
Non-convex entropic mean-field optimization via Best Response flow

Razvan-Andrei Lascu
RIKEN AIP
razvan-andrei.lascu@riken.jp

Mateusz B. Majka
Heriot-Watt University
m.majka@hw.ac.uk

Abstract

We study the problem of minimizing non-convex functionals on the space of probability measures, regularized by the relative entropy (KL divergence) with respect to a fixed reference measure, as well as the corresponding problem of solving entropy-regularized non-convex-non-concave min-max problems. We utilize the Best Response flow (also known in the literature as the fictitious play flow) and study how its convergence is influenced by the relation between the degree of non-convexity of the functional under consideration, the regularization parameter and the tail behaviour of the reference measure. In particular, we demonstrate how to choose the regularizer, given the non-convex functional, so that the Best Response operator becomes a contraction with respect to the L^1 -Wasserstein distance, which ensures the existence of its unique fixed point that is then shown to be the unique global minimizer for our optimization problem. This extends recent results where the Best Response flow was applied to solve convex optimization problems regularized by the relative entropy with respect to arbitrary reference measures, and with arbitrary values of the regularization parameter. Our results explain precisely how the assumption of convexity can be relaxed, at the expense of making a specific choice of the regularizer. Additionally, we demonstrate how these results can be applied in reinforcement learning in the context of policy optimization for Markov Decision Processes and Markov games with softmax parametrized policies in the mean-field regime.

1 Introduction

We consider the problem of minimizing an entropy-regularized, non-convex functional $F : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ over the Wasserstein space $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$, that is,

$$\min_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} F^\sigma(\nu), \text{ with } F^\sigma(\nu) := F(\nu) + \sigma \text{KL}(\nu|\xi), \quad (1)$$

where the function F is bounded below, i.e., $\inf_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} F(\nu) > -\infty$, $\xi \in \mathcal{P}_1(\mathbb{R}^d)$ is a fixed reference probability measure with finite first moment and $\sigma > 0$ is a regularization parameter, while KL and \mathcal{W}_1 denote the KL-divergence (relative entropy) and the L^1 -Wasserstein distance, respectively.

In recent years, there has been considerable interest in such problems, motivated by their applications in machine learning, including the task of training two-layer neural networks (NNs) in the mean-field regime [18, 33, 9, 10, 31, 37, 35] and Reinforcement Learning (RL) [1, 24, 43, 28]. In particular, several works studied problem (1) in a setting where F is assumed to be convex, by utilizing various gradient flows such as the Best Response/fictitious play flow [8, 32], the Wasserstein gradient flow [18, 33], or the Fisher-Rao gradient flow [25].

In these works, the convergence of the flow to the minimizer of (1) holds for any value of $\sigma > 0$ (although the convergence rate may degenerate when σ approaches zero) and for an arbitrary reference

measure ξ , which indicates that for a convex F , one needs to add *any* regularization to ensure convergence. The exact form of the regularizer matters only for the convergence rate, but not for the sheer fact that the convergence holds. In the present paper we focus on the Best Response flow and we make an observation that, in order to ensure its convergence, the assumption about the convexity of F is superfluous. One can demonstrate convergence for a non-convex F , as long as a specific choice of an appropriate regularizer is made. This will be precisely characterized by a formula that links the degree of non-convexity of F , the value of σ and the first moment of the reference measure ξ , see (10) in Section 2 and the discussion therein. Hence we extend the results from [8] to non-convex functions F .

On the other hand, as motivation for this extension, we study applications in RL in the context of policy optimization for Markov Decision Processes (MDPs) with softmax parametrized policies with single-hidden-layer neural network in the mean-field regime [1, 24, 43, 28]. In order to explain the main ideas, let us focus on the simplified setting of a one-state MDP (also known in the literature as the bandit problem [1, 30]; for the general setting, see Section 2.2), where our goal is to minimize the function

$$F(\nu) = \int_A c(a)\pi_\nu(da), \quad (2)$$

for a cost function $c : A \rightarrow \mathbb{R}$, where A is a Polish space denoting the player's action space and the measure π_ν represents the policy. Note that in the context of the bandit problem, one typically considers the goal of maximizing the reward, but here for the sake of notational consistency with [8] and the related optimization literature discussed above, we consider the equivalent problem of minimizing the cost. In the setting of (2), for any measure $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, the mean-field softmax policy $\pi_\nu \in \mathcal{P}^\eta(A)$ is defined by

$$\pi_\nu(da) \propto \exp\left(\int_{\mathbb{R}^d} f(z, a)\nu(dz)\right)\eta(da), \quad (3)$$

where $\mathcal{P}^\eta(A)$ denotes the set of probability measures that are absolutely continuous with respect to the strictly positive finite reference measure η on A . In formula (3), the function

$$\mathcal{P}_1(\mathbb{R}^d) \times A \ni (\nu, a) \mapsto \int_{\mathbb{R}^d} f(z, a)\nu(dz)$$

represents a mean-field neural network with a bounded, continuous, non-constant activation function $f : \mathbb{R}^d \times A \rightarrow \mathbb{R}$, whereas $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ denotes a measure representing the parameters of the network. For a detailed discussion about the mean-field approach to MDPs, see [24]. Note that the normalization constant in (3) renders the objective in (2) non-convex. As a result, standard mean-field optimization results [33, 9, 8, 25], which rely on the convexity of F do not guarantee convergence to a global minimizer.

In this work, we address (1) using the Best Response flow, a learning algorithm originally proposed in [14, 29] for computing Nash equilibria in games with finite-dimensional strategy spaces (see also [16, 17] for the analysis of its convergence in that setting). Only recently has it been introduced as an optimization method for minimizing entropy-regularized convex functionals over the space of probability measures [8], and for computing mixed Nash equilibria (MNEs) in two-player and multi-player zero-sum games with entropy-regularized convex-concave payoff functions [23, 21, 27]. The Best Response flow is defined by

$$d\nu_t = \alpha (\Psi_\sigma[\nu_t] - \nu_t) dt, \quad \nu|_{t=0} := \nu_0 \in \mathcal{P}_1(\mathbb{R}^d), \quad (4)$$

where $\alpha > 0$ is the learning rate of the flow and Ψ_σ is the Best Response operator.

Definition 1.1 (Best Response operator). *For any $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and any $\sigma > 0$, define the Best Response operator $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1^\xi(\mathbb{R}^d)$ by*

$$\Psi_\sigma[\nu](dx) \propto \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x)\right) \xi(dx). \quad (5)$$

Here $\frac{\delta F}{\delta \nu}(\nu, \cdot)$ represents the flat derivative of F at $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ (cf. Definition F.1), whereas $\mathcal{P}_1^\xi(\mathbb{R}^d)$ is the space of probability measures on \mathbb{R}^d that are absolutely continuous with respect to ξ and have finite first moment. The Best Response operator (5) was introduced as part of the definition of the flow (4) in [8], but it appeared also in [33, 9] under the name *proximal operator* as a tool for studying convergence of mean-field Langevin dynamics.

Remark 1.2. We note that there exists a dynamic known as the fictitious play flow (originally proposed in [4] for finite-dimensional games), which is closely related to the flow (4), and which can be interpreted as the best response computed with respect to the historical average of strategies. We refer to [23, Remark 2.6] for a detailed comparison between the two flows in the context of mean-field min-max problems.

Following [8, Definition 3], we observe that for a given $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\sigma > 0$, the map Ψ_σ satisfies the following variational representation:

$$\Psi_\sigma[\nu] = \operatorname{argmin}_{\nu' \in \mathcal{P}_1^\xi(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \frac{\delta F}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) + \sigma \operatorname{KL}(\nu'|\xi) \right\}.$$

This is a direct consequence of the first-order condition characterizing the minimizers of mean-field optimization problems (see [18, Proposition 2.5] and the discussion in [8, Section 3]). Hence Ψ_σ is the unique minimizer (i.e., *Best Response*) of entropy-regularized linearization of F . As Proposition C.1 will demonstrate, every minimizer ν_σ^* of F^σ must satisfy the fixed-point equation $\nu = \Psi_\sigma[\nu]$.

Our results in the MDP setting of Section 2.2 are closely related to [24], which considers the same MDP under the assumption that $\xi \propto e^{-U^\xi}$, where $U^\xi : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function, and utilizes the Wasserstein gradient flow to solve (1). In the present paper we will demonstrate that, compared to the results in [24], one can obtain exponential convergence to the minimizer for the Best Response flow, under weaker assumptions than the ones required in [24] for the Wasserstein flow. The comparison between our results and [24] will be discussed in detail in Section 2.3.

1.1 Non-convex-non-concave min-max problems

We extend our approach from the single-agent optimization setting to the two-player min-max setting by studying an entropy-regularized, non-convex–non-concave, zero-sum game:

$$\min_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} \max_{\mu \in \mathcal{P}_1(\mathbb{R}^d)} F^{\sigma_\nu, \sigma_\mu}(\nu, \mu), \text{ where } F^{\sigma_\nu, \sigma_\mu}(\nu, \mu) := F(\nu, \mu) + \sigma_\nu \operatorname{KL}(\nu|\xi) - \sigma_\mu \operatorname{KL}(\mu|\rho), \quad (6)$$

with a function $F : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$, reference measures $\xi, \rho \in \mathcal{P}_1(\mathbb{R}^d)$ and regularization parameters $\sigma_\nu, \sigma_\mu > 0$. A solution $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)$ to (6) is referred to as a mixed Nash equilibrium (MNE) and can be characterized by the condition

$$\nu_{\sigma_\nu}^* \in \operatorname{argmin}_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} F^{\sigma_\nu, \sigma_\mu}(\nu, \mu_{\sigma_\mu}^*), \quad \mu_{\sigma_\mu}^* \in \operatorname{argmax}_{\mu \in \mathcal{P}_1(\mathbb{R}^d)} F^{\sigma_\nu, \sigma_\mu}(\nu_{\sigma_\nu}^*, \mu).$$

Problem (6) arises in a variety of machine learning contexts, including the training of Generative Adversarial Networks (GANs) [13, 42, 26, 22, 2, 5, 23], adversarial robustness [38] or Multi-Agent Reinforcement Learning (MARL) [44, 12, 7, 21], and has been the subject of extensive theoretical investigation. In particular, several works considered problem (6) under the assumption that F is either bilinear or convex-concave, employing various gradient flow dynamics, including the Best Response/fictitious play flow [23, 21], the Wasserstein gradient flow [26, 42, 13, 2, 5], and the Fisher–Rao gradient flow [22]. Following the single-agent optimization setting presented above, we consider the Best Response flow for (6) and demonstrate that the convexity-concavity of F is not needed for convergence. Specifically, we show that convergence can still be guaranteed in the non-convex-non-concave setting, provided that appropriately chosen regularizers are employed. This relationship is made precise through a formula that connects the degree of non-convexity-non-concavity of F , the regularization parameters σ_ν and σ_μ , the first moments of the reference measures ξ and ρ and, in comparison to the single-agent optimization case, additionally the learning rates of the players' flows α_ν and α_μ . From this standpoint, our results can be interpreted as an extension of the analysis in [23] to the case of non-convex-non-concave objective functions F with distinct regularization parameters and learning rates for each player.

Another key motivation for investigating the non-convex–non-concave min–max problem stems from applications in Multi-Agent Reinforcement Learning (MARL), particularly in the setting of policy optimization for two-player zero-sum Markov games [7, 21]. Specifically, we extend the examples presented in [21], which cover the convex–concave setting of a two-player zero-sum Markov game without policy parametrization, to the more general case of non-convex–non-concave objective functions with softmax parametrized policies, analogous to those encountered in the single-agent

optimization framework presented above. A detailed comparison between our results and those of [21] is provided in Section E.2. In the simplified setting of a two-player bandit problem, we take

$$F(\nu, \mu) = \int_A \int_A c(a, b) \pi_\nu(da) \pi_\mu(db) \quad (7)$$

for a cost function $c : A \times A \rightarrow \mathbb{R}$, where A is the players' action space. In (7), for any measures $\nu, \mu \in \mathcal{P}_1(\mathbb{R}^d)$ the policies $\pi_\nu, \pi_\mu \in \mathcal{P}^\eta(A)$ are given by (3) for π_ν and by an analogous formula for π_μ . Similarly to the single-agent bandit problem, the objective function F is non-convex-non-concave due to the parametrization. As a result, existing mean-field optimization results for computing mixed Nash equilibria (MNEs) in min-max games with convex-concave payoff functions [26, 5, 2, 23, 21, 22] do not, in general, guarantee convergence to a global MNE.

Following the optimization setting, we tackle (6) by utilizing the Best Response flow, which is defined as

$$\begin{cases} d\nu_t = \alpha_\nu (\Psi_{\sigma_\nu}(\nu_t, \mu_t) - \nu_t) dt, \\ d\mu_t = \alpha_\mu (\Phi_{\sigma_\mu}(\nu_t, \mu_t) - \mu_t) dt, \end{cases} \quad (\nu, \mu)|_{t=0} = (\nu_0, \mu_0) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d), \quad (8)$$

where $\alpha_\nu, \alpha_\mu > 0$ are the learning rates of each player and $\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}$ are the Best Response operators defined in Section B analogously to Definition 1.1. As Proposition C.4 will show, any MNE $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)$ of $F^{\sigma_\nu, \sigma_\mu}$ must satisfy the coupled fixed-point equations $\nu = \Psi_{\sigma_\nu}[\nu, \mu]$ and $\mu = \Phi_{\sigma_\mu}[\nu, \mu]$.

As showed in [23, Corollary 1], when $\alpha_\nu = \alpha_\mu = \alpha$, flow (8) converges exponentially at rate $\mathcal{O}(e^{-\alpha t})$ in KL divergence and TV^2 distance to the unique MNE of (6), under the assumption that F is convex-concave. A key observation is that the convergence rate depends solely on the learning rate α and is independent of the regularization parameter $\sigma = \sigma_\nu = \sigma_\mu > 0$. Our results extend those of [23] by removing the convexity-concavity assumption on F , the requirement of uniform bounds on the second-order flat derivatives of F , and allowing for different regularization parameters and learning rates for each player. A detailed discussion about the differences between the single-agent optimization and the min-max settings can be found in Remark 3.5.

1.2 Our contribution

Our results extend those of [8, 24] in the single-agent optimization setting, and [23, 21] in the mix-max setting. We summarize our contribution as follows:

- In the general setting of the problem (1), we extend the analysis of [8] by addressing the case where F is non-convex. In contrast to [8], we show in Theorem 2.3 that when $\sigma > 0$ is sufficiently large, the Best Response map $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d)$ becomes a contraction in the L^1 -Wasserstein metric, which guarantees the existence of a unique global minimizer for (1) (cf. Corollary 2.4). Moreover, in Theorem 2.5 we derive a stability estimate for the flow, with respect to both the initial condition and the regularization parameter σ . As a consequence, in Theorem 2.6 we obtain exponential convergence of the Best Response flow to the unique minimizer in the L^1 -Wasserstein distance — an improvement over [8, Theorem 9], which establishes convergence in the L^1 -Wasserstein metric without an explicit rate. In the context of MDPs, compared to [24], we employ the Best Response flow instead of the Wasserstein gradient flow and achieve a similar convergence rate under less restrictive assumptions on F and U^ξ . Hence our result can be interpreted as providing an alternative method of choosing an optimal policy for MDPs with softmax parametrized policies in the mean field regime. We also briefly discuss the implementation of the Best Response flow in Section 2.4, however, a precise analysis of the corresponding numerical methods still remains an open problem and is left for future work.
- We extend the results of [23] by removing the convexity-concavity assumption on F , the requirement of uniformly bounded second-order flat derivatives and allowing for different regularization parameters σ_ν, σ_μ and learning rates α_ν, α_μ for each player. Specifically, in Theorem B.2, we prove that, when σ_ν, σ_μ are sufficiently large, the joint Best Response operator $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}) : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is a contraction with respect to the L^1 -Wasserstein distance. Consequently, we show that the game (6) admits a unique global MNE (cf. Corollary B.3). Additionally, in Theorem 3.3, we establish that under this strong regularization regime, the Best Response flow (8) converges exponentially in the L^1 -Wasserstein metric to the equilibrium. Lastly, in Subsection E.2, we illustrate

how our theoretical results for min-max games apply to policy optimization in two-player zero-sum Markov games, thereby generalizing the examples in [21, Section 5] from the convex-concave to the non-convex-non-concave setting with softmax parametrized policies.

2 Assumptions and main results for single-agent optimization

In this section, we first introduce the necessary notations and assumptions used throughout the paper, followed by our main results.

2.1 Notation and main results

By $\mathcal{P}_1(\mathbb{R}^d)$ we denote the space of probability measures with finite first moment. We equip $\mathcal{P}_1(\mathbb{R}^d)$ with the L^1 -Wasserstein distance \mathcal{W}_1 . The set $\mathcal{P}_1(\mathbb{R}^d)$ endowed with the topology induced by the \mathcal{W}_1 distance is a complete separable metric space [40, Theorem 6.18]. Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel σ -algebra over \mathbb{R}^d . For any measure ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, the set of absolutely continuous measures in $\mathcal{P}_1(\mathbb{R}^d)$ with respect to ν is denoted by $\mathcal{P}_1^\nu(\mathbb{R}^d)$. In particular, we will use this notation for $\nu = \lambda$, where λ is the d -dimensional Lebesgue measure. The Kantorovich–Rubinstein duality theorem [40, Theorem 5.10] implies that for all $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\mathcal{W}_1(\nu, \nu') = \sup_{\phi \in \text{Lip}_1(\mathbb{R}^d)} \int_{\mathbb{R}^d} \phi(x)(\nu - \nu')(dx), \quad (9)$$

where $\text{Lip}_1(\mathbb{R}^d)$ denotes the space of all functions $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ with Lipschitz constant equal to 1. For $\xi \in \mathcal{P}_1(\mathbb{R}^d)$, the relative entropy $\text{KL}(\cdot|\xi) : \mathcal{P}_1(\mathbb{R}^d) \rightarrow [0, \infty]$ with respect to ξ is given for any $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ by

$$\text{KL}(\nu|\xi) := \begin{cases} \int_{\mathbb{R}^d} \log \frac{d\nu}{d\xi}(x) \nu(dx), & \text{if } \nu \in \mathcal{P}_1^\xi(\mathbb{R}^d), \\ +\infty, & \text{otherwise.} \end{cases}$$

and analogously for $\text{KL}(\cdot|\rho)$.

Assumption 2.1 (Lipschitzness and boundedness of the flat derivative). *Assume that $F \in \mathcal{C}^1$ (cf. Definition F.1) and there exist $C_F, L_F > 0$ such that for all $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$ and all $x, x' \in \mathbb{R}^d$, we have*

$$\left| \frac{\delta F}{\delta \nu}(\nu, x) \right| \leq C_F, \quad \left| \frac{\delta F}{\delta \nu}(\nu', x') - \frac{\delta F}{\delta \nu}(\nu, x) \right| \leq L_F (|x' - x| + \mathcal{W}_1(\nu', \nu)).$$

Assumption 2.2. *Assume that $\xi(dx) := e^{-U^\xi(x)} dx^1$ for a measurable function $U^\xi : \mathbb{R}^d \rightarrow \mathbb{R}$ such that U^ξ is bounded below and has at least linear growth, i.e.,*

$$\text{ess inf}_{x \in \mathbb{R}^d} U^\xi(x) > -\infty, \quad \liminf_{x \rightarrow \infty} \frac{U^\xi(x)}{|x|} > 0.$$

We are ready to show that, under Assumptions 2.1 and 2.2, the Best Response map $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d)$ is \mathcal{W}_1 -Lipschitz for any $\sigma > 0$, and becomes a contraction for sufficiently large σ .

Theorem 2.3 (L^1 -Wasserstein contraction). *Let Assumptions 2.1, 2.2 hold. For any $\sigma > 0$, let*

$$L_{\Psi_\sigma, U^\xi} := \frac{L_F}{\sigma} \exp\left(\frac{2C_F}{\sigma}\right) \left(1 + \exp\left(\frac{2C_F}{\sigma}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0.$$

Then

$$\mathcal{W}_1(\Psi_\sigma[\nu], \Psi_\sigma[\nu']) \leq L_{\Psi_\sigma, U^\xi} \mathcal{W}_1(\nu, \nu'),$$

for any $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$. If

$$\sigma > 2C_F + e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx, \quad (10)$$

then $L_{\Psi_\sigma, U^\xi} \in (0, 1)$, hence $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d)$ is an L^1 -Wasserstein contraction.

¹For simplicity, we assume that $Z_\xi := \int_{\mathbb{R}^d} e^{-U^\xi(x)} dx = 1$ since we adopt the convention that the potential function U^ξ is shifted by $\log Z_\xi$.

Note that (10) is a condition that relates the degree of non-convexity of F , the regularization parameter and the tail behaviour of the reference measure. Indeed, the convexity of F can be expressed in terms of $\frac{\delta F}{\delta \nu}$ as

$$F(\nu') - F(\nu) \geq \int_{\mathbb{R}^d} \frac{\delta F}{\delta \nu}(\nu, x)(\nu' - \nu)(dx),$$

for any $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Hence if F can be represented as $F_1 + F_2$, where F_1 is convex and F_2 is non-convex, then the upper bound on $|\frac{\delta F_2}{\delta \nu}|$ will measure the degree of non-convexity of the problem. Moreover, the quantity $\int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx$ is the first moment of the reference measure ξ . Evidently from (10), making this quantity smaller helps to achieve contractivity of Ψ_σ and this corresponds to choosing ξ with lighter tails (or equivalently, U^ξ with higher growth).

Note that the lower bound on σ in condition (10) does not directly imply that the functional F^σ in (1) is strongly convex, see Remark A.1 for details. Moreover, as shown in Remark A.1, enforcing strong convexity of F^σ would require a lower bound on σ that is independent of the choice of ξ . In contrast, a key insight of our work is that the proof of \mathcal{W}_1 -contractivity of the Best Response operator allows us to identify a link between σ and the reference measure ξ , given by (10), which ensures convergence of the Best Response flow to the global minimizer of (1). This connection enables the use of smaller values of σ when ξ is chosen appropriately.

Having proven that Ψ_σ is an L^1 -Wasserstein contraction in the high-regularization regime, we can conclude that it has a unique fixed point. Combining this with the fact that any minimizer of (1) has to be a fixed point of Ψ_σ , we arrive at the following corollary.

Corollary 2.4 (Existence and uniqueness of the minimizer). *Let Assumptions 2.1, 2.2 hold. If (10) holds, then (1) has a unique global minimizer.*

The proofs of Theorem 2.3 and Corollary 2.4 can be found in Section A. Before presenting the stability estimate for the flow, both with respect to the initial condition and the regularization parameter σ , we note that Proposition C.2 ensures the existence and uniqueness of a solution $(\nu_t)_{t \geq 0} \in C([0, \infty]; \mathcal{P}_1(\mathbb{R}^d))$ to the Best Response flow (4).

Theorem 2.5 (Stability of the flow with respect to σ and ν_0 in \mathcal{W}_1). *Let Assumptions 2.1, 2.2 hold. Let $(\nu_t)_{t \geq 0}, (\nu'_t)_{t \geq 0} \subset \mathcal{P}_1(\mathbb{R}^d)$ be the solutions of (4) with parameters $\sigma, \sigma' > 0$ and initial conditions $\nu_0, \nu'_0 \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$, respectively. Let $\nu_\sigma^*, \nu_{\sigma'}^*$ be the unique minimizers of (1) with parameters σ, σ' , respectively. Let*

$$L_{\Psi, \sigma, \sigma', U^\xi} := \frac{C_F}{\sigma \sigma'} \exp\left(C_F \left(\min\{\sigma, \sigma'\} + \frac{1}{\sigma'}\right)\right) \left(1 + \exp\left(\frac{2C_F}{\sigma}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0.$$

If (10) holds, then for all $t \geq 0$,

$$\mathcal{W}_1(\nu_t, \nu'_t) \leq e^{-\alpha t(1-L_{\Psi, \sigma, \sigma', U^\xi})} \mathcal{W}_1(\nu_0, \nu'_0) + |\sigma' - \sigma| \frac{L_{\Psi, \sigma, \sigma', U^\xi}}{1 - L_{\Psi, \sigma, \sigma', U^\xi}} \left(1 - e^{-\alpha t(1-L_{\Psi, \sigma, \sigma', U^\xi})}\right),$$

$$\mathcal{W}_1(\nu_\sigma^*, \nu_{\sigma'}^*) \leq |\sigma - \sigma'| \frac{L_{\Psi, \sigma, \sigma', U^\xi}}{1 - L_{\Psi, \sigma, \sigma', U^\xi}}.$$

An immediate consequence of Theorem 2.5 is our main convergence result for single-agent optimization.

Theorem 2.6 (Convergence of (4) in the high-regularized regime). *Let Assumptions 2.1, 2.2 hold. Let $(\nu_t)_{t \geq 0} \subset \mathcal{P}_1(\mathbb{R}^d)$ be the solution of (4) with parameter $\sigma > 0$ and initial condition $\nu_0 \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$. If (10) holds, then for all $t \geq 0$,*

$$\mathcal{W}_1(\nu_t, \nu_\sigma^*) \leq e^{-\alpha t(1-L_{\Psi, \sigma, \sigma', U^\xi})} \mathcal{W}_1(\nu_0, \nu_\sigma^*).$$

We next show how our general results can be applied to the MDP setting of [24], followed by a detailed comparison with their work, and conclude Section 2 with a numerical scheme for the Best Response flow (4) in the bandit case.

2.2 Markov Decision Processes

We demonstrate how our results on single-agent optimization can be applied to infinite-horizon discounted MDPs studied in [1, 24]. We refer the reader to [34, 3, 15] for a thorough introduction to MDPs.

We consider an MDP defined by the seven-tuple $\mathcal{M} = (S, A, P, c, \delta, \tau, \eta)$ with Polish spaces S and A representing the state and the action space, respectively, a transition kernel $P : S \times A \rightarrow \mathcal{P}(S)$, a cost function $c \in B_b(S \times A)$, a discount factor $\delta \in [0, 1)$, a regularization parameter $\tau > 0$ and a fixed reference measure $\eta \in \mathcal{M}_+(A)$. For a discussion on possible choices of η , see Remark E.4. Below we follow the kernel notation from [24] (see also Section D), i.e., we denote $P \in \mathcal{P}(S|S \times A)$, which means that $P(\cdot|s, a) \in \mathcal{P}(S)$ for any $(s, a) \in S \times A$, and similarly we consider policies $\pi \in \mathcal{P}(A|S)$, i.e, $\pi(\cdot|s) \in \mathcal{P}(A)$ for any $s \in S$.

While prior works such as [1, 24] focus on maximizing expected rewards, we adopt a cost minimization perspective to remain consistent with the general optimization framework in (1). In an MDP, at each time step, the agent observes a state $s \in S$, and based on this state, it chooses an action $a \sim \pi(\cdot|s)$ according to the policy. Then the environment returns a cost $c(s, a)$ and transitions to a new state $s' \sim P(\cdot|s, a)$ according to the transition kernel.

For a given policy $\pi \in \mathcal{P}^\eta(A|S)$ and $s \in S$, define the τ -entropy regularized expected cumulative cost by

$$V_\tau^\pi(s) := \mathbb{E}_s^\pi \left[\sum_{n=0}^{\infty} \delta^n \left(c(s_n, a_n) + \tau \log \frac{d\pi}{d\eta}(a_n|s_n) \right) \right] \in \mathbb{R},$$

where \mathbb{E}_s^π denotes the expectation over the state-action trajectory $(s_0, a_0, s_1, a_1, \dots)$ generated by policy π and kernel P such that $s_0 := s$, $a_n \sim \pi(\cdot|s_n)$ and $s_{n+1} \sim P(\cdot|s_n, a_n)$, for all $n \geq 0$. For any measure $\gamma \in \mathcal{P}(S)$ and any policy $\pi \in \mathcal{P}(A|S)$, we rigorously define \mathbb{E}_γ^π in Appendix D. For a given initial distribution $\gamma \in \mathcal{P}(S)$, the agent's goal is to solve

$$\min_{\pi \in \mathcal{P}^\eta(A|S)} V_\tau^\pi(\gamma), \text{ with } V_\tau^\pi(\gamma) := \int_S V_\tau^\pi(s) \gamma(ds). \quad (11)$$

Due to τ -strong-convexity of the value function $\pi(\cdot|s) \mapsto V_\tau^\pi(s)$, it follows from [24, Theorem 2.1] (see also [18, Proposition 2.5]) that (11) admits a unique policy $\pi_\tau^* \in \mathcal{P}^\eta(A|S)$ independent of γ such that

$$\pi_\tau^*(da|s) \propto \exp \left(-\frac{1}{\tau} Q_\tau^{\pi_\tau^*}(s, a) \right) \eta(da),$$

where for all $\pi \in \mathcal{P}^\eta(A|S)$, the state-value function is defined by

$$Q_\tau^\pi(s, a) = c(s, a) + \delta \int_S V_\tau^\pi(s') P(ds'|s, a). \quad (12)$$

To prevent the agent from having to search for the optimal policy $\pi_\tau^*(\cdot|s)$ at each state $s \in S$ over the entire space of probability measures $\mathcal{P}^\eta(A)$, the expression of π_τ^* suggests that it suffices to search for the minimizer of (11) among the class of softmax policies $\{\pi_\nu(\cdot|s) | \nu \in \mathcal{P}_1(\mathbb{R}^d)\} \subset \mathcal{P}^\eta(A)$, where for each $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, the kernel $\pi_\nu \in \mathcal{P}^\eta(A|S)$ is defined by

$$\pi_\nu(da|s) \propto \exp(f_\nu(s, a)) \eta(da) \quad (13)$$

and

$$f_\nu(s, a) := \int_{\mathbb{R}^d} f(\theta, s, a) \nu(d\theta)$$

is a mean-field neural network with $f : \mathbb{R}^d \times S \times A \rightarrow \mathbb{R}$ denoting a bounded, differentiable, non-constant activation function and $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ denoting a measure which represents the parameters of the network. Note that various activation functions such as softplus, tanh, sigmoid satisfy these assumptions.

Assuming that the policy $\pi_\nu(\cdot|s)$ is of the softmax form (13) for all states $s \in S$, the agent's objective is now to learn the optimal distribution of the network's parameters which solve the problem

$$\min_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} V_\tau^{\pi_\nu}(\gamma).$$

Due to the normalization constant in (13), the value function $\nu \mapsto V_\tau^{\pi\nu}(\gamma)$ is non-convex (see also [30]). Therefore, one should not expect convergence of gradient flows to a global minimizer unless further entropic regularization at the level of the network's parameters is added, which leads us to consider our initial optimization problem (1) with $F(\nu) := V_\tau^{\pi\nu}(\gamma)$ and an appropriately chosen regulariser. Note that, in contrast to the general MDP problem presented above, in the bandit problem the actions and the cost are state-independent and the general setting of (11)-(13) simplifies to the setting (2)-(3) presented in Section 1.

In Proposition E.3 we will verify that the function $F(\nu) := V_\tau^{\pi\nu}(\gamma)$ satisfies Assumption 2.1 and hence all our main results (Theorem 2.3, Corollary 2.4, Theorem 2.5) and Theorem 2.6 apply to the minimization problems (1) corresponding to energy functions $F^\sigma(\nu) := V_\tau^{\pi\nu}(\gamma) + \sigma \text{KL}(\nu|\xi)$ for any reference measure ξ satisfying Assumption 2.2. In particular, we will provide explicit expressions for the constants C_F and L_F that appear in Assumption 2.1, in terms of the parameters of the MDP under consideration, and we will explain how to choose these parameters to ensure that the crucial condition (10) is satisfied, cf. Remark E.4.

2.3 Comparison to the Wasserstein gradient flow

Our work is closely related to [24], which also considers the MDP presented in Section 2.2, and aims to minimize the function $\nu \mapsto V_\tau^{\pi\nu}(\gamma) + \sigma \text{KL}(\nu|\xi)$ with $\xi \propto e^{-U^\xi}$. In [24, Theorem 2.12], it is shown that the Wasserstein gradient flow $(\nu_t)_{t \geq 0}$ defined by

$$\partial_t \nu_t = \nabla \cdot \left(\left(\nabla \frac{\delta F}{\delta \nu}(\nu_t, \cdot) + \sigma \nabla U^\xi \right) \nu_t \right) + \sigma \Delta \nu_t, \quad \nu|_{t=0} := \nu_0 \in \mathcal{P}_2(\mathbb{R}^d), \quad (14)$$

converges exponentially at rate $\mathcal{O}(e^{-\beta t})$ to the unique global minimizer of (1) in the L^2 -Wasserstein distance, provided that $\sigma > 0$ is sufficiently large relative to the regularity constants of F and U^ξ .

More precisely, the convergence rate $\beta := \sigma\kappa - C_2 - L$ is determined by the strong convexity constant $\kappa > 0$ of U^ξ , i.e.,

$$(\nabla U^\xi(x') - \nabla U^\xi(x)) \cdot (x' - x) \geq \kappa |x' - x|^2,$$

the uniform bound $C_2 > 0$ on the Hessian of the flat derivative of F and the Lipschitz constant $L > 0$ of the map $\nu \mapsto \nabla \frac{\delta F}{\delta \nu}(\nu, \cdot)$ with respect to the L^1 -Wasserstein distance, i.e.,

$$\left| \nabla^2 \frac{\delta F}{\delta \nu}(\nu, \cdot) \right| \leq C_2, \quad \left| \nabla \frac{\delta F}{\delta \nu}(\nu', \cdot) - \nabla \frac{\delta F}{\delta \nu}(\nu, \cdot) \right| \leq L \mathcal{W}_1(\nu', \nu).$$

In principle, choosing either σ or κ sufficiently large relative to C_2 and L ensures that $\beta > 0$. However, additional assumptions on U^ξ required for the well-posedness of the flow (14), such as Lipschitzness and at most linear growth of ∇U^ξ , impose an upper bound on κ (see Assumptions 2.5, 2.8 in [24]). Consequently, the quadratic potential $U^\xi(x) = |x|^2$ becomes essentially the only viable choice.

In contrast, under significantly weaker regularity assumptions on F and U^ξ , we establish exponential convergence of the Best Response flow to the unique global minimizer of (1) in the L^1 -Wasserstein distance, with rate $\mathcal{O}\left(e^{-\alpha(1-L_{\Psi_\sigma, U^\xi})t}\right)$, provided that $\sigma > 0$ is sufficiently large. In particular, taking

$$\sigma > 2C_F + e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx,$$

where $C_F, L_F > 0$ are the constants in Assumption 2.1 ensures that

$$L_{\Psi_\sigma, U^\xi} := \frac{L_F}{\sigma} \exp\left(\frac{2C_F}{\sigma}\right) \left(1 + \exp\left(\frac{2C_F}{\sigma}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx < 1.$$

Furthermore, in comparison to [24], we only require U^ξ to be bounded below and to have at least linear growth (cf. Assumption 2.2). Therefore, selecting a potential U^ξ with sufficiently rapid growth can effectively reduce the lower bound required on σ .

Regarding the possible choice of α , we would like to remark that, while in the continuous-time setting, a larger value of $\alpha > 0$ may accelerate the convergence rate of (4), it was shown in [32, Theorem 7] that convergence of explicit Euler discretization of the flow is only guaranteed as long as the step size $\tau > 0$ satisfies $\alpha\tau \leq 1$, effectively imposing a restriction on the range of α .

2.4 Numerical simulation

We now describe a numerical scheme for the Best Response flow (4) in the bandit setting ($S = \emptyset$ in Section 2.2). The idea is to update the parameter distribution ν of the mean-field neural network in (13) using a two-loop Langevin particle algorithm.

The algorithm takes as input the cost function c , activation function f , potential U^ξ and reference measure η (as in Remark E.4), together with the regularization parameter σ prescribed by (10) (with C_F, L_F from Proposition E.3), the regularization $\tau > 0$, the initial distribution of parameters ν_0 , outer and inner step sizes $h_{\text{out}}, h_{\text{in}}$, horizons T, K , learning rate $\alpha > 0$ and number of parameters N .

In the outer loop, we approximate ν_t by the empirical measure

$$\nu_t^N(\mathrm{d}\theta) = \frac{1}{N} \sum_{i=1}^N \delta_{\theta_t^i},$$

where $(\theta_t^i)_{i=1}^N$ are the network parameters. Since $\Psi_\sigma[\nu_t^N]$ has the exponential form (5), we compute it via an inner-loop Langevin dynamics. For K steps,

$$\theta_{t,k+1}^i = \theta_{t,k}^i - h_{\text{in}} \left(\nabla_\theta \frac{\delta V_\tau^{\pi_\nu}}{\delta \nu}(\nu_t^N, \theta_{t,k}^i) + \sigma \nabla_\theta U^\xi(\theta_{t,k}^i) \right) + \sqrt{2h_{\text{in}}\sigma} \mathcal{N}_{t,k}^i,$$

for $i = 1, \dots, N$, where $(\mathcal{N}_{t,k}^i)_{i=1}^N$ are i.i.d normally distributed random variables and $h_{\text{in}} > 0$ is the step size. Note that $\frac{\delta V_\tau^{\pi_\nu}}{\delta \nu}(\nu, \cdot)$ is explicitly computed in Lemma E.2. For K large enough, we set

$$\Psi_\sigma[\nu_t^N] = \frac{1}{N} \sum_{i=1}^N \delta_{\theta_{t,K}^i}.$$

The Best Response flow (4) can be computed in the outer loop via the Euler step

$$\nu_{t+1}^N = (1 - \alpha h_{\text{out}}) \nu_t^N + \alpha h_{\text{out}} \Psi_\sigma[\nu_t^N],$$

where $h_{\text{out}} > 0$ is the step size. The algorithm outputs ν_T^N , which by Theorem 2.6 converges in \mathcal{W}_1 to ν_σ^* , provided that T, N are sufficiently large, yielding a near-optimal policy $\pi_{\nu_T^N} \approx \pi_\tau^*$. The description of the algorithm is summarized in Algorithm 1 in Appendix E.

The continuous-time analysis omits a key feature of the algorithm. In continuous time, as expressed in condition (10), the parameter σ is determined solely by the cost function c , activation f , potential U^ξ , regularization τ and reference measure η (cf. Proposition E.3). By contrast, in discrete time one must also account for the interaction of σ with the step sizes h_{in} and h_{out} . A rigorous analysis of this discrete-time scheme lies beyond the scope of the present paper and will be studied in future work.

3 Assumptions and main results for min-max problems

We now turn our attention to the class of min-max problems given by (6). We start with the necessary assumptions, and then formulate the main results. Let us first consider the metric space $(\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d), \widetilde{\mathcal{W}}_1)$ equipped with the L^1 -Wasserstein distance

$$\widetilde{\mathcal{W}}_1((\nu, \mu), (\nu', \mu')) := \mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu'),$$

for any $(\nu, \mu), (\nu', \mu') \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$.

Assumption 3.1 (Lipschitzness and boundedness of the flat derivative). *Assume that $F \in C^1$ (cf. Definition F.1) and there exist $C_F, \bar{C}_F, L_F, \bar{L}_F > 0$ such that for all $\nu, \mu, \nu', \mu' \in \mathcal{P}_1(\mathbb{R}^d)$ and all $x, y, x', y' \in \mathbb{R}^d$, we have*

$$\begin{aligned} \left| \frac{\delta F}{\delta \nu}(\nu, \mu, x) \right| &\leq C_F, & \left| \frac{\delta F}{\delta \nu}(\nu', \mu', x') - \frac{\delta F}{\delta \nu}(\nu, \mu, x) \right| &\leq L_F \left(|x' - x| + \widetilde{\mathcal{W}}_1((\nu, \mu), (\nu', \mu')) \right), \\ \left| \frac{\delta F}{\delta \mu}(\nu, \mu, y) \right| &\leq \bar{C}_F, & \left| \frac{\delta F}{\delta \mu}(\nu', \mu', y') - \frac{\delta F}{\delta \mu}(\nu, \mu, y) \right| &\leq \bar{L}_F \left(|y' - y| + \widetilde{\mathcal{W}}_1((\nu, \mu), (\nu', \mu')) \right), \end{aligned}$$

Assumption 3.2. Assume that $\xi(dx) := e^{-U^\xi(x)}dx$ and $\rho(dy) := e^{-U^\rho(y)}dy$ for measurable functions $U^\xi, U^\rho : \mathbb{R}^d \rightarrow \mathbb{R}$ such that U^ξ, U^ρ are bounded below and have at least linear growth, i.e.,

$$\operatorname{ess\,inf}_{x \in \mathbb{R}^d} U^\xi(x) > -\infty, \operatorname{ess\,inf}_{y \in \mathbb{R}^d} U^\rho(y) > -\infty, \liminf_{x \rightarrow \infty} \frac{U^\xi(x)}{|x|} > 0, \liminf_{y \rightarrow \infty} \frac{U^\rho(y)}{|y|} > 0.$$

In Theorem B.2 we show that, under Assumptions 3.1 and 3.2, the pair of Best Response maps $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}) : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is $\widetilde{\mathcal{W}}_1$ -Lipschitz for any $\sigma_\nu, \sigma_\mu > 0$, and for sufficiently large σ_ν, σ_μ (cf. conditions (29)-(30)) becomes a contraction on $(\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d), \widetilde{\mathcal{W}}_1)$.

Having proven that $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})$ is an L^1 -Wasserstein contraction in the high-regularization regime, we can conclude that it has a unique fixed point. Combining this with the fact that any MNE of (6) has to be a fixed point of $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})$, we arrive at uniqueness of the MNE of (6) via Corollary B.3. Proposition C.5 ensures the existence and uniqueness of a solution $(\nu_t, \mu_t)_{t \geq 0} \in C([0, \infty]; \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d))$ to the Best Response flow (8).

Subsequently, we prove in Theorem B.4 a stability estimate for the flow, both with respect to the initial condition and the regularization parameters σ_ν, σ_μ . An immediate consequence of Theorem B.4 is our main convergence result for min-max problems.

Theorem 3.3 (Convergence of (8) in the high-regularized regime). *Let Assumptions 3.1, 3.2 hold. Let $(\nu_t, \mu_t)_{t \geq 0} \subset \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ be the solution of (8) with parameters $\sigma_\nu, \sigma_\mu > 0$ and initial condition $(\nu_0, \mu_0) \in \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$. If (41) and (42) hold, then for all $t \geq 0$,*

$$\widetilde{\mathcal{W}}_1((\nu_t, \mu_t), (\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)) \leq e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U^\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U^\rho})} \widetilde{\mathcal{W}}_1((\nu_0, \mu_0), (\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)).$$

Remark 3.4. Note that, similarly to the case of single-agent optimization, the Wasserstein contractivity of the Best Response flow in Theorem B.2 does not depend on the choice of the learning rates α_ν, α_μ , however, these rates influence the stability estimates and convergence rates of (8) in Theorems B.4 and 3.3. Unlike in the single-agent case, in the min-max setting, in the stability and convergence results we work with learning rate-dependent counterparts of the lower bounds on σ_ν, σ_μ . Note that (41)-(42) imply (29)-(30) since $\alpha_\nu, \alpha_\mu \geq \min\{\alpha_\nu, \alpha_\mu\}$. Moreover, in the case where $\alpha := \alpha_\nu = \alpha_\mu$, (29)-(30) and (41)-(42) coincide.

Remark 3.5. Evidently, our results for the min-max problems (6) are a direct counterpart to our results for the minimization problem (1). However, working with the Best Response flow in the context of min-max problems leads to additional difficulties compared to its application in single-agent optimization.

We note that each player is allowed to have distinct regularization parameters, σ_ν and σ_μ , as well as different learning rates, α_ν and α_μ . The presence of differing learning rates, in particular, plays a significant role in determining suitable regularizers to ensure convergence (cf. Theorems B.4, 3.3), in contrast to the single-agent setting, where the choices of σ and α are independent, and the choice of α is unrestricted in the continuous-time setting, cf. the remark at the end of Section 2.3.

Problem (1) was previously examined in [8] under the assumption that F is convex. It was shown that the Best Response flow (4) exhibits exponential convergence in the value function F^σ at rate $\mathcal{O}(\sigma e^{-\alpha t})$ for any $\sigma > 0$. Subsequently, [23] extended the results of [8] to the min-max setting (6), assuming F is convex-concave and choosing $\sigma := \sigma_\nu = \sigma_\mu$. It was proved that the Best Response flow (8) with $\alpha := \alpha_\nu = \alpha_\mu$ converges exponentially at rate $\mathcal{O}(e^{-\alpha t})$ in both KL divergence and TV^2 distance, and at rate $\mathcal{O}(\sigma e^{-\alpha t})$ in the Nikaidó-Isoda (NI) error (see [23, Subsection 1.2]). However, [23] did not analyze convergence in terms of the Wasserstein distance.

The present work significantly generalizes the findings of [23]. We show that exponential convergence to the MNE can be achieved without assuming convexity-concavity of F , while also allowing for different regularization parameters σ_ν, σ_μ and learning rates α_ν, α_μ for each player. Notably, the convergence result in [23, Theorem 2] fundamentally relies on the fact that the players update their strategies at the same rate, i.e., $\alpha_\nu = \alpha_\mu$, whereas Theorem 3.3 establishes that this condition is not necessary.

4 Acknowledgments and Disclosure of Funding

We thank the anonymous reviewers for their insightful comments that helped to improve the paper. R-AL gratefully acknowledges the support of the EPSRC Centre for Doctoral Training in Mathematical Modelling, Analysis and Computation (MAC-MIGS), funded by the UK Engineering and Physical Sciences Research Council (grant EP/S023291/1), Heriot-Watt University, and the University of Edinburgh. This work was carried out while R-AL was a PhD student at Heriot-Watt University.

References

- [1] A. Agazzi and J. Lu. Global optimality of softmax policy gradient with single hidden layer neural networks in the mean-field regime. In *International Conference on Learning Representations*, 2021.
- [2] J. An and J. Lu. Convergence of two-timescale gradient descent ascent dynamics: finite-dimensional and mean-field perspectives, 2025. arXiv:2501.17122.
- [3] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press, Inc., USA, 1978.
- [4] G. W. Brown. Iterative solution of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*. Wiley, New York, 1951.
- [5] Y. Cai, S. Mitra, X. Wang, and A. Wibisono. On the convergence of min-max Langevin dynamics and algorithm, 2025. arXiv:2412.20471.
- [6] R. A. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I: Mean Field FBSDEs, Control, and Games*. Springer International Publishing, 2018.
- [7] S. Cen, Y. Chi, S. S. Du, and L. Xiao. Faster last-iterate convergence of policy optimization in zero-sum Markov games. In *ICLR*, 2023.
- [8] F. Chen, Z. Ren, and S. Wang. Entropic fictitious play for mean field optimization problem. *Journal of Machine Learning Research*, 24(211):1–36, 2023.
- [9] L. Chizat. Mean-field Langevin dynamics: Exponential convergence and annealing. *Transactions on Machine Learning Research*, 2022.
- [10] L. Chizat and F. R. Bach. On the global convergence of gradient descent for over-parameterized models using optimal transport. In *NeurIPS*, 2018.
- [11] G. Conforti, A. Kazeykina, and Z. Ren. Game on random environment, mean-field Langevin system, and neural networks. *Mathematics of Operations Research*, 48(1):78–99, 2023.
- [12] C. Daskalakis, D. J. Foster, and N. Golowich. Independent policy gradient methods for competitive reinforcement learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20*, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [13] C. Domingo-Enrich, S. Jelassi, A. Mensch, G. Rotskoff, and J. Bruna. A mean-field analysis of two-player zero-sum games. In *Advances in Neural Information Processing Systems*, volume 33, pages 20215–20226. Curran Associates, Inc., 2020.
- [14] I. Gilboa and A. Matsui. Social stability and equilibrium. *Econometrica*, 59(3):859–867, 1991.
- [15] O. Hernandez-Lerma and J. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Stochastic Modelling and Applied Probability. Springer New York, 2012.
- [16] J. Hofbauer. Stability for the best response dynamics. *Unpublished paper*, 1995.
- [17] J. Hofbauer and S. Sorin. Best response dynamics for continuous zero-sum games. *Discrete and Continuous Dynamical Systems - B*, 6(1):215–224, 2006.

- [18] K. Hu, Z. Ren, D. Šiška, and Ł. Szpruch. Mean-field Langevin dynamics and energy landscape of neural networks. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 57(4):2043 – 2065, 2021.
- [19] B. Jourdain and A. Tse. Central limit theorem over non-linear functionals of empirical measures with applications to the mean-field fluctuation of interacting diffusions. *Electronic Journal of Probability*, 26:1 – 34, 2021.
- [20] B. Kerimkulov, J.-M. Leahy, D. Siska, L. Szpruch, and Y. Zhang. A Fisher-Rao gradient flow for entropy-regularised Markov decision processes in Polish spaces, 2024. arXiv:2310.02951.
- [21] J. Kim, K. Yamamoto, K. Oko, Z. Yang, and T. Suzuki. Symmetric mean-field Langevin dynamics for distributional minimax problems. In *The Twelfth International Conference on Learning Representations*, 2024.
- [22] R.-A. Lascu, M. B. Majka, and Ł. Szpruch. A Fisher-Rao gradient flow for entropic mean-field min-max games. *Transactions on Machine Learning Research*, 2024.
- [23] R.-A. Lascu, M. B. Majka, and Ł. Szpruch. Entropic mean-field min–max problems via Best Response flow. *Appl. Math. Optim.*, 91(2), Mar. 2025.
- [24] J.-M. Leahy, B. Kerimkulov, D. Šiška, and Ł. Szpruch. 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, 17–23 Jul 2022.
- [25] L. Liu, M. Majka, and Ł. Szpruch. Polyak–Łojasiewicz inequality on the space of measures and convergence of mean-field birth-death processes. *Applied Mathematics and Optimization*, 87(3), Mar. 2023.
- [26] Y. Lu. Two-scale gradient descent ascent dynamics finds Mixed Nash Equilibria of continuous games: A mean-field perspective. In *Proceedings of the 40th International Conference on Machine Learning*, 2023.
- [27] Y. Lu and P. Monmarché. Convergence of time-averaged mean field gradient descent dynamics for continuous multi-player zero-sum games, 2025. arXiv:2505.07642.
- [28] P. Malo, L. Viitasaari, A. Suominen, E. Vilkkumaa, and O. Tahvonen. Convex regularization and convergence of policy gradient flows under safety constraints, 2024. arXiv:2411.19193.
- [29] A. Matsui. Best response dynamics and socially stable strategies. *Journal of Economic Theory*, 57(2):343–362, 1992.
- [30] J. Mei, C. Xiao, C. Szepesvari, and D. Schuurmans. On the global convergence rates of softmax policy gradient methods. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6820–6829. PMLR, 13–18 Jul 2020.
- [31] S. Mei, A. Montanari, and P.-M. Nguyen. A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, 115:E7665 – E7671, 2018.
- [32] A. Nitanda, K. Oko, D. Wu, N. Takenouchi, and T. Suzuki. Primal and dual analysis of entropic fictitious play for finite-sum problems. In *International Conference on Machine Learning*, 2023.
- [33] A. Nitanda, D. Wu, and T. Suzuki. Convex Analysis of the Mean Field Langevin Dynamics. *arXiv e-prints*, page arXiv:2201.10469, Jan. 2022.
- [34] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., USA, 1st edition, 1994.
- [35] G. Rotskoff and E. Vanden-Eijnden. Trainability and accuracy of artificial neural networks: An interacting particle system approach. *Communications on Pure and Applied Mathematics*, 75(9):1889–1935, 2022.

- [36] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100, 1953.
- [37] J. Sirignano and K. Spiliopoulos. Mean field analysis of neural networks: A central limit theorem. *Stochastic Processes and their Applications*, 130(3):1820–1852, 2020.
- [38] C. G. Trillos and N. G. Trillos. On adversarial robustness and the use of Wasserstein ascent-descent dynamics to enforce it, 2023. arXiv:2301.03662.
- [39] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York, 2008.
- [40] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.
- [41] D. Šiška and Ł. Szpruch. Gradient Flows for Regularized Stochastic Control Problems. *arXiv e-prints*, page arXiv:2006.05956, June 2020.
- [42] G. Wang and L. Chizat. An exponentially converging particle method for the Mixed Nash Equilibrium of continuous games, 2023. arXiv:2211.01280.
- [43] K. Yamamoto, K. Oko, Z. Yang, and T. Suzuki. Mean field Langevin actor-critic: Faster convergence and global optimality beyond lazy learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 55706–55738. PMLR, 21–27 Jul 2024.
- [44] K. Zhang, S. M. Kakade, T. Basar, and L. F. Yang. Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity. *Journal of Machine Learning Research*, 24(175):1–53, 2023.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [\[Yes\]](#)

Justification: See abstract and introduction.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: See Sections 1.2, 2.2, 3.1.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.

- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: See Sections 2, 3 and Appendix.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.

- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper does not include experiments requiring code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: We reviewed the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: This is a theory paper.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: The paper does not use existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

A Proofs of the main results in Section 2

In this section we present the proofs of the results from Section 2. We start with the proof of Theorem 2.3.

Proof of Theorem 2.3. The proof is an adaptation of [8, Proposition 14, Corollary 15] for the L^1 -Wasserstein distance. We give the proof in two steps.

Step 1: First, we show that the the Best Response map $\Psi_\sigma[\nu]$ belongs to $\mathcal{P}_1(\mathbb{R}^d)$, and that $\frac{d\Psi_\sigma[\nu]}{d\xi}(x)$ is uniformly bounded from below and above. From Assumption 2.1, we have

$$\exp\left(-\frac{1}{\sigma}C_F - U^\xi(x)\right) \leq \exp\left(-\frac{1}{\sigma}\frac{\delta F}{\delta\nu}(\nu, x) - U^\xi(x)\right) \leq \exp\left(\frac{1}{\sigma}C_F - U^\xi(x)\right). \quad (15)$$

Integrating over \mathbb{R}^d and using the fact that $\int_{\mathbb{R}^d} e^{-U^\xi(x)} dx = 1$, we obtain

$$\exp\left(-\frac{C_F}{\sigma}\right) \leq Z_\sigma(\nu) := \int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma}\frac{\delta F}{\delta\nu}(\nu, x) - U^\xi(x)\right) dx \leq \exp\left(\frac{C_F}{\sigma}\right). \quad (16)$$

Thus, we obtain

$$k_{\Psi_\sigma} e^{-U^\xi(x)} \leq \Psi_\sigma[\nu](x) \leq K_{\Psi_\sigma} e^{-U^\xi(x)}, \quad (17)$$

with constant $K_{\Psi_\sigma} = \frac{1}{k_{\Psi_\sigma}} = \exp\left(\frac{2C_F}{\sigma}\right) > 1$, where, by an abuse of notation, $\Psi_\sigma[\nu](x)$ denotes the density of $\Psi_\sigma[\nu]$ with respect to λ on \mathbb{R}^d . Moreover, by definition,

$$\int_{\mathbb{R}^d} \Psi_\sigma[\nu](dx) = \int_{\mathbb{R}^d} \Psi_\sigma[\nu](x) dx = \frac{1}{Z_\sigma(\nu)} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma}\frac{\delta F}{\delta\nu}(\nu, x) - U^\xi(x)\right) dx = 1,$$

and using Assumption 2.2,

$$\int_{\mathbb{R}^d} |x| \Psi_\sigma[\nu](dx) = \int_{\mathbb{R}^d} |x| \Psi_\sigma[\nu](x) dx \leq K_{\Psi_\sigma} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx < \infty.$$

Therefore, $\Psi_\sigma[\nu] \in \mathcal{P}_1(\mathbb{R}^d)$.

Step 2: We show that the map $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d)$ is \mathcal{W}_1 -Lipschitz and then that for sufficiently large σ it is actually a \mathcal{W}_1 -contraction. From Assumption 2.1 and the estimate $|e^x - e^y| \leq e^{\max\{x, y\}} |x - y|$, we have

$$\begin{aligned} \left| \exp\left(-\frac{1}{\sigma}\frac{\delta F}{\delta\nu}(\nu, x) - U^\xi(x)\right) - \exp\left(-\frac{1}{\sigma}\frac{\delta F}{\delta\nu}(\nu', x) - U^\xi(x)\right) \right| \\ \leq \frac{L_F}{\sigma} \exp\left(\frac{1}{\sigma}C_F\right) e^{-U^\xi(x)} \mathcal{W}_1(\nu, \nu'). \end{aligned} \quad (18)$$

Integrating the previous inequality with respect to x , we obtain

$$|Z_\sigma(\nu) - Z_\sigma(\nu')| \leq \frac{L_F}{\sigma} \exp\left(\frac{1}{\sigma}C_F\right) \mathcal{W}_1(\nu, \nu'). \quad (19)$$

Therefore, we have that

$$\begin{aligned}
\left| \Psi_\sigma[\nu](x) - \Psi_\sigma[\nu'](x) \right| &= \left| \frac{1}{Z_\sigma(\nu)} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right. \\
&\quad - \frac{1}{Z_\sigma(\nu')} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \\
&\quad + \frac{1}{Z_\sigma(\nu')} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \\
&\quad \left. - \frac{1}{Z_\sigma(\nu')} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu', x) - U^\xi(x)\right) \right| \\
&\leq \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \frac{|Z_\sigma(\nu') - Z_\sigma(\nu)|}{Z_\sigma(\nu)Z_\sigma(\nu')} \\
&\quad + \frac{1}{Z_\sigma(\nu')} \left| \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right. \\
&\quad \left. - \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu', x) - U^\xi(x)\right) \right|.
\end{aligned}$$

Using estimates (15), (16), (18) and (19), we arrive at the Lipschitz property

$$|\Psi_\sigma[\nu](x) - \Psi_\sigma[\nu'](x)| \leq L_{\Psi_\sigma} e^{-U^\xi(x)} \mathcal{W}_1(\nu, \nu'), \quad (20)$$

with

$$L_{\Psi_\sigma} := \frac{L_F}{\sigma} \exp\left(\frac{2C_F}{\sigma}\right) \left(1 + \exp\left(\frac{2C_F}{\sigma}\right)\right) > 0.$$

Now, applying [8, Lemma 16] with $p = 1$ and $\mu(dx) = e^{-U^\xi(x)} dx$ gives

$$\mathcal{W}_1(\Psi_\sigma[\nu], \Psi_\sigma[\nu']) \leq \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx \left\| \frac{\Psi_\sigma[\nu](\cdot)}{e^{-U^\xi(\cdot)}} - \frac{\Psi_\sigma[\nu'](\cdot)}{e^{-U^\xi(\cdot)}} \right\|_{L^\infty(\mathbb{R}^d)}.$$

Hence, using (20) and setting

$$L_{\Psi_\sigma, U^\xi} := L_{\Psi_\sigma} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0$$

gives

$$\mathcal{W}_1(\Psi_\sigma[\nu], \Psi_\sigma[\nu']) \leq L_{\Psi_\sigma, U^\xi} \mathcal{W}_1(\nu, \nu').$$

If

$$\sigma > 2C_F + e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx,$$

then immediately $L_{\Psi_\sigma, U^\xi} \in (0, 1)$, hence $\Psi_\sigma : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d)$ is a \mathcal{W}_1 -contraction. \square

Remark A.1. *The lower bound on σ in Theorem 2.3 does not directly imply that the functional F^σ in (1) is strongly convex with respect to TV^2 , as the following formal computation demonstrates. Note that, for any $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$,*

$$\begin{aligned}
&F^\sigma(\nu') - F^\sigma(\nu) - \int_{\mathbb{R}^d} \frac{\delta F^\sigma}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) \\
&= F(\nu') - F(\nu) - \int_{\mathbb{R}^d} \frac{\delta F}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) + \\
&\quad + \sigma \text{KL}(\nu'|\xi) - \sigma \text{KL}(\nu|\xi) - \sigma \int_{\mathbb{R}^d} \frac{\delta \text{KL}(\cdot|\xi)}{\delta \nu}(\nu, x)(\nu' - \nu)(dx).
\end{aligned}$$

Remark F.2 and Assumption 2.1 give

$$\begin{aligned}
& \left| F(\nu') - F(\nu) - \int_{\mathbb{R}^d} \frac{\delta F}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) \right| \\
&= \left| \int_0^1 \int \left(\frac{\delta F}{\delta \nu}(\nu + \varepsilon(\nu' - \nu), x) - \frac{\delta F}{\delta \nu}(\nu, x) \right) (\nu' - \nu)(dx) d\varepsilon \right| \\
&\leq L_F \int_0^1 \int \mathcal{W}_1(\nu + \varepsilon(\nu' - \nu), \nu) |\nu' - \nu|(dx) d\varepsilon \\
&= 2L_F \text{TV}(\nu', \nu) \int_0^1 \mathcal{W}_1(\nu + \varepsilon(\nu' - \nu), \nu) d\varepsilon \\
&\leq 2L_F \text{TV}(\nu', \nu) \mathcal{W}_1(\nu', \nu) \int_0^1 \varepsilon d\varepsilon \\
&= L_F \text{TV}(\nu', \nu) \mathcal{W}_1(\nu', \nu),
\end{aligned}$$

where the second equality is due to [39, Lemma 2.1] and the second inequality uses the convexity of the Wasserstein distance.

On the other hand, since

$$\frac{\delta \text{KL}(\cdot|\xi)}{\delta \nu}(\nu, x) = \log \frac{d\nu}{d\xi}(x) - \text{KL}(\nu|\xi),$$

we have

$$\text{KL}(\nu'|\xi) - \text{KL}(\nu|\xi) - \int_{\mathbb{R}^d} \frac{\delta \text{KL}(\cdot|\xi)}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) = \text{KL}(\nu'|\nu). \quad (21)$$

Moreover, by Pinsker's inequality, $\text{TV}^2(\nu', \nu) \leq \frac{1}{2} \text{KL}(\nu'|\nu)$, so we obtain

$$\text{KL}(\nu'|\xi) - \text{KL}(\nu|\xi) - \int_{\mathbb{R}^d} \frac{\delta \text{KL}(\cdot|\xi)}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) \geq 2 \text{TV}^2(\nu', \nu). \quad (22)$$

Therefore,

$$F^\sigma(\nu') - F^\sigma(\nu) - \int \frac{\delta F^\sigma}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) \geq -L_F \text{TV}(\nu', \nu) \mathcal{W}_1(\nu', \nu) + 2\sigma \text{TV}^2(\nu', \nu).$$

This shows that while the KL divergence has natural strong convexity in terms of TV^2 (cf. equation (22)), this does not align with the \mathcal{W}_1 -based Lipschitz continuity of F . Therefore, it does not seem possible to directly deduce strong convexity of F^σ in the TV^2 sense simply by choosing σ large enough relative to L_F . This is one of the motivations for proving convergence via contraction of the Best Response operator rather than by attempting to enforce strong convexity of the objective function.

As a further remark, it would be possible to modify Assumption 2.1 to assume Lipschitz continuity with respect to TV rather than \mathcal{W}_1 . Then we would obtain

$$F^\sigma(\nu') - F^\sigma(\nu) - \int \frac{\delta F^\sigma}{\delta \nu}(\nu, x)(\nu' - \nu)(dx) \geq -L_F \text{TV}^2(\nu', \nu) + 2\sigma \text{TV}^2(\nu', \nu),$$

so choosing $\sigma > \frac{1}{2}L_F$ would ensure strong convexity of F^σ relative to TV^2 . Even though existence and uniqueness of a minimizer could now be established, it is not obvious how to prove that $\text{TV}^2(\nu_t, \nu_\sigma^*) \rightarrow 0$ for the Best Response flow $(\nu_t)_{t \geq 0}$. Moreover, this would change the setting of the problem, which would no longer be comparable to [8, 24], which work in the Wasserstein space.

Finally, we emphasize that this argument eliminates any dependence on the choice of the reference measure ξ (cf. equation (21)) and hence any lower bound on σ obtained this way, which would guarantee strong convexity of F^σ with respect to TV^2 , would be independent of the choice of ξ . On the other hand, one of the important conclusions from our paper is that by following the proof via \mathcal{W}_1 -contractivity of the Best Response operator, we can identify the connection between the choice of σ and the reference measure ξ expressed in terms of (10), that guarantees convergence of the BR flow to the global minimizer. This allows one to choose a smaller value of σ if the measure ξ is chosen appropriately.

Proof of Corollary 2.4. Since $\sigma > 2C_F + e(e+1)L_F \int_{\mathbb{R}^d} |x|e^{-U^\xi(x)} dx$, we can apply Banach's fixed point theorem for the contraction Ψ_σ on the complete metric space $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$ and deduce that the fixed point problem $\Psi_\sigma[\nu] = \nu$ admits a unique solution. We also know from Proposition C.1 that any local minimizer $\nu_\sigma^* \in \mathcal{P}_1(\mathbb{R}^d)$ of F^σ is equivalent to λ , and for λ -a.a. $x \in \mathbb{R}^d$,

$$\nu_\sigma^*(dx) = \frac{e^{-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_\sigma^*, x) - U^\xi(x)}}{\int_{\mathbb{R}^d} e^{-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_\sigma^*, x') - U^\xi(x')} dx'} dx.$$

Therefore, $\Psi_\sigma[\nu_\sigma^*] = \nu_\sigma^*$ due to the definition of Ψ_σ . Since $\Psi_\sigma[\nu] = \nu$ admits a unique solution, it follows that F^σ has a unique minimizer, that is ν_σ^* , and hence it is actually the global minimizer. \square

Proof of Theorem 2.5. Step 1: First, we show that the map Ψ_σ is \mathcal{W}_1 -Lipschitz with respect to σ . From Assumption 2.1 and the estimate $|e^x - e^y| \leq e^{\max\{x, y\}} |x - y|$, we have

$$\begin{aligned} \left| \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) - \exp\left(-\frac{1}{\sigma'} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right| \\ \leq C_F \frac{|\sigma - \sigma'|}{\sigma \sigma'} \exp(C_F \min\{\sigma, \sigma'\}) e^{-U^\xi(x)}. \end{aligned} \quad (23)$$

Integrating the previous inequality with respect to x , we obtain

$$|Z_\sigma(\nu) - Z_{\sigma'}(\nu)| \leq C_F \frac{|\sigma - \sigma'|}{\sigma \sigma'} \exp(C_F \min\{\sigma, \sigma'\}). \quad (24)$$

Therefore, we have that

$$\begin{aligned} \left| \Psi_\sigma[\nu](x) - \Psi_{\sigma'}[\nu](x) \right| &= \left| \frac{1}{Z_\sigma(\nu)} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right. \\ &\quad - \frac{1}{Z_{\sigma'}(\nu)} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \\ &\quad + \frac{1}{Z_{\sigma'}(\nu)} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \\ &\quad \left. - \frac{1}{Z_{\sigma'}(\nu)} \exp\left(-\frac{1}{\sigma'} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right| \\ &\leq \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \frac{|Z_{\sigma'}(\nu) - Z_\sigma(\nu)|}{Z_\sigma(\nu) Z_{\sigma'}(\nu)} \\ &\quad + \frac{1}{Z_{\sigma'}(\nu)} \left| \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right. \\ &\quad \left. - \exp\left(-\frac{1}{\sigma'} \frac{\delta F}{\delta \nu}(\nu, x) - U^\xi(x)\right) \right|. \end{aligned}$$

Using estimates (15), (16), (23) and (24), we arrive at

$$|\Psi_\sigma[\nu](x) - \Psi_{\sigma'}[\nu](x)| \leq L_{\Psi, \sigma, \sigma'} |\sigma - \sigma'| e^{-U^\xi(x)}, \quad (25)$$

with

$$L_{\Psi, \sigma, \sigma'} := \frac{C_F}{\sigma \sigma'} \exp\left(C_F \left(\min\{\sigma, \sigma'\} + \frac{1}{\sigma'}\right)\right) \left(1 + \exp\left(\frac{2C_F}{\sigma}\right)\right) > 0.$$

Now, applying [8, Lemma 16] with $p = 1$ and $\mu(dx) = e^{-U^\xi(x)} dx$ gives

$$\mathcal{W}_1(\Psi_\sigma[\nu], \Psi_{\sigma'}[\nu]) \leq \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx \left\| \frac{\Psi_\sigma[\nu](\cdot)}{e^{-U^\xi(\cdot)}} - \frac{\Psi_{\sigma'}[\nu](\cdot)}{e^{-U^\xi(\cdot)}} \right\|_{L^\infty(\mathbb{R}^d)}.$$

Hence, using (25) and setting

$$L_{\Psi, \sigma, \sigma', U^\xi} := L_{\Psi, \sigma, \sigma'} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0$$

gives

$$\mathcal{W}_1(\Psi_\sigma[\nu], \Psi_{\sigma'}[\nu]) \leq L_{\Psi, \sigma, \sigma', U\xi} |\sigma - \sigma'|.$$

Step 2: Now, we prove the stability of the solution to (4) with respect to σ and ν_0 . Since $\nu_0, \nu'_0 \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$, it follows by Theorem C.2 that $(\nu_t)_{t \geq 0}, (\nu'_t)_{t \geq 0} \subset \mathcal{P}_1^\lambda(\mathbb{R}^d)$. Let $\psi \in \text{Lip}_1(\mathbb{R}^d)$. Then we have that

$$\begin{aligned} & \int_{\mathbb{R}^d} \psi(x) (\nu_t(x) - \nu'_t(x)) dx \\ &= \int_{\mathbb{R}^d} \psi(x) \left(e^{-\alpha t} \nu_0(x) + \int_0^t \alpha e^{-\alpha(t-s)} \Psi_\sigma[\nu_s](x) ds \right. \\ & \quad \left. - e^{-\alpha t} \nu'_0(x) - \int_0^t \alpha e^{-\alpha(t-s)} \Psi_{\sigma'}[\nu'_s](x) ds \right) dx \\ &= \int_{\mathbb{R}^d} \psi(x) \left(e^{-\alpha t} (\nu_0(x) - \nu'_0(x)) + \int_0^t \alpha e^{-\alpha(t-s)} (\Psi_\sigma[\nu_s](x) - \Psi_{\sigma'}[\nu'_s](x)) ds \right) dx \\ &= e^{-\alpha t} \int_{\mathbb{R}^d} \psi(x) (\nu_0(x) - \nu'_0(x)) dx + \int_{\mathbb{R}^d} \psi(x) \int_0^t \alpha e^{-\alpha(t-s)} (\Psi_\sigma[\nu_s](x) - \Psi_{\sigma'}[\nu'_s](x)) ds dx \\ &= e^{-\alpha t} \int_{\mathbb{R}^d} \psi(x) (\nu_0(x) - \nu'_0(x)) dx + \int_0^t \alpha e^{-\alpha(t-s)} \int_{\mathbb{R}^d} \psi(x) (\Psi_\sigma[\nu_s](x) - \Psi_{\sigma'}[\nu'_s](x)) dx ds \\ &\leq e^{-\alpha t} \mathcal{W}_1(\nu_0, \nu'_0) + \int_0^t \alpha e^{-\alpha(t-s)} \mathcal{W}_1(\Psi_\sigma[\nu_s], \Psi_{\sigma'}[\nu'_s]) ds. \end{aligned}$$

where in the last equality we used Fubini's theorem, in the first inequality we used the definition of \mathcal{W}_1 . Now, using Theorem 2.3, *Step 1* and the triangle inequality, we have

$$\begin{aligned} \mathcal{W}_1(\Psi_\sigma[\nu_t], \Psi_{\sigma'}[\nu'_t]) &\leq \mathcal{W}_1(\Psi_\sigma[\nu_t], \Psi_\sigma[\nu'_t]) + \mathcal{W}_1(\Psi_\sigma[\nu'_t], \Psi_{\sigma'}[\nu'_t]) \\ &\leq L_{\Psi, \sigma, U\xi} \mathcal{W}_1(\nu_t, \nu'_t) + L_{\Psi, \sigma, \sigma', U\xi} |\sigma - \sigma'|. \end{aligned}$$

Taking supremum over ψ gives

$$\mathcal{W}_1(\nu_t, \nu'_t) \leq e^{-\alpha t} \mathcal{W}_1(\nu_0, \nu'_0) + \int_0^t \alpha e^{-\alpha(t-s)} [L_{\Psi, \sigma, U\xi} \mathcal{W}_1(\nu_s, \nu'_s) + L_{\Psi, \sigma, \sigma', U\xi} |\sigma - \sigma'|] ds.$$

Since $\sigma > 2C_F + e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U\xi(x)} dx$, it follows that $L_{\Psi, \sigma, U\xi} < 1$.

Therefore, applying Gronwall's lemma to the function $t \mapsto e^{\alpha t} \mathcal{W}_1(\nu_t, \nu'_t)$ yields

$$\mathcal{W}_1(\nu_t, \nu'_t) \leq e^{-\alpha t(1-L_{\Psi, \sigma, U\xi})} \mathcal{W}_1(\nu_0, \nu'_0) + |\sigma' - \sigma| \frac{L_{\Psi, \sigma, \sigma', U\xi}}{1-L_{\Psi, \sigma, U\xi}} \left(1 - e^{-\alpha t(1-L_{\Psi, \sigma, U\xi})}\right). \quad (26)$$

Due to the fact that $\nu_\sigma^* = \Psi_\sigma[\nu_\sigma^*]$, for any $\sigma > 0$, it follows that $\nu_\sigma^*, \nu_{\sigma'}^*$ are solutions to (4) with parameters σ, σ' , respectively. Then using (26) yields

$$\mathcal{W}_1(\nu_\sigma^*, \nu_{\sigma'}^*) \leq e^{-\alpha(1-L_{\Psi, \sigma, U\xi})t} \mathcal{W}_1(\nu_\sigma^*, \nu_{\sigma'}^*) + |\sigma' - \sigma| \frac{L_{\Psi, \sigma, \sigma', U\xi}}{1-L_{\Psi, \sigma, U\xi}} \left(1 - e^{-\alpha(1-L_{\Psi, \sigma, U\xi})t}\right).$$

Therefore, since $L_{\Psi, \sigma, U\xi} \in (0, 1)$, we obtain

$$\mathcal{W}_1(\nu_\sigma^*, \nu_{\sigma'}^*) \leq |\sigma - \sigma'| \frac{L_{\Psi, \sigma, \sigma', U\xi}}{1-L_{\Psi, \sigma, U\xi}}.$$

□

B Proofs of the main results in Section 3

In this section we present the proofs of the results from Section 3. Before starting with the proof of Theorem B.2, we define the Best Response operators for game (6).

Definition B.1 (Best Response operators [23]). For any $\nu, \mu \in \mathcal{P}_1(\mathbb{R}^d)$ and any $\sigma_\nu, \sigma_\mu > 0$, define the Best Response operators $\Psi_{\sigma_\nu} : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1^\xi(\mathbb{R}^d)$ and $\Phi_{\sigma_\mu} : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1^\rho(\mathbb{R}^d)$ by

$$\Psi_{\sigma_\nu}[\nu, \mu](dx) := \frac{\exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x)\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x')\right) \xi(dx')} \xi(dx), \quad (27)$$

$$\Phi_{\sigma_\mu}[\nu, \mu](dy) := \frac{\exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu, \mu, y)\right)}{\int_{\mathbb{R}^d} \exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu, \mu, y')\right) \rho(dy')} \rho(dy). \quad (28)$$

Theorem B.2 (L^1 -Wasserstein contraction). Let Assumptions 3.1, 3.2 hold. For any $\sigma_\nu, \sigma_\mu > 0$, let

$$L_{\Psi_{\sigma_\nu}, U^\xi} := \frac{L_F}{\sigma_\nu} \exp\left(\frac{2C_F}{\sigma_\nu}\right) \left(1 + \exp\left(\frac{2C_F}{\sigma_\nu}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0,$$

$$L_{\Phi_{\sigma_\mu}, U^\rho} := \frac{\bar{L}_F}{\sigma_\mu} \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right) \left(1 + \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\rho(x)} dx > 0.$$

Then

$$\widetilde{\mathcal{W}}_1((\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu, \mu], (\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu', \mu']) \leq (L_{\Psi_{\sigma_\nu}, U^\xi} + L_{\Phi_{\sigma_\mu}, U^\rho}) \widetilde{\mathcal{W}}_1((\nu, \mu), (\nu', \mu')),$$

for any $(\nu, \mu), (\nu', \mu') \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$. If

$$\sigma_\nu > 2C_F + 2e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx, \quad (29)$$

and

$$\sigma_\mu > 2\bar{C}_F + 2e(e+1)\bar{L}_F \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy, \quad (30)$$

then $L_{\Psi_{\sigma_\nu}, U^\xi} + L_{\Phi_{\sigma_\mu}, U^\rho} \in (0, 1)$, hence $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}) : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is a L^1 -Wasserstein contraction on $(\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d), \widetilde{\mathcal{W}}_1)$.

Proof. The proof is an adaptation of the proof of Theorem 2.3. As before, we give the proof in two steps.

Step 1: First, we show that the the Best Response maps $\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma_\mu}[\nu, \mu]$ belong to $\mathcal{P}_1(\mathbb{R}^d)$ and that $\frac{d\Psi_{\sigma_\nu}[\nu, \mu]}{d\xi}(x), \frac{d\Phi_{\sigma_\mu}[\nu, \mu]}{d\rho}(y)$ are uniformly bounded from below and above. From Assumption 3.1, we have

$$\exp\left(-\frac{1}{\sigma_\nu} C_F - U^\xi(x)\right) \leq \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \leq \exp\left(\frac{1}{\sigma_\nu} C_F - U^\xi(x)\right). \quad (31)$$

Integrating over \mathbb{R}^d and using the fact that $\int_{\mathbb{R}^d} e^{-U^\xi(x)} dx = 1$, we obtain

$$\exp\left(-\frac{C_F}{\sigma_\nu}\right) \leq Z_{\sigma_\nu}(\nu, \mu) := \int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) dx \leq \exp\left(\frac{C_F}{\sigma_\nu}\right). \quad (32)$$

Thus, we obtain

$$k_{\Psi_{\sigma_\nu}} e^{-U^\xi(x)} \leq \Psi_{\sigma_\nu}[\nu, \mu](x) \leq K_{\Psi_{\sigma_\nu}} e^{-U^\xi(x)}, \quad (33)$$

with constant $K_{\Psi_{\sigma_\nu}} = \frac{1}{k_{\Psi_{\sigma_\nu}}} = \exp\left(\frac{2C_F}{\sigma_\nu}\right) > 1$, where, by an abuse of notation, $\Psi_{\sigma_\nu}[\nu, \mu](x)$ denotes the density of $\Psi_{\sigma_\nu}[\nu, \mu]$ with respect to λ on \mathbb{R}^d . Moreover, by definition,

$$\begin{aligned} \int_{\mathbb{R}^d} \Psi_{\sigma_\nu}[\nu, \mu](dx) &= \int_{\mathbb{R}^d} \Psi_{\sigma_\nu}[\nu, \mu](x) dx \\ &= \frac{1}{Z_{\sigma_\nu}(\nu, \mu)} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) dx = 1, \end{aligned}$$

and using Assumption 3.2,

$$\int_{\mathbb{R}^d} |x| \Psi_{\sigma_\nu}[\nu, \mu](dx) = \int_{\mathbb{R}^d} |x| \Psi_{\sigma_\nu}[\nu, \mu](x) dx \leq K_{\Psi_{\sigma_\nu}} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx < \infty.$$

Therefore, $\Psi_{\sigma_\nu}[\nu, \mu] \in \mathcal{P}_1(\mathbb{R}^d)$. One can argue similarly for $\Phi_{\sigma_\mu}[\nu, \mu]$ and obtain

$$k_{\Phi_{\sigma_\mu}} e^{-U^\rho(y)} \leq \Phi_{\sigma_\mu}[\nu, \mu](y) \leq K_{\Phi_{\sigma_\mu}} e^{-U^\rho(y)}, \quad (34)$$

with constant $K_{\Phi_{\sigma_\mu}} = \frac{1}{k_{\Phi_{\sigma_\mu}}} = \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right) > 1$.

Step 2: We show that the map $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}) : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is $\widetilde{\mathcal{W}}_1$ -Lipschitz and then that for sufficiently large σ_ν, σ_μ it is actually a \mathcal{W}_1 -contraction. From Assumption 3.1 and the estimate $|e^x - e^y| \leq e^{\max\{x, y\}} |x - y|$, we have

$$\begin{aligned} & \left| \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) - \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu', \mu', x) - U^\xi(x)\right) \right| \\ & \leq \frac{L_F}{\sigma_\nu} \exp\left(\frac{1}{\sigma_\nu} C_F\right) e^{-U^\xi(x)} (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')). \end{aligned} \quad (35)$$

Integrating the previous inequality with respect to x , we obtain

$$|Z_{\sigma_\nu}(\nu, \mu) - Z_{\sigma_\nu}(\nu', \mu')| \leq \frac{L_F}{\sigma_\nu} \exp\left(\frac{1}{\sigma_\nu} C_F\right) (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')). \quad (36)$$

Therefore, we have that

$$\begin{aligned} \left| \Psi_{\sigma_\nu}[\nu, \mu](x) - \Psi_{\sigma_\nu}[\nu', \mu'](x) \right| &= \left| \frac{1}{Z_{\sigma_\nu}(\nu, \mu)} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right. \\ & \quad - \frac{1}{Z_{\sigma_\nu}(\nu', \mu')} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \\ & \quad + \frac{1}{Z_{\sigma_\nu}(\nu', \mu')} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \\ & \quad \left. - \frac{1}{Z_{\sigma_\nu}(\nu', \mu')} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu', \mu', x) - U^\xi(x)\right) \right| \\ & \leq \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \frac{|Z_{\sigma_\nu}(\nu', \mu') - Z_{\sigma_\nu}(\nu, \mu)|}{Z_{\sigma_\nu}(\nu, \mu) Z_{\sigma_\nu}(\nu', \mu')} \\ & \quad + \frac{1}{Z_{\sigma_\nu}(\nu', \mu')} \left| \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right. \\ & \quad \left. - \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu', \mu', x) - U^\xi(x)\right) \right|. \end{aligned}$$

Using estimates (31), (32), (35) and (36), we arrive at the Lipschitz property

$$|\Psi_{\sigma_\nu}[\nu, \mu](x) - \Psi_{\sigma_\nu}[\nu', \mu'](x)| \leq L_{\Psi_{\sigma_\nu}} e^{-U^\xi(x)} (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')), \quad (37)$$

with

$$L_{\Psi_{\sigma_\nu}} := \frac{L_F}{\sigma_\nu} \exp\left(\frac{2C_F}{\sigma_\nu}\right) \left(1 + \exp\left(\frac{2C_F}{\sigma_\nu}\right)\right) > 0.$$

Proving that

$$|\Phi_{\sigma_\mu}[\nu, \mu](y) - \Phi_{\sigma_\mu}[\nu', \mu'](y)| \leq L_{\Phi_{\sigma_\mu}} e^{-U^\rho(y)} (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')), \quad (38)$$

follows the same steps as above but with

$$L_{\Phi_{\sigma_\mu}} := \frac{\bar{L}_F}{\sigma_\mu} \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right) \left(1 + \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right)\right).$$

Now, applying [8, Lemma 16] with $p = 1$ and $\mu(dx) = e^{-U^\xi(x)} dx$ gives

$$\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma_\nu}[\nu', \mu']) \leq \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx \left\| \frac{\Psi_{\sigma_\nu}[\nu, \mu](\cdot)}{e^{-U^\xi(\cdot)}} - \frac{\Psi_{\sigma_\nu}[\nu', \mu'](\cdot)}{e^{-U^\xi(\cdot)}} \right\|_{L^\infty(\mathbb{R}^d)}.$$

Hence, using (37) and setting

$$L_{\Psi_{\sigma_\nu}, U^\xi} := L_{\Psi_{\sigma_\nu}} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0$$

gives

$$\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma_\nu}[\nu', \mu']) \leq L_{\Psi_{\sigma_\nu}, U^\xi} (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')). \quad (39)$$

One similarly obtains that Φ_{σ_μ} is \mathcal{W}_1 -Lipschitz, i.e.,

$$\mathcal{W}_1(\Phi_{\sigma_\mu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu', \mu']) \leq L_{\Phi_{\sigma_\mu}, U^\rho} (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')), \quad (40)$$

with constant

$$L_{\Phi_{\sigma_\mu}, U^\rho} := L_{\Phi_{\sigma_\mu}} \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy > 0.$$

Therefore, we obtain

$$\begin{aligned} & \mathcal{W}_1(\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma_\nu}[\nu', \mu']) + \mathcal{W}_1(\Phi_{\sigma_\mu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu', \mu']) \\ & \leq \left(L_{\Psi_{\sigma_\nu}, U^\xi} + L_{\Phi_{\sigma_\mu}, U^\rho} \right) (\mathcal{W}_1(\nu, \nu') + \mathcal{W}_1(\mu, \mu')), \end{aligned}$$

and using the definition of $\widetilde{\mathcal{W}}_1$ and the notation $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu, \mu] := (\Psi_{\sigma_\nu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu, \mu])$ gives

$$\widetilde{\mathcal{W}}_1((\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu, \mu], (\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu', \mu']) \leq \left(L_{\Psi_{\sigma_\nu}, U^\xi} + L_{\Phi_{\sigma_\mu}, U^\rho} \right) \widetilde{\mathcal{W}}_1((\nu, \mu), (\nu', \mu')).$$

If $\sigma_\nu > 2C_F + 2e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx$ and $\sigma_\mu > 2\bar{C}_F + 2e(e+1)\bar{L}_F \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy$, then immediately $L_{\Psi_{\sigma_\nu}, U^\xi} + L_{\Phi_{\sigma_\mu}, U^\rho} \in (0, 1)$, hence $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}) : \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is a $\widetilde{\mathcal{W}}_1$ -contraction. \square

Corollary B.3 (Existence and uniqueness of the MNE). *Let Assumptions 3.1, 3.2 hold. If (29) and (30) hold, then (6) has a unique global MNE.*

Proof. Since $\sigma_\nu > 2C_F + 2e(e+1)L_F \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx$ and $\sigma_\mu > 2\bar{C}_F + 2e(e+1)\bar{L}_F \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy$, we can apply Banach's fixed point theorem for the contraction $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})$ on the complete metric space $(\mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d), \widetilde{\mathcal{W}}_1)$ and deduce that the fixed point problem $(\Psi_{\sigma_\nu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu, \mu]) = (\nu, \mu)$ admits a unique solution. We also know from Proposition C.4 that any MNE $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ of $F^{\sigma_\nu, \sigma_\mu}$ is equivalent to λ , and for λ -a.a. $x, y \in \mathbb{R}^d$,

$$\begin{aligned} \nu_{\sigma_\nu}^*(dx) &= \frac{\exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, x) - U^\xi(x)\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, x') - U^\xi(x')\right) dx'} dx. \\ \mu_{\sigma_\mu}^*(dy) &= \frac{\exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, y) - U^\rho(y)\right)}{\int_{\mathbb{R}^d} \exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, y') - U^\rho(y')\right) dy'} dy. \end{aligned}$$

Therefore, $(\Psi_{\sigma_\nu}[\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*], \Phi_{\sigma_\mu}[\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*]) = (\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)$ due to the definition of $\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}$. Since $(\Psi_{\sigma_\nu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu, \mu]) = (\nu, \mu)$ admits a unique solution, it follows that $F^{\sigma_\nu, \sigma_\mu}$ has a unique MNE, that is $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)$, and hence it is actually the global MNE. \square

Theorem B.4 (Stability of the flow with respect to σ_ν, σ_μ and (ν_0, μ_0) in $\widetilde{\mathcal{W}}_1$). *Let Assumptions 3.1, 3.2 hold. Let $(\nu_t, \mu_t)_{t \geq 0}, (\nu'_t, \mu'_t)_{t \geq 0} \subset \mathcal{P}_1(\mathbb{R}^d)$ be the solutions of (8) with parameters $(\sigma_\nu, \sigma_\mu), (\sigma'_\nu, \sigma'_\mu)$ and initial conditions $(\nu_0, \mu_0), (\nu'_0, \mu'_0) \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$, respectively. Let $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*)$ and $(\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)$ be the unique MNE of (6) with parameters (σ_ν, σ_μ) and $(\sigma'_\nu, \sigma'_\mu)$, respectively. If*

$$\sigma_\nu > 2C_F + 2e(e+1)L_F \frac{\alpha_\nu}{\min\{\alpha_\nu, \alpha_\mu\}} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx, \quad (41)$$

and

$$\sigma_\mu > 2\bar{C}_F + 2e(e+1)\bar{L}_F \frac{\alpha_\mu}{\min\{\alpha_\nu, \alpha_\mu\}} \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy, \quad (42)$$

there exists

$$\begin{aligned} & L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} \\ & := \frac{C_F}{\sigma_\nu \sigma'_\nu} \exp\left(C_F \left(\min\{\sigma_\nu, \sigma'_\nu\} + \frac{1}{\sigma'_\nu}\right)\right) \left(1 + \exp\left(\frac{2C_F}{\sigma_\nu}\right)\right) \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0, \end{aligned}$$

$$\begin{aligned} & L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} \\ & := \frac{\bar{C}_F}{\sigma_\mu \sigma'_\mu} \exp\left(\bar{C}_F \left(\min\{\sigma_\mu, \sigma'_\mu\} + \frac{1}{\sigma'_\mu}\right)\right) \left(1 + \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right)\right) \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy > 0, \end{aligned}$$

such that for all $t \geq 0$,

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_t, \mu_t), (\nu'_t, \mu'_t)) & \leq e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho}))} \widetilde{\mathcal{W}}_1((\nu_0, \mu_0), (\nu'_0, \mu'_0)) \\ & \quad + \frac{\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma'_\nu - \sigma_\nu| + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} |\sigma'_\mu - \sigma_\mu|}{\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho})} \times \\ & \quad \times \left(1 - e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho}))}\right), \end{aligned}$$

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)) & \leq \frac{\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma'_\nu - \sigma_\nu| + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} |\sigma'_\mu - \sigma_\mu|}{\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho})} \times \\ & \quad \times \left(1 - e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} + \alpha_\mu L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho}))}\right). \end{aligned}$$

Proof. Step 1: First, we show that the maps $\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}$ are \mathcal{W}_1 -Lipschitz with respect to σ_ν, σ_μ , respectively. From Assumption 3.1 and the estimate $|e^x - e^y| \leq e^{\max\{x, y\}} |x - y|$, we have

$$\begin{aligned} & \left| \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) - \exp\left(-\frac{1}{\sigma'_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right| \\ & \leq C_F \frac{|\sigma_\nu - \sigma'_\nu|}{\sigma_\nu \sigma'_\nu} \exp(C_F \min\{\sigma_\nu, \sigma'_\nu\}) e^{-U^\xi(x)}. \quad (43) \end{aligned}$$

Integrating the previous inequality with respect to x , we obtain

$$|Z_{\sigma_\nu}(\nu, \mu) - Z_{\sigma'_\nu}(\nu, \mu)| \leq C_F \frac{|\sigma_\nu - \sigma'_\nu|}{\sigma_\nu \sigma'_\nu} \exp(C_F \min\{\sigma_\nu, \sigma'_\nu\}). \quad (44)$$

Therefore, we have that

$$\begin{aligned}
\left| \Psi_{\sigma_\nu}[\nu, \mu](x) - \Psi_{\sigma'_\nu}[\nu, \mu](x) \right| &= \left| \frac{1}{Z_{\sigma_\nu}(\nu, \mu)} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right. \\
&\quad - \frac{1}{Z_{\sigma'_\nu}(\nu, \mu)} \exp\left(-\frac{1}{\sigma'_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \\
&\quad + \frac{1}{Z_{\sigma'_\nu}(\nu, \mu)} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \\
&\quad \left. - \frac{1}{Z_{\sigma_\nu}(\nu, \mu)} \exp\left(-\frac{1}{\sigma'_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right| \\
&\leq \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \frac{|Z_{\sigma'_\nu}(\nu, \mu) - Z_{\sigma_\nu}(\nu, \mu)|}{Z_{\sigma_\nu}(\nu, \mu) Z_{\sigma'_\nu}(\nu, \mu)} \\
&\quad + \frac{1}{Z_{\sigma'_\nu}(\nu, \mu)} \left| \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right. \\
&\quad \left. - \exp\left(-\frac{1}{\sigma'_\nu} \frac{\delta F}{\delta \nu}(\nu, \mu, x) - U^\xi(x)\right) \right|.
\end{aligned}$$

Using estimates (31), (32), (43) and (44), we arrive at

$$\left| \Psi_{\sigma_\nu}[\nu, \mu](x) - \Psi_{\sigma'_\nu}[\nu, \mu](x) \right| \leq L_{\Psi, \sigma_\nu, \sigma'_\nu} |\sigma_\nu - \sigma'_\nu| e^{-U^\xi(x)}, \quad (45)$$

with

$$L_{\Psi, \sigma_\nu, \sigma'_\nu} := \frac{C_F}{\sigma_\nu \sigma'_\nu} \exp\left(C_F \left(\min\{\sigma_\nu, \sigma'_\nu\} + \frac{1}{\sigma'_\nu}\right)\right) \left(1 + \exp\left(\frac{2C_F}{\sigma_\nu}\right)\right) > 0.$$

Proving

$$\left| \Phi_{\sigma_\mu}[\nu, \mu](y) - \Phi_{\sigma'_\mu}[\nu, \mu](y) \right| \leq L_{\Phi, \sigma_\mu, \sigma'_\mu} |\sigma_\mu - \sigma'_\mu| e^{-U^\rho(y)} \quad (46)$$

follows the same steps as above but with

$$L_{\Phi, \sigma_\mu, \sigma'_\mu} := \frac{\bar{C}_F}{\sigma_\mu \sigma'_\mu} \exp\left(\bar{C}_F \left(\min\{\sigma_\mu, \sigma'_\mu\} + \frac{1}{\sigma'_\mu}\right)\right) \left(1 + \exp\left(\frac{2\bar{C}_F}{\sigma_\mu}\right)\right).$$

Now, applying [8, Lemma 16] with $p = 1$ and $\mu(dx) = e^{-U^\xi(x)} dx$ gives

$$\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma'_\nu}[\nu, \mu]) \leq \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx \left\| \frac{\Psi_{\sigma_\nu}[\nu, \mu](\cdot)}{e^{-U^\xi(\cdot)}} - \frac{\Psi_{\sigma'_\nu}[\nu, \mu](\cdot)}{e^{-U^\xi(\cdot)}} \right\|_{L^\infty(\mathbb{R}^d)}.$$

Hence, using (45) and setting

$$L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} := L_{\Psi, \sigma_\nu, \sigma'_\nu} \int_{\mathbb{R}^d} |x| e^{-U^\xi(x)} dx > 0$$

gives

$$\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu, \mu], \Psi_{\sigma'_\nu}[\nu, \mu]) \leq L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma_\nu - \sigma'_\nu|.$$

We similarly obtain

$$\mathcal{W}_1(\Phi_{\sigma_\mu}[\nu, \mu], \Phi_{\sigma'_\mu}[\nu, \mu]) \leq L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} |\sigma_\mu - \sigma'_\mu|$$

with constant

$$L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} := L_{\Phi, \sigma_\mu, \sigma'_\mu} \int_{\mathbb{R}^d} |y| e^{-U^\rho(y)} dy > 0.$$

Step 2: Now, we prove the stability of the solution to (8) with respect to σ_ν, σ_μ and (ν_0, μ_0) . Since $(\nu_0, \mu_0), (\nu'_0, \mu'_0) \in \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$, it follows by Theorem C.5 that $(\nu_t, \mu_t)_{t \geq 0}, (\nu'_t, \mu'_t)_{t \geq 0} \subset \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$. Let $\psi \in \text{Lip}_1(\mathbb{R}^d)$. Then we have that

$$\begin{aligned}
& \int_{\mathbb{R}^d} \psi(x) (\nu_t(x) - \nu'_t(x)) dx \\
&= \int_{\mathbb{R}^d} \psi(x) \left(e^{-\alpha_\nu t} \nu_0(x) + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu}[\nu_s, \mu_s](x) ds \right. \\
&\quad \left. - e^{-\alpha_\nu t} \nu'_0(x) - \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s](x) ds \right) dx \\
&= \int_{\mathbb{R}^d} \psi(x) \left(e^{-\alpha_\nu t} (\nu_0(x) - \nu'_0(x)) \right. \\
&\quad \left. + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} (\Psi_{\sigma_\nu}[\nu_s, \mu_s](x) - \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s](x)) ds \right) dx \\
&= e^{-\alpha_\nu t} \int_{\mathbb{R}^d} \psi(x) (\nu_0(x) - \nu'_0(x)) dx \\
&\quad + \int_{\mathbb{R}^d} \psi(x) \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} (\Psi_{\sigma_\nu}[\nu_s, \mu_s](x) - \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s](x)) ds dx \\
&= e^{-\alpha_\nu t} \int_{\mathbb{R}^d} \psi(x) (\nu_0(x) - \nu'_0(x)) dx \\
&\quad + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \int_{\mathbb{R}^d} \psi(x) (\Psi_{\sigma_\nu}[\nu_s, \mu_s](x) - \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s](x)) dx ds \\
&\leq e^{-\alpha_\nu t} \mathcal{W}_1(\nu_0, \nu'_0) + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \mathcal{W}_1(\Psi_{\sigma_\nu}[\nu_s, \mu_s], \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s]) ds \\
&\leq e^{-\alpha_\nu t} \mathcal{W}_1(\nu_0, \nu'_0) + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \left(\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu_s, \mu_s], \Psi_{\sigma_\nu}[\nu'_s, \mu'_s]) \right. \\
&\quad \left. + \mathcal{W}_1(\Psi_{\sigma_\nu}[\nu'_s, \mu'_s], \Psi_{\sigma'_\nu}[\nu'_s, \mu'_s]) \right) ds \\
&\leq e^{-\alpha_\nu t} \mathcal{W}_1(\nu_0, \nu'_0) \\
&\quad + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} (\mathcal{W}_1(\Psi_{\sigma_\nu}[\nu_s, \mu_s], \Psi_{\sigma_\nu}[\nu'_s, \mu'_s]) + L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma_\nu - \sigma'_\nu|) ds \\
&\leq e^{-\alpha_\nu t} \mathcal{W}_1(\nu_0, \nu'_0) \\
&\quad + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} (L_{\Psi, \sigma_\nu, U^\xi} (\mathcal{W}_1(\nu_s, \nu'_s) + \mathcal{W}_1(\mu_s, \mu'_s)) + L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma_\nu - \sigma'_\nu|) ds,
\end{aligned}$$

where in the last equality we used Fubini's theorem, in the first inequality we used the definition of \mathcal{W}_1 and the last three inequalities follow from the triangle inequality, *Step 1* and (39), respectively. Taking supremum over ψ gives

$$\begin{aligned}
\mathcal{W}_1(\nu_t, \nu'_t) &\leq e^{-\alpha_\nu t} \mathcal{W}_1(\nu_0, \nu'_0) \\
&\quad + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} (L_{\Psi, \sigma_\nu, U^\xi} (\mathcal{W}_1(\nu_s, \nu'_s) + \mathcal{W}_1(\mu_s, \mu'_s)) + L_{\Psi, \sigma_\nu, \sigma'_\nu, U^\xi} |\sigma_\nu - \sigma'_\nu|) ds.
\end{aligned}$$

Similarly using (40), we obtain

$$\begin{aligned}
\mathcal{W}_1(\mu_t, \mu'_t) &\leq e^{-\alpha_\mu t} \mathcal{W}_1(\mu_0, \mu'_0) \\
&\quad + \int_0^t \alpha_\mu e^{-\alpha_\mu(t-s)} \left(L_{\Phi, \sigma_\mu, U^\rho} (\mathcal{W}_1(\nu_s, \nu'_s) + \mathcal{W}_1(\mu_s, \mu'_s)) + L_{\Phi, \sigma_\mu, \sigma'_\mu, U^\rho} |\sigma_\mu - \sigma'_\mu| \right) ds.
\end{aligned}$$

Using the bound $\alpha_\nu, \alpha_\mu \geq \min\{\alpha_\nu, \alpha_\mu\}$ we can add the previous two inequalities, and using the definition of $\widetilde{\mathcal{W}}_1$ and the notation $(\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu})[\nu, \mu] := (\Psi_{\sigma_\nu}[\nu, \mu], \Phi_{\sigma_\mu}[\nu, \mu])$ we obtain

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_t, \mu_t), (\nu'_t, \mu'_t)) &\leq e^{-\min\{\alpha_\nu, \alpha_\mu\}t} \widetilde{\mathcal{W}}_1((\nu_0, \mu_0), (\nu'_0, \mu'_0)) \\ &+ \int_0^t e^{-\min\{\alpha_\nu, \alpha_\mu\}(t-s)} \left(\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho} \right) \widetilde{\mathcal{W}}_1((\nu_s, \mu_s), (\nu'_s, \mu'_s)) ds \\ &+ \int_0^t e^{-\min\{\alpha_\nu, \alpha_\mu\}(t-s)} \left(\alpha_\nu L_{\Psi_{\sigma_\nu}, \sigma'_\nu, U\xi} |\sigma_\nu - \sigma'_\nu| + \alpha_\mu L_{\Phi_{\sigma_\mu}, \sigma'_\mu, U\rho} |\sigma_\mu - \sigma'_\mu| \right) ds. \end{aligned}$$

Since $\sigma_\nu > 2C_F + 2e(e+1)L_F \frac{\alpha_\nu}{\min\{\alpha_\nu, \alpha_\mu\}} \int_{\mathbb{R}^d} |x| e^{-U\xi(x)} dx$ and $\sigma_\mu > 2\bar{C}_F + 2e(e+1)\bar{L}_F \frac{\alpha_\mu}{\min\{\alpha_\nu, \alpha_\mu\}} \int_{\mathbb{R}^d} |y| e^{-U\rho(y)} dy$, it follows that $\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho} < \min\{\alpha_\nu, \alpha_\mu\}$.

Therefore, applying Gronwall's lemma to the function $t \mapsto e^{\min\{\alpha_\nu, \alpha_\mu\}t} \widetilde{\mathcal{W}}_1((\nu_t, \mu_t), (\nu'_t, \mu'_t))$ yields

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_t, \mu_t), (\nu'_t, \mu'_t)) &\leq e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}))} \widetilde{\mathcal{W}}_1((\nu_0, \mu_0), (\nu'_0, \mu'_0)) \\ &+ \frac{\alpha_\nu L_{\Psi_{\sigma_\nu}, \sigma'_\nu, U\xi} |\sigma_\nu - \sigma'_\nu| + \alpha_\mu L_{\Phi_{\sigma_\mu}, \sigma'_\mu, U\rho} |\sigma_\mu - \sigma'_\mu|}{\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho})} \times \\ &\times \left(1 - e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}))} \right). \end{aligned} \quad (47)$$

Since $\alpha_\nu, \alpha_\mu \geq \min\{\alpha_\nu, \alpha_\mu\}$, it follows that the lower bounds for σ_ν, σ_μ in Corollary B.3 hold, hence $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)$ are the MNE of (6) with parameters $(\sigma_\nu, \sigma_\mu), (\sigma'_\nu, \sigma'_\mu)$, respectively. Moreover, since $\nu_{\sigma_\nu}^* = \Psi_{\sigma_\nu}[\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*], \mu_{\sigma_\mu}^* = \Phi_{\sigma_\mu}[\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*]$, we deduce that $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)$ are solutions to (8) with parameters $(\sigma_\nu, \sigma_\mu), (\sigma'_\nu, \sigma'_\mu)$, respectively. Then using (47) yields

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)) &\leq e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}))} \widetilde{\mathcal{W}}_1((\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)) \\ &+ \frac{\alpha_\nu L_{\Psi_{\sigma_\nu}, \sigma'_\nu, U\xi} |\sigma_\nu - \sigma'_\nu| + \alpha_\mu L_{\Phi_{\sigma_\mu}, \sigma'_\mu, U\rho} |\sigma_\mu - \sigma'_\mu|}{\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho})} \times \\ &\times \left(1 - e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}))} \right). \end{aligned}$$

Therefore, since $\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}) > 0$, we obtain

$$\begin{aligned} \widetilde{\mathcal{W}}_1((\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*), (\nu_{\sigma'_\nu}^*, \mu_{\sigma'_\mu}^*)) &\leq \frac{\alpha_\nu L_{\Psi_{\sigma_\nu}, \sigma'_\nu, U\xi} |\sigma_\nu - \sigma'_\nu| + \alpha_\mu L_{\Phi_{\sigma_\mu}, \sigma'_\mu, U\rho} |\sigma_\mu - \sigma'_\mu|}{\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho})} \times \\ &\times \left(1 - e^{-t(\min\{\alpha_\nu, \alpha_\mu\} - (\alpha_\nu L_{\Psi_{\sigma_\nu}, U\xi} + \alpha_\mu L_{\Phi_{\sigma_\mu}, U\rho}))} \right). \end{aligned}$$

□

C Auxiliary results for the Best Response flow

C.1 Single-agent optimization

We first recall a necessary condition for the optimality of local minimizers of (1).

Proposition C.1 (Necessary condition for optimality [18, Proposition 2.5]). *Let Assumptions 2.1, 2.2 hold. If $\nu_\sigma^* \in \mathcal{P}_1(\mathbb{R}^d)$ is a minimizer of F^σ , then ν_σ^* is equivalent to λ , and for λ -a.a. $x \in \mathbb{R}^d$,*

$$\nu_\sigma^*(dx) = \frac{e^{-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_\sigma^*, x) - U^\xi(x)}}{\int_{\mathbb{R}^d} e^{-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_\sigma^*, x') - U^\xi(x')} dx'} dx.$$

Next we give the proof of the existence and uniqueness of the flow (4). While the proof is skipped in [8, Proposition 7], we present it by following a classical Picard iteration technique.

Proposition C.2 (Existence and uniqueness of the flow). *Let Assumptions 2.1, 2.2 hold and let $\nu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. Then there exists a unique solution $(\nu_t)_{t \geq 0} \in C([0, \infty]; \mathcal{P}_1(\mathbb{R}^d))$ to (4) and the solution depends continuously on the initial condition. Moreover, if $\nu_0 \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$, then $(\nu_t)_{t \geq 0}$ admits the density $(\nu_t(\cdot))_{t \geq 0} \subset \mathcal{P}_1^\lambda(\mathbb{R}^d)$ such that, for every $x \in \mathbb{R}^d$, it holds that $t \mapsto \nu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$, and*

$$d\nu_t(x) = \alpha (\Psi_\sigma[\nu_t](x) - \nu_t(x)) dt, \quad t > 0, \quad (48)$$

for some initial condition $\nu_0(x) \in \mathcal{P}_1^\lambda(\mathbb{R}^d)$.

Proof of Theorem C.2. Step 1: Existence of flow on $[0, T]$. We will define a Picard iteration scheme as follows. Fix $T > 0$ and for each $n \geq 1$, fix $\nu_0^{(n)} = \nu_0^{(0)} = \nu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. Then define the flow $(\nu_t^{(n)})_{t \in [0, T]}$ by

$$\nu_t^{(n)} := e^{-\alpha t} \nu_0 + \int_0^t \alpha e^{-\alpha(t-s)} \Psi_\sigma[\nu_s^{(n-1)}] ds, \quad t \in [0, T]. \quad (49)$$

For fixed $T > 0$, we consider the sequence of flows $\left((\nu_t^{(n)})_{t \in [0, T]} \right)_{n=0}^\infty$ in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$, where, for any $(\nu_t)_{t \in [0, T]} \in \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}$, the distance $\mathcal{W}_1^{[0, T]}$ is defined by

$$\mathcal{W}_1^{[0, T]} \left((\nu_t)_{t \in [0, T]}, (\nu'_t)_{t \in [0, T]} \right) := \int_0^T \mathcal{W}_1(\nu_t, \nu'_t) dt.$$

Since $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$ is a complete metric space, we can apply the argument from [41, Lemma A.5] with $p = 1$ to conclude that $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \int_0^T \mathcal{W}_1(\nu_t, \nu'_t) dt \right)$ is a complete metric space.

We consider the Picard iteration mapping $\varphi \left((\nu_t^{(n-1)})_{t \in [0, T]} \right) := (\nu_t^{(n)})_{t \in [0, T]}$ defined via (49) and show that φ admits a unique fixed point $(\nu_t)_{t \in [0, T]}$ in the complete space $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$. Then this fixed point is the solution to (4).

Lemma C.3. *The mapping $\varphi \left((\nu_t^{(n-1)})_{t \in [0, T]} \right) := (\nu_t^{(n)})_{t \in [0, T]}$ defined via (49) admits a unique fixed point in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$.*

Proof of Lemma C.3. Step 1: The sequence of flows $\left((\nu_t^{(n)})_{t \in [0, T]} \right)_{n=0}^\infty$ is a Cauchy sequence in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$.

Let $\psi \in \text{Lip}_1(\mathbb{R}^d)$. Then we have that

$$\begin{aligned}
& \int_{\mathbb{R}^d} \psi(x) \left(\nu_t^{(n)} - \nu_t^{(n-1)} \right) (dx) \\
&= \int_{\mathbb{R}^d} \alpha \psi(x) \left(\int_0^t e^{-\alpha(t-s)} \Psi_\sigma \left[\nu_s^{(n-1)} \right] ds - \int_0^t e^{-\alpha(t-s)} \Psi_\sigma \left[\nu_s^{(n-2)} \right] ds \right) (dx) \\
&= \int_{\mathbb{R}^d} \alpha \psi(x) \left(\int_0^t e^{-\alpha(t-s)} \left(\Psi_\sigma \left[\nu_s^{(n-1)} \right] - \Psi_\sigma \left[\nu_s^{(n-2)} \right] \right) ds \right) (dx) \\
&= \int_0^t \alpha e^{-\alpha(t-s)} \int_{\mathbb{R}^d} \psi(x) \left(\Psi_\sigma \left[\nu_s^{(n-1)} \right] - \Psi_\sigma \left[\nu_s^{(n-2)} \right] \right) (dx) ds \\
&\leq \int_0^t \alpha e^{-\alpha(t-s)} \mathcal{W}_1 \left(\Psi_\sigma \left[\nu_s^{(n-1)} \right], \Psi_\sigma \left[\nu_s^{(n-2)} \right] \right) ds \\
&\leq \alpha L_{\Psi_\sigma, U^\varepsilon} \int_0^t \mathcal{W}_1 \left(\nu_s^{(n-1)}, \nu_s^{(n-2)} \right) ds,
\end{aligned} \tag{50}$$

where in the third equality we used Fubini's theorem, in the first inequality we used the definition of \mathcal{W}_1 , and in the last inequality we used Theorem 2.3 and the fact that $e^{-\alpha(t-s)} \leq 1$ for $s \in [0, t]$.

Taking supremum over ψ in (50) gives

$$\begin{aligned}
\mathcal{W}_1 \left(\nu_t^{(n)}, \nu_t^{(n-1)} \right) &\leq \alpha L_{\Psi_\sigma, U^\varepsilon} \int_0^t \mathcal{W}_1 \left(\nu_s^{(n-1)}, \nu_s^{(n-2)} \right) ds \\
&\leq (\alpha L_{\Psi_\sigma, U^\varepsilon})^{n-1} \int_0^t \int_0^{t_1} \dots \int_0^{t_{n-2}} \mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) dt_{n-1} \dots dt_2 dt_1 \\
&\leq (\alpha L_{\Psi_\sigma, U^\varepsilon})^{n-1} \frac{t^{n-2}}{(n-2)!} \int_0^t \mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) dt_{n-1},
\end{aligned}$$

where in the third inequality we used the bound $\int_0^{t_{n-2}} dt_{n-1} \leq \int_0^t dt_{n-1}$. Hence, we obtain

$$\int_0^T \mathcal{W}_1 \left(\nu_t^{(n)}, \nu_t^{(n-1)} \right) dt \leq (\alpha L_{\Psi_\sigma, U^\varepsilon})^{n-1} \frac{T^{n-1}}{(n-1)!} \int_0^T \mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) dt_{n-1}.$$

Using the definition of $\mathcal{W}_1^{[0, T]}$, the last inequality becomes

$$\begin{aligned}
& \mathcal{W}_1^{[0, T]} \left((\nu_t^{(n)})_{t \in [0, T]}, (\nu_t^{(n-1)})_{t \in [0, T]} \right) \\
&\leq (\alpha L_{\Psi_\sigma, U^\varepsilon})^{n-1} \frac{T^{n-1}}{(n-1)!} \mathcal{W}_1^{[0, T]} \left((\nu_t^{(1)})_{t \in [0, T]}, (\nu_t^{(0)})_{t \in [0, T]} \right).
\end{aligned}$$

By choosing n sufficiently large, we conclude that $\left((\nu_t^{(n)})_{t \in [0, T]} \right)_{n=0}^\infty$ is a Cauchy sequence.

By completeness of $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$, the sequence admits a limit point $(\nu_t)_{t \in [0, T]} \in \left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]} \right)$.

Step 2: The limit point $(\nu_t)_{t \in [0, T]}$ is a fixed point of φ . From Step 1, we obtain that for Lebesgue-almost all $t \in [0, T]$ we have

$$\mathcal{W}_1(\nu_t^{(n)}, \nu_t) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Therefore, by Theorem 2.3, for Lebesgue-almost all $t \in [0, T]$, we have that

$$\mathcal{W}_1 \left(\Psi_\sigma \left[\nu_t^{(n)} \right], \Psi_\sigma \left[\nu_t \right] \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Hence, letting $n \rightarrow \infty$ in (49) and using the dominated convergence theorem (which is possible since Ψ_σ is uniformly bounded in n due to Assumption 2.1), we conclude that $(\nu_t)_{t \in [0, T]}$ is a fixed point of φ .

Step 3: The fixed point $(\nu_t)_{t \in [0, T]}$ of φ is unique. Suppose, for the contrary, that φ admits two fixed points $(\nu_t)_{t \in [0, T]}$ and $(\bar{\nu}_t)_{t \in [0, T]}$ such that $\nu_0 = \bar{\nu}_0$. Then repeating the same calculations from (50), we arrive at

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) \leq \alpha L_{\Psi_\sigma, U^\xi} \int_0^t \mathcal{W}_1(\nu_s, \bar{\nu}_s) ds.$$

For each $t \in [0, T]$, denote $f(t) := \int_0^t \mathcal{W}_1(\nu_s, \bar{\nu}_s) ds$. Observe that $f \geq 0$ and $f(0) = 0$. Then, by Gronwall's lemma, we obtain

$$0 \leq f(t) \leq e^{t\alpha L_{\Psi_\sigma, U^\xi}} f(0) = 0,$$

and hence

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) = 0,$$

for Lebesgue-almost all $t \in [0, T]$, which implies

$$\nu_t = \bar{\nu}_t,$$

for Lebesgue-almost all $t \in [0, T]$. Therefore, the fixed point $(\nu_t)_{t \in [0, T]}$ of φ must be unique.

From Steps 1, 2 and 3, we obtain the existence and uniqueness of a flow $(\nu_t)_{t \in [0, T]}$ satisfying (4) for any $T > 0$. \square

Having proved Lemma C.3, we return to the proof of Theorem C.2.

Step 2: Existence of the flow on $[0, \infty)$.

From Lemma C.3, for any $T > 0$, there exists unique flow $(\nu_t)_{t \in [0, T]}$ satisfying (4). It remains to prove that the existence of this flow could be extended to $[0, \infty)$. Let $(\nu_t)_{t \in [0, T]}, (\nu'_t)_{t \in [0, T]} \in \mathcal{P}_1(\mathbb{R}^d)$. Then, using the calculations from Lemma C.3, we have that

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) \leq \alpha L_{\Psi_\sigma, U^\xi} \int_0^t \mathcal{W}_1(\nu_s, \bar{\nu}_s) ds,$$

which shows that $(\nu_t)_{t \in [0, T]}$ does not blow up in any finite time, and therefore we can extend $(\nu_t)_{t \in [0, T]}$ globally to $(\nu_t)_{t \in [0, \infty)}$. By definition, $\Psi_\sigma[\nu]$ admits a density of the form

$$\Psi_\sigma[\nu_t](x) = \frac{\exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_t, x) - U^\xi(x)\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_t, x') - U^\xi(x')\right) dx'}.$$

For any fixed $x \in \mathbb{R}^d$, the flat derivative $t \mapsto \frac{\delta F}{\delta \nu}(\nu_t, x)$ is continuous on $[0, \infty)$ due to the fact that $\nu_t \in C([0, \infty), \mathcal{P}_1(\mathbb{R}^d))$ and $\nu \mapsto \frac{\delta F}{\delta \nu}(\nu, x)$ is continuous. Moreover, the flat derivative $\frac{\delta F}{\delta \nu}(\nu_t, x)$ is bounded for every $x \in \mathbb{R}^d$ and all $t \geq 0$ due to Assumption 2.1. Therefore, both terms $Z_\sigma(\nu_t)$ and $\exp\left(-\frac{1}{\sigma} \frac{\delta F}{\delta \nu}(\nu_t, x) - U^\xi(x)\right)$ are continuous in t and bounded for every $x \in \mathbb{R}^d$ and every $t \geq 0$. Hence, we have that $\Psi_\sigma[\nu_t](x)$ is continuous in t and bounded for every $x \in \mathbb{R}^d$. Since by assumption $\nu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ admits a density $\nu_0(x)$, we define the density of ν_t by

$$\nu_t(x) := e^{-\alpha t} \nu_0(x) + \int_0^t \alpha e^{-\alpha(t-s)} \Psi_\sigma[\nu_s](x) ds.$$

By definition $[0, \infty) \ni t \mapsto \nu_t(x)$ is continuous for every $x \in \mathbb{R}^d$. Since $s \mapsto e^{-\alpha(t-s)} \Psi_\sigma[\nu_s](x)$ is continuous and bounded in s for every $t \geq 0$, it follows that $t \mapsto \nu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$ and $\nu_t(x)$ satisfies (48). \square

C.2 Min-max problems

We recall a necessary condition for the optimality of MNEs of (6).

Proposition C.4 (Necessary condition for optimality; [11, Theorem 3.1]). *Let Assumptions 3.1, 3.2 hold and let $\sigma_\nu, \sigma_\mu > 0$. If the pair $(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$ is an MNE, then $\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*$ are equivalent to λ , and for λ -a.a. $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$,*

$$\nu_{\sigma_\nu}^*(dx) = \frac{\exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, x) - U^\xi(x)\right)}{\int_{\mathbb{R}^d} \exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, x') - U^\xi(x')\right) dx'} dx,$$

$$\mu_{\sigma_\mu}^*(dy) = \frac{\exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, y) - U^\rho(y)\right)}{\int_{\mathbb{R}^d} \exp\left(\frac{1}{\sigma_\mu} \frac{\delta F}{\delta \mu}(\nu_{\sigma_\nu}^*, \mu_{\sigma_\mu}^*, y') - U^\rho(y')\right) dy'} dy.$$

Now we give the proof of the existence and uniqueness of the flow (8), which is an extension of the proof of Proposition C.2. Note that the proof closely follows the argument of [23, Theorem 1], with the key difference being that we use the Wasserstein distance in place of the Total Variation distance. We include the proof here for completeness.

Proposition C.5 (Existence and uniqueness of the flow). *Let Assumptions 3.1, 3.2 hold and let $(\nu_0, \mu_0) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d)$. Then there exists a unique solution $(\nu_t, \mu_t)_{t \geq 0} \in C([0, \infty); \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d))$ to (8) and the solution depends continuously on the initial condition. Moreover, if $(\nu_0, \mu_0) \in \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$, then $(\nu_t, \mu_t)_{t \geq 0}$ admits the density $(\nu_t(\cdot), \mu_t(\cdot))_{t \geq 0} \subset \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$ such that, for every $x, y \in \mathbb{R}^d$, it holds that $t \mapsto \nu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$, $t \mapsto \mu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$ and*

$$\begin{cases} d\nu_t(x) = \alpha_\nu (\Psi_{\sigma_\nu}[\nu_t, \mu_t](x) - \nu_t(x)) dt, \\ d\mu_t(y) = \alpha_\mu (\Phi_{\sigma_\mu}[\nu_t, \mu_t](y) - \mu_t(y)) dt, \quad t > 0, \end{cases} \quad (51)$$

for some initial condition $(\nu_0(x), \mu_0(y)) \in \mathcal{P}_1^\lambda(\mathbb{R}^d) \times \mathcal{P}_1^\lambda(\mathbb{R}^d)$.

Proof of Theorem C.5. Step 1: Existence of flow on $[0, T]$. We will define a Picard iteration scheme as follows. Fix $T > 0$ and for each $n \geq 1$, fix $\nu_0^{(n)} = \nu_0^{(0)} = \nu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ and

$$\mu_0^{(n)} = \mu_0^{(0)} = \mu_0 \in \mathcal{P}_1(\mathbb{R}^d).$$

Then define the flow $(\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}$ by

$$\nu_t^{(n)} := e^{-\alpha_\nu t} \nu_0 + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu}[\nu_s^{(n-1)}, \mu_s^{(n-1)}] ds, \quad t \in [0, T], \quad (52)$$

$$\mu_t^{(n)} := e^{-\alpha_\mu t} \mu_0 + \int_0^t \alpha_\mu e^{-\alpha_\mu(t-s)} \Phi_{\sigma_\mu}[\nu_s^{(n-1)}, \mu_s^{(n-1)}] ds, \quad t \in [0, T]. \quad (53)$$

For fixed $T > 0$, we consider the sequence of flows $\left((\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}\right)_{n=0}^\infty$ in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]}\right)$, where, for any $(\nu_t, \mu_t)_{t \in [0, T]} \in \mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}$, the distance $\mathcal{W}_1^{[0, T]}$ is defined by

$$\mathcal{W}_1^{[0, T]}((\nu_t, \mu_t)_{t \in [0, T]}, (\nu'_t, \mu'_t)_{t \in [0, T]}) := \int_0^T \mathcal{W}_1(\nu_t, \nu'_t) dt + \int_0^T \mathcal{W}_1(\mu_t, \mu'_t) dt.$$

Since $(\mathcal{P}_1(\mathbb{R}^d), \mathcal{W}_1)$ is a complete metric space, we can apply the argument from [41, Lemma A.5] with $p = 1$ to conclude that $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \int_0^T \mathcal{W}_1(\nu_t, \nu'_t) dt\right)$ and $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \int_0^T \mathcal{W}_1(\mu_t, \mu'_t) dt\right)$ are a complete metric spaces, and hence so is $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]}\right)$.

We consider the Picard iteration mapping $\varphi \left((\nu_t^{(n-1)}, \mu_t^{(n-1)})_{t \in [0, T]}\right) := (\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}$ defined via (52) and (53) and show that φ admits a unique fixed point $(\nu_t, \mu_t)_{t \in [0, T]}$ in the complete space $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]}\right)$. Then this fixed point is the solution to (8).

Lemma C.6. *The mapping $\varphi \left((\nu_t^{(n-1)}, \mu_t^{(n-1)})_{t \in [0, T]}\right) := (\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}$ defined via (52) and (53) admits a unique fixed point in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]}\right)$.*

Proof of Lemma C.6. Step 1: The sequence of flows $\left((\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0, T]}\right)_{n=0}^{\infty}$ is a Cauchy sequence in $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0, T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0, T]}, \mathcal{W}_1^{[0, T]}\right)$.

Let $\psi \in \text{Lip}_1(\mathbb{R}^d)$. Then we have that

$$\begin{aligned}
& \int_{\mathbb{R}^d} \psi(x) \left(\nu_t^{(n)} - \nu_t^{(n-1)} \right) (dx) \\
&= \int_{\mathbb{R}^d} \psi(x) \left(\int_0^t e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu} \left[\nu_s^{(n-1)}, \mu_s^{(n-1)} \right] ds \right. \\
&\quad \left. - \int_0^t e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu} \left[\nu_s^{(n-2)}, \mu_s^{(n-2)} \right] ds \right) (dx) \\
&= \int_{\mathbb{R}^d} \psi(x) \left(\int_0^t e^{-\alpha_\nu(t-s)} \left(\Psi_{\sigma_\nu} \left[\nu_s^{(n-1)}, \mu_s^{(n-1)} \right] - \Psi_{\sigma_\nu} \left[\nu_s^{(n-2)}, \mu_s^{(n-2)} \right] \right) ds \right) (dx) \quad (54) \\
&= \int_0^t e^{-\alpha_\nu(t-s)} \int_{\mathbb{R}^d} \psi(x) \left(\Psi_{\sigma_\nu} \left[\nu_s^{(n-1)}, \mu_s^{(n-1)} \right] - \Psi_{\sigma_\nu} \left[\nu_s^{(n-2)}, \mu_s^{(n-2)} \right] \right) (dx) ds \\
&\leq \int_0^t e^{-\alpha_\nu(t-s)} \mathcal{W}_1 \left(\Psi_{\sigma_\nu} \left[\nu_s^{(n-1)}, \mu_s^{(n-1)} \right], \Psi_{\sigma_\nu} \left[\nu_s^{(n-2)}, \mu_s^{(n-2)} \right] \right) ds \\
&\leq L_{\Psi_{\sigma_\nu}, U^\varepsilon} \int_0^t \left(\mathcal{W}_1 \left(\nu_s^{(n-1)}, \nu_s^{(n-2)} \right) + \mathcal{W}_1 \left(\mu_s^{(n-1)}, \mu_s^{(n-2)} \right) \right) ds,
\end{aligned}$$

where in the third equality we used Fubini's theorem, in the first inequality we used the definition of \mathcal{W}_1 , and in the last inequality we used (39) and the fact that $e^{-\alpha(t-s)} \leq 1$ for $s \in [0, t]$.

A similar argument using (40) yields

$$\begin{aligned}
& \int_{\mathbb{R}^d} \psi(x) \left(\mu_t^{(n)} - \mu_t^{(n-1)} \right) (dx) \\
&\leq L_{\Phi_{\sigma_\mu}, U^\rho} \int_0^t \left(\mathcal{W}_1 \left(\nu_s^{(n-1)}, \nu_s^{(n-2)} \right) + \mathcal{W}_1 \left(\mu_s^{(n-1)}, \mu_s^{(n-2)} \right) \right) ds. \quad (55)
\end{aligned}$$

Taking supremum over ψ in (54) and (55) and adding both together gives

$$\begin{aligned}
& \mathcal{W}_1 \left(\nu_t^{(n)}, \nu_t^{(n-1)} \right) + \mathcal{W}_1 \left(\mu_t^{(n)}, \mu_t^{(n-1)} \right) \\
&\leq \left(L_{\Psi_{\sigma_\nu}, U^\varepsilon} + L_{\Phi_{\sigma_\mu}, U^\rho} \right) \int_0^t \left(\mathcal{W}_1 \left(\nu_s^{(n-1)}, \nu_s^{(n-2)} \right) + \mathcal{W}_1 \left(\mu_s^{(n-1)}, \mu_s^{(n-2)} \right) \right) ds \\
&\leq \left(L_{\Psi_{\sigma_\nu}, U^\varepsilon} + L_{\Phi_{\sigma_\mu}, U^\rho} \right)^{n-1} \times \\
&\quad \times \int_0^t \int_0^{t_1} \dots \int_0^{t_{n-2}} \left(\mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) + \mathcal{W}_1 \left(\mu_{t_{n-1}}^{(1)}, \mu_{t_{n-1}}^{(0)} \right) \right) dt_{n-1} \dots dt_2 dt_1 \\
&\leq \left(L_{\Psi_{\sigma_\nu}, U^\varepsilon} + L_{\Phi_{\sigma_\mu}, U^\rho} \right)^{n-1} \frac{t^{n-2}}{(n-2)!} \int_0^t \left(\mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) + \mathcal{W}_1 \left(\mu_{t_{n-1}}^{(1)}, \mu_{t_{n-1}}^{(0)} \right) \right) dt_{n-1},
\end{aligned}$$

where in the third inequality we used the bound $\int_0^t \int_0^{t_{n-2}} dt_{n-1} \leq \int_0^t dt_{n-1}$. Hence, we obtain

$$\begin{aligned}
& \int_0^T \left(\mathcal{W}_1 \left(\nu_t^{(n)}, \nu_t^{(n-1)} \right) + \mathcal{W}_1 \left(\mu_t^{(n)}, \mu_t^{(n-1)} \right) \right) dt \\
&\leq \left(L_{\Psi_{\sigma_\nu}, U^\varepsilon} + L_{\Phi_{\sigma_\mu}, U^\rho} \right)^{n-1} \frac{T^{n-1}}{(n-1)!} \int_0^T \left(\mathcal{W}_1 \left(\nu_{t_{n-1}}^{(1)}, \nu_{t_{n-1}}^{(0)} \right) + \mathcal{W}_1 \left(\mu_{t_{n-1}}^{(1)}, \mu_{t_{n-1}}^{(0)} \right) \right) dt_{n-1}.
\end{aligned}$$

Using the definition of $\mathcal{W}_1^{[0,T]}$, the last inequality becomes

$$\begin{aligned} & \mathcal{W}_1^{[0,T]} \left((\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0,T]}, (\nu_t^{(n-1)}, \mu_t^{(n-1)})_{t \in [0,T]} \right) \\ & \leq \left(L_{\Psi_{\sigma_\nu, U^\xi}} + L_{\Phi_{\sigma_\mu, U^\rho}} \right)^{n-1} \frac{T^{n-1}}{(n-1)!} \mathcal{W}_1^{[0,T]} \left((\nu_t^{(1)}, \mu_t^{(1)})_{t \in [0,T]}, (\nu_t^{(0)}, \mu_t^{(0)})_{t \in [0,T]} \right). \end{aligned}$$

By choosing n sufficiently large, we conclude that $\left((\nu_t^{(n)}, \mu_t^{(n)})_{t \in [0,T]} \right)_{n=0}^\infty$ is a Cauchy sequence. By completeness of $\left(\mathcal{P}_1(\mathbb{R}^d)^{[0,T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0,T]}, \mathcal{W}_1^{[0,T]} \right)$, the sequence admits a limit point $(\nu_t, \mu_t)_{t \in [0,T]} \in \left(\mathcal{P}_1(\mathbb{R}^d)^{[0,T]} \times \mathcal{P}_1(\mathbb{R}^d)^{[0,T]}, \mathcal{W}_1^{[0,T]} \right)$.

Step 2: The limit point $(\nu_t, \mu_t)_{t \in [0,T]}$ is a fixed point of φ . From Step 1, we obtain that for Lebesgue-almost all $t \in [0, T]$ we have

$$\mathcal{W}_1(\nu_t^{(n)}, \nu_t) \rightarrow 0, \quad \mathcal{W}_1(\mu_t^{(n)}, \mu_t) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Therefore, by (39) and (40), for Lebesgue-almost all $t \in [0, T]$, we have that

$$\mathcal{W}_1 \left(\Psi_{\sigma_\nu} \left[\nu_t^{(n)}, \mu_t^{(n)} \right], \Psi_{\sigma_\nu} \left[\nu_t, \mu_t \right] \right) \rightarrow 0, \quad \mathcal{W}_1 \left(\Phi_{\sigma_\mu} \left[\nu_t^{(n)}, \mu_t^{(n)} \right], \Phi_{\sigma_\mu} \left[\nu_t, \mu_t \right] \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Hence, letting $n \rightarrow \infty$ in (52) and in (53) and using the dominated convergence theorem (which is possible since $\Psi_{\sigma_\nu}, \Phi_{\sigma_\mu}$ are uniformly bounded in n due to Assumption 3.1), we conclude that $(\nu_t, \mu_t)_{t \in [0,T]}$ is a fixed point of φ .

Step 3: The fixed point $(\nu_t, \mu_t)_{t \in [0,T]}$ of φ is unique. Suppose, for the contrary, that φ admits two fixed points $(\nu_t, \mu_t)_{t \in [0,T]}$ and $(\bar{\nu}_t, \bar{\mu}_t)_{t \in [0,T]}$ such that $\nu_0 = \bar{\nu}_0$ and $\mu_0 = \bar{\mu}_0$. Then repeating the same calculations from (54) and (55), we arrive at

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) + \mathcal{W}_1(\mu_t, \bar{\mu}_t) \leq \left(L_{\Psi_{\sigma_\nu, U^\xi}} + L_{\Phi_{\sigma_\mu, U^\rho}} \right) \int_0^t (\mathcal{W}_1(\nu_s, \bar{\nu}_s) + \mathcal{W}_1(\mu_s, \bar{\mu}_s)) ds.$$

For each $t \in [0, T]$, denote $f(t) := \int_0^t (\mathcal{W}_1(\nu_s, \bar{\nu}_s) + \mathcal{W}_1(\mu_s, \bar{\mu}_s)) ds$. Observe that $f \geq 0$ and $f(0) = 0$. Then, by Gronwall's lemma, we obtain

$$0 \leq f(t) \leq e^{\left(L_{\Psi_{\sigma_\nu, U^\xi}} + L_{\Phi_{\sigma_\mu, U^\rho}} \right) t} f(0) = 0,$$

and hence

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) + \mathcal{W}_1(\mu_t, \bar{\mu}_t) = 0,$$

for Lebesgue-almost all $t \in [0, T]$, which implies

$$\nu_t = \bar{\nu}_t, \quad \mu_t = \bar{\mu}_t,$$

for Lebesgue-almost all $t \in [0, T]$. Therefore, the fixed point $(\nu_t, \mu_t)_{t \in [0,T]}$ of φ must be unique.

From Steps 1, 2 and 3, we obtain the existence and uniqueness of a flow $(\nu_t, \mu_t)_{t \in [0,T]}$ satisfying (8) for any $T > 0$. \square

Having proved Lemma C.6, we return to the proof of Theorem C.5.

Step 2: Existence of the flow on $[0, \infty)$.

From Lemma C.6, for any $T > 0$, there exists unique flow $(\nu_t, \mu_t)_{t \in [0,T]}$ satisfying (8). It remains to prove that the existence of this flow could be extended to $[0, \infty)$. Let $(\nu_t, \mu_t)_{t \in [0,T]}, (\nu'_t, \mu'_t)_{t \in [0,T]} \in \mathcal{P}_1(\mathbb{R}^d)$. Then, using the calculations from Lemma C.6, we have that

$$\mathcal{W}_1(\nu_t, \bar{\nu}_t) + \mathcal{W}_1(\mu_t, \bar{\mu}_t) \leq \left(L_{\Psi_{\sigma_\nu, U^\xi}} + L_{\Phi_{\sigma_\mu, U^\rho}} \right) \int_0^t (\mathcal{W}_1(\nu_s, \bar{\nu}_s) + \mathcal{W}_1(\mu_s, \bar{\mu}_s)) ds,$$

which shows that $(\nu_t, \mu_t)_{t \in [0,T]}$ does not blow up in any finite time, and therefore we can extend $(\nu_t, \mu_t)_{t \in [0,T]}$ globally to $(\nu_t, \mu_t)_{t \in [0, \infty)}$. By definition, $\Psi_{\sigma_\nu}[\nu, \mu]$ admits a density of the form

$$\Psi_{\sigma_\nu}[\nu_t, \mu_t](x) = \frac{\exp \left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_t, \mu_t, x) - U^\xi(x) \right)}{\int_{\mathbb{R}^d} \exp \left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_t, \mu_t, x') - U^\xi(x') \right) dx'}.$$

For any fixed $x \in \mathbb{R}^d$, the flat derivative $t \mapsto \frac{\delta F}{\delta \nu}(\nu_t, \mu_t, x)$ is continuous on $[0, \infty)$ due to the fact that $\nu_t, \mu_t \in C([0, \infty), \mathcal{P}_1(\mathbb{R}^d))$ and $(\nu, \mu) \mapsto \frac{\delta F}{\delta \nu}(\nu, \mu, x)$ is continuous. Moreover, the flat derivative $\frac{\delta F}{\delta \nu}(\nu_t, \mu_t, x)$ is bounded for every $x \in \mathbb{R}^d$ and all $t \geq 0$ due to Assumption 3.1. Therefore, both terms $Z_{\sigma_\nu}(\nu_t, \mu_t)$ and $\exp\left(-\frac{1}{\sigma_\nu} \frac{\delta F}{\delta \nu}(\nu_t, \mu_t, x) - U^\xi(x)\right)$ are continuous in t and bounded for every $x \in \mathbb{R}^d$ and every $t \geq 0$. Hence, we have that $\Psi_{\sigma_\nu}[\nu_t, \mu_t](x)$ is continuous in t and bounded for every $x \in \mathbb{R}^d$. Since by assumption $\nu_0 \in \mathcal{P}_1(\mathbb{R}^d)$ admits a density $\nu_0(x)$, we define the density of ν_t by

$$\nu_t(x) := e^{-\alpha_\nu t} \nu_0(x) + \int_0^t \alpha_\nu e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu}[\nu_s, \mu_s](x) ds.$$

By definition $[0, \infty) \ni t \mapsto \nu_t(x)$ is continuous for every $x \in \mathbb{R}^d$. Since $s \mapsto e^{-\alpha_\nu(t-s)} \Psi_{\sigma_\nu}[\nu_s, \mu_s](x)$ is continuous and bounded in s for every $t \geq 0$, it follows that $t \mapsto \nu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$ and $\nu_t(x)$ satisfies (51). The same argument gives that $\Phi_{\sigma_\mu}[\nu_t, \mu_t](y)$ is continuous in t and bounded for every $y \in \mathbb{R}^d$, consequently that μ_t admits a density $\mu_t(y)$ that satisfies (51) and $t \mapsto \mu_t \in C^1([0, \infty), \mathcal{P}_1^\lambda(\mathbb{R}^d))$. \square

D Notation for MDPs

In this section, we present standard notation for MDPs, following the conventions used in [24]. Let (E, d) denote a Polish space, i.e., a complete separable metric space. We endow E with its Borel σ -algebra $\mathcal{B}(E)$. Let $B_b(E)$ denote the space of bounded measurable functions $f : E \rightarrow \mathbb{R}$ endowed with the supremum norm $\|f\|_{B_b(E)} = \sup_{x \in E} |f(x)|$. Let $\mathcal{M}(E)$ denote the Banach space of finite signed measures m on E endowed with the total variation norm $\|m\|_{\mathcal{M}(E)} = |m|(E)$, where $|m|$ is the total-variation measure. We denote by $\mathcal{M}_+(E) \subset \mathcal{M}(E)$ the subset of finite positive measures. For a measure $m \in \mathcal{M}_+(E)$ and measurable function $f : E \rightarrow \mathbb{R}$, let

$$\text{ess sup}_{x \in E} f := \inf\{c \in \mathbb{R} : m\{x \in E : f(x) > c\} = 0\}.$$

For given Polish spaces (E_1, d_1) and (E_2, d_2) , denote by $b\mathcal{K}(E_1|E_2)$ the Banach space of bounded signed kernels $k : E_2 \rightarrow \mathcal{M}(E_1)$ endowed with the norm $\|k\|_{b\mathcal{K}(E_1|E_2)} = \sup_{x \in E_2} \|k(x)\|_{\mathcal{M}(E_1)}$; that is, $k(U|\cdot) : E_2 \rightarrow \mathbb{R}$ is measurable for all $U \in \mathcal{M}(E_1)$ and $k(\cdot|x) \in \mathcal{M}(E_1)$ for all $x \in E_2$. For a fixed reference measure $\eta \in \mathcal{M}_+(E_1)$, we denote by $b\mathcal{K}_\eta(E_1|E_2)$ the space of bounded kernels that are absolutely continuous with respect to η . Every kernel $k \in b\mathcal{K}(E_1|E_2)$ induces a bounded linear operator $T_k \in \mathcal{L}(\mathcal{M}(E_2), \mathcal{M}(E_1))$ defined by

$$T_k \eta(dy) = \eta k(dy) = \int_{E_2} \eta(dx) k(dy|x).$$

Moreover, we have

$$\|k\|_{b\mathcal{K}(E_1|E_2)} = \sup_{x \in E_2} \sup_{\substack{h \in B_b(E_1) \\ \|h\|_{B_b(E_1)} \leq 1}} \int_{E_1} h(y) k(dy|x) = \|T_k\|_{\mathcal{L}(\mathcal{M}(E_2), \mathcal{M}(E_1))}, \quad (56)$$

where the latter is the operator norm. Thus, $b\mathcal{K}(E|E)$ is a Banach algebra with the product defined via composition of the corresponding linear operators. In particular, for given $k \in b\mathcal{K}(E|E)$,

$$T_k^n \mu(dy) = \mu k^n(dy) = \int_{E^n} \mu(dx_0) k(dx_1|x_0) \cdots k(dx_{n-1}|x_{n-2}) k(dy|x_{n-1}). \quad (57)$$

Let $\mathcal{P}(E_1|E_2)$ denote the convex subspace of $b\mathcal{K}(E_1|E_2)$ such that $P(\cdot|x) \in \mathcal{P}(E_1)$ for all $x \in E_2$.

We denote by $((S \times A)^\mathbb{N}, \mathcal{F})$ a sample space, where the elements of $(S \times A)^\mathbb{N}$ are state-action pairs $(s_n, a_n)_{n=0}^\infty$ with $(s_n, a_n) \in S \times A$, for each $n \in \mathbb{N}$, and \mathcal{F} is the associated σ -algebra. By [3, Proposition 7.28], for a given initial distribution $\gamma \in \mathcal{P}(S)$ and policy $\pi \in \mathcal{P}(A|S)$, there exists a unique product probability measure \mathbb{P}_γ^π on $((S \times A)^\mathbb{N}, \mathcal{F})$ such that for every $n \in \mathbb{N}$, we have

1. $\mathbb{P}_\gamma^\pi(s_0 \in S) = \gamma(S)$,
2. $\mathbb{P}_\gamma^\pi(a_n \in \mathcal{A} | (s_0, a_0, \dots, s_n)) = \pi(a_n | s_n)$,

$$3. \mathbb{P}_\gamma^\pi(s_{n+1} \in \mathcal{S} | (s_0, a_0, \dots, s_n, a_n)) = P(\mathcal{S} | s_n, a_n),$$

for all $\mathcal{S} \in \mathcal{B}(S)$ and $\mathcal{A} \in \mathcal{B}(A)$. Thus, $\{s_n\}_{n \geq 0}$ is a Markov chain with transition kernel $P_\pi \in \mathcal{P}(S|S)$ defined by

$$P_\pi(ds'|s) := \int_A P(ds'|s, a')\pi(da'|s). \quad (58)$$

The expectation corresponding to \mathbb{P}_γ^π is denoted by \mathbb{E}_γ^π . For given $s \in S$, we denote $\mathbb{E}_s^\pi := \mathbb{E}_{\delta_s}^\pi$, where $\delta_s \in \mathcal{P}(S)$ denotes the Dirac measure at $s \in S$.

E Auxiliary results for MDPs

In this section, we present two sets of auxiliary results, each corresponding to Section 2 and Section 3, respectively.

E.1 Single-agent optimization

First, we prove that the value function V_τ in the MDP setting of Section 2.2 satisfies Assumption 2.1. Our proof closely follows the argument in [24, Theorem 2.4], but we include it here for completeness.

Before we give the proof, we need to derive an alternative expression of V_τ to (2.2). Again, for full details on notation, we refer to Appendix D. By the Bellman principle (see, e.g., [20, Lemma B.2]), for all $\pi \in \mathcal{P}^\eta(A|S)$ and $s \in S$, it follows that

$$V_\tau^\pi(s) = \int_A \left(Q_\tau^\pi(s, a) + \tau \log \frac{d\pi}{d\eta}(a|s) \right) \pi(da|s). \quad (59)$$

For a given policy $\pi \in \mathcal{P}^\eta(A|S)$, the occupancy kernel $d^\pi \in \mathcal{P}(S|S)$ is defined by

$$d^\pi(ds'|s) = (1 - \delta) \sum_{n=0}^{\infty} \delta^n P_\pi^n(ds'|s), \quad (60)$$

where $P_\pi^0(ds'|s) := \delta_s(ds')$, for the Dirac measure δ_s at $s \in S$, P_π^n is a product of kernels in the sense of (57), and the convergence of the series is understood in $b\mathcal{K}(S|S)$. For a given initial distribution $\gamma \in \mathcal{P}(S)$, the state occupancy measure $d_\gamma^\pi \in \mathcal{P}(S)$ is defined by

$$d_\gamma^\pi(ds') := \int_S d^\pi(ds'|s)\gamma(ds).$$

Using (12) and (58) in (59) gives for all $s \in S$ that

$$V_\tau^\pi(s) = \int_A \left(c(s, a) + \tau \log \frac{d\pi}{d\eta}(a|s) \right) \pi(da|s) + \delta \int_S V_\tau^\pi(s')P_\pi(ds'|s).$$

Applying this identity recursively and using (60) yields for all $s \in S$ that

$$V_\tau^\pi(s) = \frac{1}{1 - \delta} \int_S \int_A \left[c(s', a) + \tau \log \frac{d\pi}{d\eta}(a|s') \right] \pi(da|s')d^\pi(ds'|s).$$

Finally, integrating over $\gamma \in \mathcal{P}(S)$ leads to

$$V_\tau^\pi(\gamma) = \frac{1}{1 - \delta} \int_S \int_A \left[c(s', a) + \tau \log \frac{d\pi}{d\eta}(a|s') \right] \pi(da|s')d_\gamma^\pi(ds'). \quad (61)$$

Now, we introduce the following necessary notation for the proof. For $k \in \{0, 1\}$, let \mathfrak{C}^k be the set of jointly measurable functions $h : \mathbb{R}^d \times S \times A \rightarrow \mathbb{R}$ such that for Lebesgue-almost all $\theta \in \mathbb{R}^d$, all $s \in S$, and η -almost all $a \in A$, h is k -times differentiable in θ , and satisfies

$$|h|_{\mathfrak{C}^k} := \max_{k \in \{0, 1\}} \operatorname{ess\,sup}_{a \in A} \operatorname{ess\,sup}_{s \in S} \operatorname{ess\,sup}_{\theta \in \mathbb{R}^d} |\nabla_\theta^k h(\theta, s, a)| < \infty,$$

where the essential supremum over A is defined relative to the reference measure η and the essential supremum over \mathbb{R}^d is defined relative to the Lebesgue measure.

For each $\nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\gamma \in \mathcal{P}(S)$ we define the maps

$$\mathcal{P}_1(\mathbb{R}^d) \ni \nu \mapsto \Pi(\nu)(da|s) := \pi_\nu(da|s) \in \mathcal{P}_\eta(A|S),$$

and

$$\mathcal{P}_1(\mathbb{R}^d) \ni \nu \mapsto \hat{V}_\tau(\nu)(\gamma) := V_\tau^{\pi_\nu}(\gamma) \in \mathbb{R},$$

where $\pi_\nu(da|s)$ and $\bar{V}_\tau(\nu)(\gamma)$ are given by (13) and (61), respectively.

The proof will depend on the following key lemmas, which we state without proof, as they are available in [24]. The first is Lemma [24, Lemma 2.2].

Lemma E.1 (Flat derivative of Π). *If $f \in \mathfrak{C}^0$, then the function $\Pi : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathcal{P}_\eta(A|S)$ has a flat derivative $\frac{\delta \Pi}{\delta \nu} : \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow b\mathcal{K}_\eta(A|S)$ given by*

$$\frac{\delta \Pi}{\delta \nu}(\nu, \theta)(da|s) = \left(f(\theta, s, a) - \int_A f(\theta, s, a') \Pi(\nu)(da'|s) \right) \Pi(\nu)(da|s). \quad (62)$$

The second is [24, Lemma 2.3], which can be viewed as a policy gradient theorem.

Lemma E.2 (Flat derivative of \hat{V}_τ). *If $f \in \mathfrak{C}^1$, then the function $\hat{V}_\tau : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ has a flat derivative $\frac{\delta \hat{V}_\tau}{\delta \nu} : \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$ given by*

$$\frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) = \frac{1}{1-\delta} \int_S \int_A \left(Q_\tau^{\Pi(\nu)}(s, a) + \tau \log \frac{d\Pi(\nu)}{d\eta}(a|s) \right) \frac{\delta \Pi}{\delta \nu}(\nu, \theta)(da|s) d_\gamma^{\Pi(\nu)}(ds). \quad (63)$$

We are ready to prove that \hat{V}_τ satisfies Assumption 2.1.

Proposition E.3. *Let S and A be Polish spaces, $c \in B_b(S \times A)$, $\eta \in \mathcal{M}_+(A)$, and $\tau > 0$. Then for all $\gamma \in \mathcal{P}(S)$ the function $F(\cdot) := \hat{V}_\tau(\cdot)(\gamma)$ satisfies Assumption 2.1 with*

$$C_F := \frac{2}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau (2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) |f|_{\mathfrak{C}^0},$$

$$L_F := |f|_{\mathfrak{C}^1} \left(\frac{1}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau (2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) \max \left\{ 2, \frac{5}{1-\delta} |f|_{\mathfrak{C}^0} \right\} + 4\tau |f|_{\mathfrak{C}^0} \right).$$

Proof. For given $(s, a) \in S \times A$ and $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, let

$$\bar{Q}_\tau^{\Pi(\nu)}(s, a) := \frac{1}{1-\delta} \left(Q_\tau^{\Pi(\nu)}(s, a) + \tau \log \frac{d\Pi(\nu)}{d\eta}(a|s) \right).$$

Since $c \in B_b(S \times A)$, $f \in \mathfrak{C}^0$ and $\eta \in \mathcal{M}_+(A)$, it follows by [24, Lemma A.4] that

$$\left| \bar{Q}_\tau^{\Pi(\nu)}(s, a) \right| \leq \frac{1}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau (2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)), \quad (64)$$

for all $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, $s \in S$ and $a \in A$ η -a.e. Therefore, by Lemma E.2, we obtain

$$\begin{aligned} & \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \right| \\ & \leq \int_S \int_A \left| \bar{Q}_\tau^{\Pi(\nu)}(s, a) \right| \left| f(\theta, s, a) - \int_A f(\theta, s, a') \Pi(\nu)(da'|s) \right| \Pi(\nu)(da|s) d_\gamma^{\Pi(\nu)}(ds) \\ & \leq \frac{2}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau (2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) |f|_{\mathfrak{C}^0} =: C_F, \end{aligned}$$

for all $\gamma \in \mathcal{P}(S)$ and $(\nu, \theta) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d$, which proves that $\frac{\delta \hat{V}_\tau}{\delta \nu}$ is uniformly bounded.

Next we show that, given $\gamma \in \mathcal{P}(S)$, $(\nu, \theta) \mapsto \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma)$ is Lipschitz. For any $(\nu', \theta'), (\nu, \theta) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d$, we have

$$\begin{aligned} & \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta')(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \right| \\ & \leq \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta')(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta)(\gamma) \right| + \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta)(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \right|. \end{aligned} \quad (65)$$

Note that since $f \in \mathfrak{C}^1$, we have

$$|f(\theta', s, a) - f(\theta, s, a)| \leq |f|_{\mathfrak{C}^1} |\theta' - \theta|,$$

for all $\theta \in \mathbb{R}^d$, $s \in S$ and $a \in A$ η -a.e. Then, for the first term in (65), we obtain

$$\begin{aligned} & \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta')(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta)(\gamma) \right| \\ & = \left| \int_S \int_A \bar{Q}_\tau^{\Pi(\nu')}(s, a) \left(\frac{\delta \Pi}{\delta \nu}(\nu', \theta')(da|s) - \frac{\delta \Pi}{\delta \nu}(\nu', \theta)(da|s) \right) d_\gamma^{\Pi(\nu')}(ds) \right| \\ & = \left| \int_S \int_A \bar{Q}_\tau^{\Pi(\nu')}(s, a) \left(f(\theta', s, a) - f(\theta, s, a) \right) \right. \\ & \quad \left. - \int_A (f(\theta', s, a') - f(\theta, s, a')) \Pi(\nu')(da'|s) \right) \Pi(\nu')(da|s) d_\gamma^{\Pi(\nu')}(ds) \right| \quad (66) \\ & \leq \int_S \int_A |\bar{Q}_\tau^{\Pi(\nu')}(s, a)| \left(|f(\theta', s, a) - f(\theta, s, a)| \right. \\ & \quad \left. + \int_A |f(\theta', s, a') - f(\theta, s, a')| \Pi(\nu')(da'|s) \right) \Pi(\nu')(da|s) d_\gamma^{\Pi(\nu')}(ds) \\ & \leq \frac{2|f|_{\mathfrak{C}^1}}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) |\theta' - \theta|. \end{aligned}$$

for all $\nu' \in \mathcal{P}_1(\mathbb{R}^d)$ and all $\theta', \theta \in \mathbb{R}^d$.

For the second term in (65), for $\theta \in \mathbb{R}^d$ and $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$, we have

$$\begin{aligned} & \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta)(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \\ & = \underbrace{\int_S \int_A \bar{Q}_\tau^{\Pi(\nu')}(s, a) \frac{\delta \Pi}{\delta \nu}(\nu', \theta)(da|s) [d_\gamma^{\Pi(\nu')} - d_\gamma^{\Pi(\nu)}]}_{:=I_1}(ds) \\ & \quad + \underbrace{\int_S \int_A [\bar{Q}_\tau^{\Pi(\nu')}(s, a) - \bar{Q}_\tau^{\Pi(\nu)}(s, a)] \frac{\delta \Pi}{\delta \nu}(\nu', \theta)(da|s) d_\gamma^{\Pi(\nu)}(ds)}_{:=I_2} \\ & \quad + \int_S \int_A \bar{Q}_\tau^{\Pi(\nu)}(s, a) \left[\frac{\delta \Pi}{\delta \nu}(\nu', \theta)(da|s) - \frac{\delta \Pi}{\delta \nu}(\nu, \theta)(da|s) \right] d_\gamma^{\Pi(\nu)}(ds) \\ & = I_1 + I_2 + \underbrace{\int_S \int_A \bar{Q}_\tau^{\Pi(\nu)}(s, a) \int_A f(\theta, s, a') [\Pi(\nu) - \Pi(\nu')](da'|s) \Pi(\nu)(da|s) d_\gamma^{\Pi(\nu)}(ds)}_{:=I_3} \\ & \quad + \underbrace{\int_S \int_A \bar{Q}_\tau^{\Pi(\nu)}(s, a) \left(f(\theta, s, a) - \int_A f(\theta, s, a') \Pi(\nu')(da'|s) \right) [\Pi(\nu') - \Pi(\nu)](da|s) d_\gamma^{\Pi(\nu)}(ds)}_{:=I_4}. \end{aligned}$$

By [24, Corollary A.5], we have for all $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$ that

$$|d^{\Pi(\nu')} - d^{\Pi(\nu)}|_{b\mathcal{K}(S|S)} \leq 2|f|_{\mathfrak{E}^1} \frac{\delta}{1-\delta} \mathcal{W}_1(\nu', \nu).$$

Hence, using (64), we obtain

$$|I_1| \leq \frac{4\delta}{(1-\delta)^3} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{E}^0} + |\log \eta(A)|)) |f|_{\mathfrak{E}^0} |f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu). \quad (67)$$

By the proof of [24, Corollary A.5], we have for all $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$ that

$$|\Pi(\nu') - \Pi(\nu)|_{b\mathcal{K}(A|S)} \leq 2|f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu). \quad (68)$$

Hence, by the definition of the $b\mathcal{K}(A|S)$ -norm (see (56)), for any measurable $h : \mathbb{R}^d \times S \times A \rightarrow \mathbb{R}$, we have

$$\int_A h(\theta, s, a)(\Pi(\nu') - \Pi(\nu))(da|s) \leq 2|f|_{\mathfrak{E}^1} |h|_{B_b(A)} \mathcal{W}_1(\nu', \nu).$$

Thus, using (68) and (64), we deduce

$$|I_3 + I_4| \leq \frac{4}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{E}^0} + |\log \eta(A)|)) |f|_{\mathfrak{E}^0} |f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu). \quad (69)$$

To estimate I_2 , applying Definition F.1 we observe that for all $(s, a) \in S \times A$, we have

$$\begin{aligned} & \log \frac{d\Pi(\nu')}{d\Pi}(a|s) - \log \frac{d\Pi(\nu)}{d\Pi}(a|s) \\ &= \int_0^1 \int_{\mathbb{R}^d} \left(f(\theta, s, a) - \int_A f(\theta, s, a') \Pi(\nu + \varepsilon(\nu' - \nu))(da'|s) \right) (\nu' - \nu)(d\theta) d\varepsilon \\ &\leq 2|f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu). \end{aligned}$$

Now, taking $\gamma = \delta_s$ in $\hat{V}_\tau(\nu)(\gamma) = \int_S \hat{V}_\tau(\nu)(s) \gamma(ds)$ gives

$$\hat{V}_\tau(\nu)(s) = \hat{V}_\tau(\nu)(\delta_s),$$

for all $s \in S$, and therefore

$$\begin{aligned} & Q_\tau^{\Pi(\nu')}(s, a) - Q_\tau^{\Pi(\nu)}(s, a) \\ &= \delta \int_S (\hat{V}_\tau(\nu')(s') - \hat{V}_\tau(\nu)(s')) P(ds'|s, a) \\ &= \delta \int_S \int_0^1 \int_{\mathbb{R}^d} \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu + \varepsilon(\nu' - \nu), \theta)(\delta_{s'}) (\nu' - \nu)(d\theta) d\varepsilon P(ds'|s, a) \\ &\leq \delta \mathcal{W}_1(\nu', \nu) \frac{2}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{E}^0} + |\log \eta(A)|)) |f|_{\mathfrak{E}^1} \int_S P(ds'|s, a) \\ &= \frac{2\delta}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{E}^0} + |\log \eta(A)|)) |f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu), \end{aligned}$$

where the first inequality follows from the fact that the map

$$\theta \mapsto \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu + \varepsilon(\nu' - \nu), \theta)(\delta_{s'})$$

is Lipschitz due to (66) and the definition of \mathcal{W}_1 (see (9)), and last equality follows from the fact that $P \in \mathcal{P}(S|S \times A)$. Hence, we obtain

$$|I_2| \leq 4 \left(\frac{\delta}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{E}^0} + |\log \eta(A)|)) + \tau \right) |f|_{\mathfrak{E}^0} |f|_{\mathfrak{E}^1} \mathcal{W}_1(\nu', \nu). \quad (70)$$

Putting together (67), (69) and (70) yields that

$$\begin{aligned}
& \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta)(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \right| \\
& \leq \left[\frac{4}{(1-\delta)^2} \left(\frac{\delta}{1-\delta} + 1 + \delta \right) \times \right. \\
& \quad \left. \times (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) + 4\tau \right] |f|_{\mathfrak{C}^0} |f|_{\mathfrak{C}^1} \mathcal{W}_1(\nu', \nu) \\
& = \left[\frac{4}{(1-\delta)^3} \left(-\left(\frac{1}{2} - \delta\right)^2 + \frac{5}{4} \right) \times \right. \\
& \quad \left. \times (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) + 4\tau \right] |f|_{\mathfrak{C}^0} |f|_{\mathfrak{C}^1} \mathcal{W}_1(\nu', \nu) \\
& \leq \left(\frac{5}{(1-\delta)^3} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) + 4\tau \right) |f|_{\mathfrak{C}^0} |f|_{\mathfrak{C}^1} \mathcal{W}_1(\nu', \nu).
\end{aligned} \tag{71}$$

for all $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\theta \in \mathbb{R}^d$. Combining (66) and (71) with (65) gives

$$\begin{aligned}
& \left| \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu', \theta')(\gamma) - \frac{\delta \hat{V}_\tau}{\delta \nu}(\nu, \theta)(\gamma) \right| \\
& \leq \frac{2|f|_{\mathfrak{C}^1}}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) |\theta' - \theta| \\
& \quad + \left(\frac{5}{(1-\delta)^3} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) + 4\tau \right) |f|_{\mathfrak{C}^0} |f|_{\mathfrak{C}^1} \mathcal{W}_1(\nu', \nu) \\
& \leq |f|_{\mathfrak{C}^1} \max \left\{ \frac{2}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)), \right. \\
& \quad \left. \left(\frac{5}{(1-\delta)^3} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) + 4\tau \right) |f|_{\mathfrak{C}^0} \right\} (|\theta' - \theta| + \mathcal{W}_1(\nu', \nu)) \\
& \leq |f|_{\mathfrak{C}^1} \left(\max \left\{ \frac{2}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)), \right. \right. \\
& \quad \left. \left. \frac{5}{(1-\delta)^3} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) |f|_{\mathfrak{C}^0} \right\} + 4\tau |f|_{\mathfrak{C}^0} \right) (|\theta' - \theta| + \mathcal{W}_1(\nu', \nu)) \\
& = |f|_{\mathfrak{C}^1} \left(\frac{1}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) \max \left\{ 2, \frac{5}{1-\delta} |f|_{\mathfrak{C}^0} \right\} + 4\tau |f|_{\mathfrak{C}^0} \right) \times \\
& \quad \times (|\theta' - \theta| + \mathcal{W}_1(\nu', \nu)),
\end{aligned}$$

for all $\nu', \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\theta', \theta \in \mathbb{R}^d$, where the last inequality follows from the standard estimate $\max\{a, c+d\} \leq d + \max\{a, c\}$. Hence, we set

$$L_F := |f|_{\mathfrak{C}^1} \left(\frac{1}{(1-\delta)^2} (|c|_{B_b(S \times A)} + \tau(2|f|_{\mathfrak{C}^0} + |\log \eta(A)|)) \max \left\{ 2, \frac{5}{1-\delta} |f|_{\mathfrak{C}^0} \right\} + 4\tau |f|_{\mathfrak{C}^0} \right).$$

□

Remark E.4. If A is a finite action space, a natural choice for η is the uniform distribution over A . In contrast, when A is continuous, η can be taken as a Gaussian or any other probability measure on A . Assuming $\eta \in \mathcal{P}(A)$, it follows that $|\log \eta(A)| = 0$. Note that the constants C_F and L_F in Proposition E.3 scale linearly with τ . Therefore, as indicated by condition (10), increasing τ

necessitates a corresponding increase in σ . This is intuitive, since the mapping $\nu \mapsto \text{KL}(\pi_\nu|\eta)$ is non-convex. As a result, choosing a large τ in (61) amplifies the non-convexity of $\hat{V}_\tau(\nu)(\gamma)$, requiring a larger σ to mitigate it.

Algorithm 1: Langevin-based Best Response for bandits

Input: cost function c , activation function f , potential U^ξ , reference measure η chosen according to Remark E.4, regularization parameter σ as prescribed by (10), where C_F, L_F are explicitly computed in Proposition E.3, τ , initial distribution of parameters ν_0 , outer and inner step sizes h_{out} and h_{in} , horizons T and K , learning rate $\alpha > 0$, and number of parameters N .

```

1 Generate i.i.d.  $(\theta_0^i)_{i=1}^N \sim \nu_0$  and set  $(\theta_{0,0}^i)_{i=1}^N := (\theta_0^i)_{i=1}^N$ ;
2 for  $t = 0, \dots, T - 1$  do
3   for  $k = 0, \dots, K - 1$  do
4     for  $i = 1, 2, \dots, N$  do
5        $\theta_{t,k+1}^i = \theta_{t,k}^i - h_{\text{in}}(\nabla_{\theta} \frac{\delta V_{\tau}^{\pi\nu}}{\delta \nu}(\nu_{t,k}^N, \theta_{t,k}^i) + \sigma \nabla_{\theta} U^\xi(\theta_{t,k}^i)) + \sqrt{2\sigma h_{\text{in}}} \mathcal{N}_{t,k}^i$ ;
6     end
7   end
8    $\nu_{t+1}^N = (1 - \alpha h_{\text{out}})\nu_{t+1}^N + \alpha h_{\text{out}} \Psi_{\sigma}[\nu_t^N]$ ;
9 end

```

Output: $\nu_T^N = \frac{1}{N} \sum_{i=1}^N \delta_{\theta_T^i}$.

E.2 Two-player zero-sum Markov game

We demonstrate how our results on min-max games can be applied to policy optimization in a two-player Markov game, thus extending the examples proposed in [21, Section 5] from convex-concave to non-convex-non-concave value function with softmax parametrized policies.

The setup is largely analogous to the MDP setting in Section 2.2. The goal of both players is to learn a policy that constitutes a Nash equilibrium of the game. For given strategies $\pi, \zeta : S \rightarrow \mathcal{P}^\eta(A)$, the (τ_1, τ_2) -entropy regularized total expected cumulative cost is given by

$$V_{\tau_1, \tau_2}^{\pi, \zeta}(s) := \mathbb{E}_s^{\pi, \zeta} \left[\sum_{n=0}^{\infty} \delta^n \left(c(s_n, a_n, b_n) + \tau_1 \log \frac{d\pi}{d\eta}(a_n|s_n) - \tau_2 \log \frac{d\zeta}{d\eta}(b_n|s_n) \right) \right].$$

For a fixed initial distribution $\gamma \in \mathcal{P}(S)$, the players' objective is then to solve

$$\min_{\pi \in \mathcal{P}^\eta(A|S)} \max_{\zeta \in \mathcal{P}^\eta(A|S)} V_{\tau_1, \tau_2}^{\pi, \zeta}(\gamma), \text{ with } V_{\tau_1, \tau_2}^{\pi, \zeta}(\gamma) := \int_S V_{\tau_1, \tau_2}^{\pi, \zeta}(s) \gamma(ds). \quad (72)$$

Assume that there exists a unique MNE $(\pi_{\tau_1}^*, \zeta_{\tau_2}^*)$ of the game (72) independent of γ , i.e., there exists $(\pi_{\tau_1}^*, \zeta_{\tau_2}^*)$ such that

$$\min_{\pi \in \mathcal{P}^\eta(A|S)} \max_{\zeta \in \mathcal{P}^\eta(A|S)} V_{\tau_1, \tau_2}^{\pi, \zeta}(\gamma) = \max_{\zeta \in \mathcal{P}^\eta(A|S)} \min_{\pi \in \mathcal{P}^\eta(A|S)} V_{\tau_1, \tau_2}^{\pi, \zeta}(\gamma) = V_{\tau_1, \tau_2}^{\pi_{\tau_1}^*, \zeta_{\tau_2}^*}(\gamma),$$

and it is given by

$$\begin{aligned} \pi_{\tau_1}^*(da|s) &\propto \exp \left(-\frac{1}{\tau_1} \int_A Q_{\tau_1, \tau_2}^{\pi_{\tau_1}^*, \zeta_{\tau_2}^*}(s, a) \zeta_{\tau_2}^*(db|s) \right) \eta(da), \\ \zeta_{\tau_2}^*(db|s) &\propto \exp \left(\frac{1}{\tau_2} \int_A Q_{\tau_1, \tau_2}^{\pi_{\tau_1}^*, \zeta_{\tau_2}^*}(s, a) \pi_{\tau_1}^*(da|s) \right) \eta(db). \end{aligned}$$

For example, such MNE exists for Markov games with finite state and action spaces [36].

To avoid requiring the players to search for the MNE $(\pi_{\tau_1}^*(\cdot|s), \zeta_{\tau_2}^*(\cdot|s))$ at each state $s \in S$ across the full space of probability measures, we assume that the strategies π_ν and ζ_μ are of the softmax form

$$\pi_\nu(da|s) \propto \exp(f_\nu(s, a)) \eta(da)$$

and

$$\zeta_\mu(db|s) \propto \exp(g_\mu(s, b))\eta(db),$$

where f_ν and g_μ are mean-field neural networks with activation functions $f, g : \mathbb{R}^d \times S \times A \rightarrow \mathbb{R}$, cf. equation (13). Thus, the goal of the players is now to find a global MNE of the game

$$\min_{\nu \in \mathcal{P}_1(\mathbb{R}^d)} \max_{\mu \in \mathcal{P}_1(\mathbb{R}^d)} V_{\tau_1, \tau_2}^{\pi_\nu, \zeta_\mu}(\gamma). \quad (73)$$

Owing to the presence of the normalization constant in (13), the value function $(\nu, \mu) \mapsto V_{\tau_1, \tau_2}^{\pi_\nu, \zeta_\mu}(\gamma)$ is non-convex-non-concave. Consequently, convergence of gradient flows to a global MNE cannot generally be expected without additional entropic regularization with respect to ν and μ (cf. (6)) as supported by our theoretical findings. In contrast, [21] considered the setting with non-parametric policies in (72), solving the problem by directly optimizing over the policies. This approach requires computing the MNE at each individual state, which can be computationally infeasible for large state spaces. By instead parametrizing the policies using the softmax form in (13) and optimizing over these parameters as in (73), the players are able to learn the MNE of (72) across all states simultaneously. Proposition E.5 states that $(\nu, \mu) \mapsto V_{\tau_1, \tau_2}^{\pi_\nu, \zeta_\mu}$ satisfies Assumption 3.1, and hence all our main results (Theorem B.2, Corollary B.3, Theorem B.4) and Theorem 3.3 apply to min-max problems (6) corresponding to energy functions $F^{\sigma_\nu, \sigma_\mu}(\nu, \mu) := V_{\tau_1, \tau_2}^{\pi_\nu, \zeta_\mu}(\gamma) + \sigma_\nu \text{KL}(\nu|\xi) - \sigma_\mu \text{KL}(\mu|\rho)$ for any reference measures ξ, ρ satisfying Assumption 3.2.

As in the previous subsection, we state that the value function V_{τ_1, τ_2} in the Markov games setting satisfies Assumption 3.1.

Proposition E.5. *Let S and A be Polish spaces, $c \in B_b(S \times A \times A)$, $\eta \in \mathcal{M}_+(A)$, and $\tau_1, \tau_2 > 0$. Then for all $\gamma \in \mathcal{P}(S)$ the function $F(\nu, \mu) := V_{\tau_1, \tau_2}^{\pi_\nu, \zeta_\mu}(\gamma)$ satisfies Assumption 3.1 with constants $C_F, L_F > 0$ depending on $\delta, \tau_1, |c|_{B_b(S \times A \times A)}, |f|_{\mathcal{C}^0}, |f|_{\mathcal{C}^1}, |\log \eta(A)|$, as in Proposition E.3, and constants $\bar{C}_F, \bar{L}_F > 0$ depending in an analogous way on $\delta, \tau_2, |c|_{B_b(S \times A \times A)}, |g|_{\mathcal{C}^0}, |g|_{\mathcal{C}^1}, |\log \eta(A)|$.*

Proof. The proof follows the same argument as the proof of Proposition E.3. \square

F Additional notation and definitions

Following [6, Definition 5.43], we recall the notion of differentiability on the space of probability measure that we utilize throughout the paper.

Definition F.1 (Flat differentiability on $\mathcal{P}_1(\mathbb{R}^d)$). *We say a function $F : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ is in \mathcal{C}^1 , if there exists a continuous function $\frac{\delta F}{\delta \mu} : \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$, with respect to the product topology on $\mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d$, called the flat derivative of F , for which there exists $\kappa > 0$ such that for all $(\mu, x) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d$, $\left| \frac{\delta F}{\delta \mu}(\mu, x) \right| \leq \kappa(1 + |x|)$, and for all $\mu' \in \mathcal{P}_1(\mathbb{R}^d)$,*

$$\lim_{\varepsilon \searrow 0} \frac{F(\mu^\varepsilon) - F(\mu)}{\varepsilon} = \int_{\mathbb{R}^d} \frac{\delta F}{\delta \mu}(\mu, x)(\mu' - \mu)(dx), \quad \text{with } \mu^\varepsilon = \mu + \varepsilon(\mu' - \mu), \quad (74)$$

and $\int_{\mathbb{R}^d} \frac{\delta F}{\delta \mu}(\mu, x)\mu(dx) = 0$.

Remark F.2. *One can show that if $F : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ admits a flat derivative $\frac{\delta F}{\delta \mu}$, then for all $\mu, \mu' \in \mathcal{P}_1(\mathbb{R}^d)$, the function $[0, 1] \ni \varepsilon \mapsto F(\mu^\varepsilon)$ is continuous on $[0, 1]$ and differentiable on $(0, 1)$ with derivative $\frac{d}{d\varepsilon} F(\mu^\varepsilon) = \int_{\mathbb{R}^d} \frac{\delta F}{\delta \mu}(\mu^\varepsilon, x)(\mu' - \mu)(dx)$ (see [19, Theorem 2.3]). Hence, by the fundamental theorem of calculus, $F(\mu') - F(\mu) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta F}{\delta \mu}(\mu^\varepsilon, x)(\mu' - \mu)(dx)d\varepsilon$, provided that $\varepsilon \mapsto \int \frac{\delta F}{\delta \mu}(\mu^\varepsilon, x)(\mu' - \mu)(dx)$ is integrable.*