K_\varepsilon*\nu(x)|\,K_\varepsilon*\mu(x)dx\\ &+\left|\int(\mu(x)-\nu(x))\,K_\varepsilon*\left(\log\frac{K_\varepsilon*\nu(x)}{K_\varepsilon*\pi(x)}\right)dx\right|\\ &\le \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}}+\frac{L_\pi}{\pi_{\min}}\right)\mathcal{W}_1(\mu,\nu),\end{aligned} \tag{53}$$

where in the last inequality, we used (E.4(2)) and (E.4(7)). Combining (51), (52), and (53) with the fact that $\mathcal{W}_1(\mu,\nu)\le\mathcal{W}_2(\mu,\nu)$, we conclude that $L^2_\varepsilon = L^1_\varepsilon$, which completes the proof.

□

E.3 For Flow (27)

Define the coefficient associated with the kernelized flow (27) by

$$a(\mu,x) := \frac{\delta F}{\delta\mu}(\mu,x) + \sigma\left(\log\frac{(K_\varepsilon*\mu)(x)}{\pi(x)} + \Big(K_\varepsilon * \frac{\mu}{K_\varepsilon*\mu}\Big)(x) - \int_{\mathcal{X}} \log\frac{(K_\varepsilon*\mu)(z)}{\pi(z)}\,\mu(z)\,dz - 1\right).$$

Under Assumptions 2, 3, and 4, and by Proposition E.1, there exist constants $C_\varepsilon^3 > 0$ and $L_\varepsilon^3 > 0$ such that, for all $\mu,\nu \in \mathcal{P}(\mathcal{X})$ and $x,y \in \mathcal{X}$,

$$|a(\mu,x)| \le C_\varepsilon^3, \qquad |a(\mu,x) - a(\nu,y)| \le L_\varepsilon^3\big(\mathcal{W}_2(\mu,\nu) + |x-y|\big).$$

Proof. For the boundedness condition, we have already established uniform bounds for $\frac{\delta F}{\delta\mu}(\mu,x)$ and $\log\frac{K_\varepsilon*\mu(x)}{\pi(x)}$ from (E.3(1)) and (E.3(2)). In particular, we recall that

$$\left|\frac{\delta F}{\delta\mu}(\mu,x)\right| + \left|\log\tfrac{K_\varepsilon*\mu(x)}{\pi(x)}\right| + \left|\int \log\tfrac{K_\varepsilon*\mu(x)}{\pi(x)}\,\mu(x)\,dx\right| \le C_F + 2\max\left(\log\frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log\frac{K_{\max,\varepsilon}}{\pi_{\min}}\right).$$

Moreover, thanks to (E.3(6)), for positive term $K_\varepsilon * \left(\frac{\mu}{K_\varepsilon*\mu}\right)$, we have

$$K_\varepsilon * \left(\frac{\mu}{K_\varepsilon*\mu}\right)(x) \le \frac{K_{\max,\varepsilon}}{K_{\min,\varepsilon}}.$$

Hence we conclude that

$$C_\varepsilon^3 = C_F + 2\sigma\max\left(\log\frac{\pi_{\max}}{K_{\min,\varepsilon}}, \log\frac{K_{\max,\varepsilon}}{\pi_{\min}}\right) + \sigma\left(\tfrac{K_{\max,\varepsilon}}{K_{\min,\varepsilon}} + 1\right).$$

We now turn to the proof of the Lipschitz condition. By the triangle inequality,

$$|a(\mu,x) - a(\nu,y)| \le (i) + \sigma(ii) + \sigma(iii) + \sigma(iv),$$

where

$$(i) := \left|\frac{\delta F}{\delta\mu}(\mu,x) - \frac{\delta F}{\delta\mu}(\nu,y)\right|,$$

$$(ii) := \left|\log\frac{K_\varepsilon*\mu(x)}{\pi(x)} - \log\frac{K_\varepsilon*\nu(y)}{\pi(y)}\right|,$$

$$(iii) := \left|K_\varepsilon * \left(\tfrac{\mu}{K_\varepsilon*\mu}\right)(x) - K_\varepsilon * \left(\tfrac{\nu}{K_\varepsilon*\nu}\right)(y)\right|,$$

$$(iv) := \left|\int \log\frac{K_\varepsilon*\mu(z)}{\pi(z)}\,\mu(z)\,dz - \int \log\frac{K_\varepsilon*\nu(z)}{\pi(z)}\,\nu(z)\,dz\right|.$$

Thanks to (E.4(1)), we have

$$(i) \le L_F\left(\mathcal{W}_2(\mu,\nu) + |x-y|\right). \tag{54}$$

For the term (ii), by triangle inequality we have

$$(ii) \le \left|\log\frac{K_\varepsilon*\mu(x)}{\pi(x)} - \log\frac{K_\varepsilon*\nu(x)}{\pi(x)}\right| + \left|\log\frac{K_\varepsilon*\nu(x)}{\pi(x)} - \log\frac{K_\varepsilon*\nu(y)}{\pi(y)}\right|.$$

We bound these two term separately by (E.4(2)) and (E.4(4)) and we get:

$$(ii) \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}}\mathcal{W}_1(\mu,\nu) + \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right)|x-y|. \tag{55}$$

Now let us check item (iii). We write

$$h_\mu(x) = \frac{1}{(K_\varepsilon*\mu)(x)}, \qquad h_\nu(x) = \frac{1}{(K_\varepsilon*\nu)(x)}.$$

Decompose

$$
\begin{aligned}
&K_\varepsilon * \left(\frac{\mu}{K_\varepsilon * \mu}\right)(y) - K_\varepsilon * \left(\frac{\nu}{K_\varepsilon * \nu}\right)(y) \\
&= \int K_\varepsilon(y-x) h_\mu(x)\, d(\mu-\nu)(x) + \int K_\varepsilon(y-x)\big(h_\mu(x) - h_\nu(x)\big)\, d\nu(x) \\
&=: T_1 + T_2.
\end{aligned}
$$

For T_1, set $f_\mu(x) = K_\varepsilon(y-x)h_\mu(x)$. Using (E.3(7)) and (E.4(9)) and the product rule for Lipschitz functions,

$$
\mathrm{Lip}(f_\mu) \le \mathrm{Lip}(K_\varepsilon)\|h_\mu\|_\infty + \|K_\varepsilon\|_\infty \,\mathrm{Lip}(h_\mu) \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{K_{\max,\varepsilon} L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}.
$$

By the Kantorovich–Rubinstein theorem,

$$
|T_1| = \left|\int f_\mu \, d(\mu-\nu)\right| \le \mathrm{Lip}(f_\mu)\, \mathcal{W}_1(\mu,\nu) \le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{K_{\max,\varepsilon} L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\right) \mathcal{W}_1(\mu,\nu).
$$

For T_2, first note that

$$
|h_\mu(x) - h_\nu(x)| = \frac{|(K_\varepsilon * (\nu-\mu))(x)|}{(K_\varepsilon * \mu)(x)(K_\varepsilon * \nu)(x)} \le \frac{1}{K_{\min,\varepsilon}^2} \sup_z \left|\int K_\varepsilon(z-w)\, d(\mu-\nu)(w)\right|.
$$

For each fixed z, the function $w \mapsto K_\varepsilon(z-w)$ is L_{K_ε}-Lipschitz. Applying the Kantorovich–Rubinstein theorem gives

$$
\sup_z \left|\int K_\varepsilon(z-w)\, d(\mu-\nu)(w)\right| \le L_{K_\varepsilon}\, \mathcal{W}_1(\mu,\nu).
$$

Hence $\|h_\mu - h_\nu\|_\infty \le \frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\mathcal{W}_1(\mu,\nu)$, and therefore

$$
|T_2| \le \|K_\varepsilon\|_\infty \, \|h_\mu - h_\nu\|_\infty \le \frac{K_{\max,\varepsilon} L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\, \mathcal{W}_1(\mu,\nu).
$$

Combining the bounds for T_1 and T_2 yields

$$
(iii) \le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{2K_{\max,\varepsilon} L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}\right) \mathcal{W}_1(\mu,\nu). \tag{56}
$$

Now let us check item (iv).

$$
\begin{aligned}
(iv) &\le \left|\int \left(\log \frac{K_\varepsilon * \mu(x)}{\pi(x)} - \log \frac{K_\varepsilon * \nu(x)}{\pi(x)}\right) \mu(x) dx\right| \\
&\quad + \left|\int (\mu(x) - \nu(x)) \log \frac{K_\varepsilon * \nu(x)}{\pi(x)} dx\right| \\
&\le \int |\log K_\varepsilon * \mu(x) - \log K_\varepsilon * \nu(x)|\, \mu(x) dx \\
&\quad + \left|\int (\mu(x) - \nu(x)) \log \frac{K_\varepsilon * \nu(x)}{\pi(x)} dx\right| \\
&\le \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right) \mathcal{W}_1(\mu,\nu),
\end{aligned} \tag{57}
$$

where in the last inequality, we used (E.4(2)) and (E.4(5)).

Combining (54), (55), (56) and (57), we could conclude that $a(m,x)$ is Lipschitz with constant

$$
L_\varepsilon^3 = L_F + \left(\frac{4L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{2L_\pi}{\pi_{\min}}\right) + \frac{2K_{\max,\varepsilon} L_{K_\varepsilon}}{K_{\min,\varepsilon}^2}.
$$

□

E.4 For Flow (28)

Define the coefficient associated with the kernelized flow (28) by

$$a(\mu, x) := \frac{\delta F}{\delta \mu}(\mu, x) + \sigma\Big(\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{\pi}\big)(x) - \mathrm{KL}(K_\varepsilon * \mu \mid \pi)\Big).$$

By Proposition E.1, a satisfies both a boundedness estimate and a Lipschitz estimate. In particular, under Assumptions 2, 3, and 4, there exist constants $C_\varepsilon^4 > 0$ and $L_\varepsilon^4 > 0$ such that, for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$,

$$|a(\mu, x)| \le C_\varepsilon^4, \qquad |a(\mu, x) - a(\nu, y)| \le L_\varepsilon^4\big(\mathcal{W}_2(\mu, \nu) + |x - y|\big).$$

Proof. Boundedness. By (E.3(2)) and $\int_{\mathcal{X}} K_\varepsilon = 1$,

$$\begin{aligned}
\Big|\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{\pi}\big)(x)\Big| &= \left|\int K_\varepsilon(x - y)\, \log \tfrac{(K_\varepsilon * \mu)(y)}{\pi(y)}\, dy\right| \\
&\le \int K_\varepsilon(x - y)\, \left|\log \tfrac{(K_\varepsilon * \mu)(y)}{\pi(y)}\right| dy \\
&\le \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}},\ \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}\right).
\end{aligned}$$

By the same reasoning, $\mathrm{KL}(K_\varepsilon * \mu|\pi)$ is bounded by the same constant. Hence we may take

$$C_\varepsilon^4 := C_F + \sigma\, \max\left(\log \frac{\pi_{\max}}{K_{\min,\varepsilon}},\ \log \frac{K_{\max,\varepsilon}}{\pi_{\min}}\right).$$

We now turn to the proof of the Lipschitz condition. By the triangle inequality,

$$|a(\mu, x) - a(\nu, y)| \le (i) + \sigma(ii) + \sigma(iii),$$

where

$$(i) := \left|\tfrac{\delta F}{\delta \mu}(\mu, x) - \tfrac{\delta F}{\delta \mu}(\nu, y)\right|, \quad (ii) := \left|\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{\pi}\big)(x) - \big(K_\varepsilon * \log \tfrac{K_\varepsilon * \nu}{\pi}\big)(y)\right|,$$

$$(iii) := \Big|\mathrm{KL}(K_\varepsilon * \mu|\pi) - \mathrm{KL}(K_\varepsilon * \nu|\pi)\Big|.$$

For (i), Assumption 2 yields

$$(i) \le L_F\big(\mathcal{W}_2(\mu, \nu) + |x - y|\big). \tag{58}$$

For (ii), split into a measure part and a space part:

$$\begin{aligned}
(ii) \ &\le \ \Big|\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \mu}{\pi}\big)(x) - \big(K_\varepsilon * \log \tfrac{K_\varepsilon * \nu}{\pi}\big)(x)\Big| \\
&\quad + \Big|\big(K_\varepsilon * \log \tfrac{K_\varepsilon * \nu}{\pi}\big)(x) - \big(K_\varepsilon * \log \tfrac{K_\varepsilon * \nu}{\pi}\big)(y)\Big| \\
&\le \int K_\varepsilon(x - y) \left|\log \frac{K_\varepsilon * \mu}{\pi}(y) - \log \frac{K_\varepsilon * \nu}{\pi}(y)\right| dy \\
&\quad + \int K_\varepsilon(z) \left|\log \frac{K_\varepsilon * \nu}{\pi}(x - z) - \log \frac{K_\varepsilon * \nu}{\pi}(y - z)\right| dx \\
&\le \left(\frac{L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right)\big(\mathcal{W}_1(\mu, \nu) + |x - y|\big),
\end{aligned} \tag{59}$$

where the last inequality is from (E.4(2)) and (E.4(5)).

For (iii),

$$(iii) \le \left(\frac{2L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{L_\pi}{\pi_{\min}}\right)\mathcal{W}_1(\mu, \nu), \tag{60}$$

because of (E.4(8)).

Combining (58), (59) and (60) we could choose

$$L_\varepsilon^4 = L_F + \sigma\left(\frac{3L_{K_\varepsilon}}{K_{\min,\varepsilon}} + \frac{2L_\pi}{\pi_{\min}}\right).$$

□

E.5 Boundness and Lipschitz Condition for Kernelized Fisher-Rao Gradient Flow Regularized by Chi Square Divergence

Proof. Using the uniform bound on $\frac{\delta F}{\delta \mu}$ and $\pi \geq \pi_{\min}$,

$$\left|K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(x)\right| = \left|\int K_\varepsilon(x-z)\,\frac{(K_\varepsilon * \mu)(z)}{\pi(z)}\,dz\right| \leq \frac{K_{\max,\varepsilon}}{\pi_{\min}},$$

and the same bound holds for $\int K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right) d\mu$. Hence

$$|a(\mu, x)| \leq C_F + 4\sigma\,\frac{K_{\max,\varepsilon}}{\pi_{\min}} =: C_\varepsilon^5.$$

We now analyze the Lipschitz condition. By the triangle inequality,

$$|a(\mu, x) - a(\nu, y)| \leq (i) + 2\sigma(ii) + 2\sigma(iii),$$

where

$$(i) := \left|\frac{\delta F}{\delta \mu}(\mu, x) - \frac{\delta F}{\delta \mu}(\nu, y)\right|,$$

$$(ii) := \left|K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(x) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right)(y)\right|,$$

$$(iii) := \left|\int K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right) d\mu - \int K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right) d\nu\right|.$$

For (i),

$$(i) \leq L_F\big(\mathcal{W}_2(\mu, \nu) + |x - y|\big).$$

For (ii), split

$$\left|K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(x) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right)(x)\right| = \left|\int \phi_x(z)\,d(\mu - \nu)(z)\right|,$$

with

$$\phi_x(z) := \int \frac{K_\varepsilon(x-w)}{\pi(w)}\,K_\varepsilon(z-w)\,dw.$$

Then $\mathrm{Lip}(\phi_x) \leq \frac{L_{K_\varepsilon}}{\pi_{\min}}$, hence by the Kantorovich–Rubinstein theorem,

$$\left|K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(x) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right)(x)\right| \leq \frac{L_{K_\varepsilon}}{\pi_{\min}}\,\mathcal{W}_1(\mu, \nu).$$

Moreover,

$$\left|K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right)(x) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right)(y)\right| \leq \frac{L_{K_\varepsilon}}{\pi_{\min}}\,|x - y|.$$

Thus

$$(ii) \leq \frac{L_{K_\varepsilon}}{\pi_{\min}}\big(\mathcal{W}_1(\mu, \nu) + |x - y|\big).$$

For (iii), set $f_\mu(z) = K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(z)$ and write

$$(iii) \leq \left| \int K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right) d(\mu - \nu) \right| + \int \left| K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}\right) \right| d\nu.$$

Since $\mathrm{Lip}(K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)) \leq \mathrm{Lip}(\frac{K_\varepsilon * \mu}{\pi}) \frac{L_{K_\varepsilon}}{\pi_{\min}}$, the first item is bounded by $\frac{L_{K_\varepsilon}}{\pi_{\min}} \mathcal{W}_1(\mu, \nu)$. Furthermore,

$$\left| K_\varepsilon * \left(\frac{K_\varepsilon * \mu}{\pi}\right)(x) - K_\varepsilon * \left(\frac{K_\varepsilon * \nu}{\pi}(x)\right) \right| \leq \left\| \frac{K_\varepsilon * (\mu - \nu)}{\pi} \right\|_\infty \leq \frac{L_{K_\varepsilon}}{\pi_{\min}} \mathcal{W}_1(\mu, \nu),$$

hence

$$(iii) \leq \frac{2L_{K_\varepsilon}}{\pi_{\min}} \mathcal{W}_1(\mu, \nu).$$

Using $\mathcal{W}_1(\mu, \nu) \leq \mathcal{W}_2(\mu, \nu)$ and collecting the bounds gives

$$|a(\mu, x) - a(\nu, y)| \leq \left(L_F + 6\sigma \frac{L_{K_\varepsilon}}{\pi_{\min}} \right) \big(\mathcal{W}_2(\mu, \nu) + |x - y| \big).$$

Therefore the Lipschitz estimate holds with

$$L_\varepsilon^5 := L_F + 6\sigma \frac{L_{K_\varepsilon}}{\pi_{\min}}.$$

□