

OSCILLATING SOLUTIONS TO THE MEAN-FIELD LANGEVIN DESCENT-ASCENT FLOW

JEAN-CHRISTOPHE MOURRAT  AND LOUCAS PILLAUD-VIVIEN

ABSTRACT. We present a counterexample to the statement of convergence of the mean-field Langevin descent-ascent flow on \mathbb{R}^2 . We consider payoff functions that are shaped as a double well in each coordinate, and for which the deterministic dynamics admits a limit cycle. When the coupling between the two coordinates is sufficiently strong and the entropic regularization sufficiently small, we show that the mean-field dynamics remains close to this cyclic behavior, and in particular, does not converge.

1. INTRODUCTION

A natural procedure for finding a saddle point of a function $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is to iteratively take small steps in the direction of $(-\nabla_x f, \nabla_y f)$. In the limit of infinitesimally small steps, we obtain a continuous dynamics called the gradient descent-ascent flow. It is well known that this gradient descent-ascent flow can fail to converge, even for simple functions such as $f(x, y) = xy$, for which the flow circles around the origin. A natural idea to remedy this is to add entropic regularization, which amounts to optimizing over probability measures rather than points, and to study the resulting mean-field Langevin descent-ascent (MFL-DA) flow. Given $\varepsilon > 0$ and a smooth function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, the MFL-DA flow is given by

$$(1.1) \quad \begin{cases} \partial_t \mu_t = \varepsilon \partial_x^2 \mu_t + \partial_x \left(\mu_t \partial_x \int f(x, y) d\nu_t(y) \right), \\ \partial_t \nu_t = \varepsilon \partial_y^2 \nu_t - \partial_y \left(\nu_t \partial_y \int f(x, y) d\mu_t(x) \right). \end{cases}$$

This is the Wasserstein gradient descent-ascent flow for the functional

$$\mathcal{F}(\mu, \nu) = \iint f d\mu d\nu + \varepsilon H(\mu) - \varepsilon H(\nu),$$

where $H(\mu) = \int \log\left(\frac{d\mu}{dx}\right) d\mu$ is the (negative) differential entropy. The laws (μ_t, ν_t) are the marginal laws of the processes (X_t, Y_t) solving

$$(1.2) \quad \begin{cases} dX_t = \left(- \int \partial_x f(X_t, y) d\nu_t(y) \right) dt + \sqrt{2\varepsilon} dB_t, \\ dY_t = \left(\int \partial_y f(x, Y_t) d\mu_t(x) \right) dt + \sqrt{2\varepsilon} dB'_t, \end{cases}$$

where B_t, B'_t are independent standard Brownian motions.

This framework was introduced and studied in [1], where it is shown that if the flow converges, then the limit must be the unique entropy-regularized mixed Nash equilibrium. Whether convergence actually holds was posed as an open problem in [12]. Part of the motivation for this question is that the entropic regularization endows the problem with strong structural properties: for instance, under mild assumptions, the partial equilibrium measures satisfy uniform log-Sobolev inequalities, which suggests that the regularization might tame the cycling behavior inherent to the deterministic dynamics. Several partial results support this intuition. Local convergence has been established independently in [11] and [13]. Global convergence is known to hold under two-timescale modifications of the flow [6, 7]; see also [5] for other modified dynamics. However, in a companion paper [8], one of us showed that in the finite-dimensional setting, gradient descent-ascent can fail to converge even under a two-sided Polyak–Łojasiewicz condition — the natural finite-dimensional analog of the uniform log-Sobolev assumption [9]. The question of global convergence in the mean-field setting, however, remained open.

The point of this paper is to show that the answer is negative: there exist functions for which the MFL-DA flow stays in a compact region and yet does not converge. More precisely, we consider functions indexed by $\alpha \geq 1$ of the form: for every $x, y \in \mathbb{R}$,

$$(1.3) \quad f_\alpha(x, y) := \alpha xy - \frac{x^2}{2} + \frac{x^4}{12} + \frac{y^2}{2} - \frac{y^4}{12}.$$

For this choice of the function f , the well-posedness of the system (1.1) follows from [3, 4]. As will be confirmed below (see Propositions 3.1 and 3.2), the quartic large-scale behavior of f ensures the confinement of the flow. Computing the partial derivatives: $\partial_x f_\alpha(x, y) = \alpha y - x + x^3/3$, $\partial_y f_\alpha(x, y) = \alpha x + y - y^3/3$ and writing $m_t := \int x d\mu_t(x) = \mathbb{E}[X_t]$ and $n_t := \int y d\nu_t(y) = \mathbb{E}[Y_t]$ for the means, the SDE system (1.2) becomes

$$(1.4) \quad \begin{cases} dX_t = (-\alpha n_t + X_t - X_t^3/3) dt + \sqrt{2\varepsilon} dB_t, \\ dY_t = (\alpha m_t + Y_t - Y_t^3/3) dt + \sqrt{2\varepsilon} dB'_t. \end{cases}$$

For this choice, the deterministic dynamics ($\varepsilon = 0$, Dirac initial condition) admits a limit cycle \mathcal{C}_α that circles around the origin and is contained in the annulus

$$(1.5) \quad \mathcal{A} := \{(m, n) \in \mathbb{R}^2 : 3 \leq m^2 + n^2 \leq 6\},$$

see Proposition 2.1. When the coupling parameter α is sufficiently large and the entropic regularization ε is sufficiently small, we show that the stochastic mean-field dynamics remains close to this limit cycle for all times, and in particular does not converge.

Theorem 1.1. *For every $\delta \in (0, 1]$, $\alpha < +\infty$ sufficiently large, and $\varepsilon > 0$ sufficiently small, the following holds. If (X_0, Y_0) is deterministic and belongs to \mathcal{C}_α , then for every $t_0 \geq 0$, there exists a parametrization $(\bar{m}_t, \bar{n}_t)_{t \in [t_0, t_0 + T_\alpha]}$*

of the curve \mathcal{C}_α such that for every $t \in [t_0, t_0 + T_\alpha]$, we have

$$\mathbb{E}[|X_t - \bar{m}_t|^2] + \mathbb{E}[|Y_t - \bar{n}_t|^2] \leq \delta.$$

In this circumstance, the means $(\mathbb{E}[X_t], \mathbb{E}[Y_t])$ therefore circle around the origin for all times while remaining away from it; in particular, the measures (μ_t, ν_t) do not converge as t tends to infinity.

We have stated Theorem 1.1 with the assumption that (X_0, Y_0) is deterministic and belongs to \mathcal{C}_α . This is only for the sake of clarity of exposition. The more general Theorem 3.3 below allows for random initial conditions, provided that they are sufficiently concentrated and with a mean sufficiently close to \mathcal{C}_α . Theorem 3.3 assumes a deterministic upper bound on $X_0^2 + Y_0^2$, but this itself can be weakened, as the proof will hopefully make clear.

The proof proceeds in two main steps. The first is a careful analysis of the deterministic dynamics, carried out in Section 2, for which we establish the existence of a unique limit cycle via the Poincaré–Bendixson theorem and construct a Lyapunov function certifying its exponential stability (Propositions 2.1 and 2.4). The second step, in Section 3, is an inductive argument (Theorem 3.3) showing that the stochastic fluctuations introduced by the noise remain small over each period of the cycle, so that they cannot accumulate and drive the system away from the periodic orbit. Finally, experiments illustrating the result are presented in Section 4.

2. ANALYSIS OF THE DETERMINISTIC DYNAMICS

In this section, we focus on the analysis of the deterministic dynamics, where we set $\varepsilon = 0$ and we impose the initial measures μ_0 and ν_0 to be Dirac masses. In this case, the evolution (1.4) becomes deterministic and reduces to the differential system

$$(2.1) \quad \partial_t m = -\alpha n + m - \frac{m^3}{3}, \quad \partial_t n = \alpha m + n - \frac{n^3}{3}.$$

We let $F_\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ denote the vector field driving this system of equations,

$$(2.2) \quad F_\alpha(m, n) := \left(-\alpha n + m - \frac{m^3}{3}, \alpha m + n - \frac{n^3}{3} \right).$$

We recall that the annulus \mathcal{A} is defined in (1.5).

Proposition 2.1 (limit cycle). *The system (2.1) admits a unique non-constant periodic orbit, which we denote by \mathcal{C}_α . This orbit circles around the origin and is contained in \mathcal{A} . It is a limit cycle of every trajectory of the system (2.1) except the one that stays put at the origin. The smallest period T_α of the motion along this cycle satisfies*

$$(2.3) \quad \frac{4\pi}{2\alpha + 1} \leq T_\alpha \leq \frac{4\pi}{2\alpha - 1}.$$

Proof. We set $r := \sqrt{m^2 + n^2}$. For (m, n) solving (2.1), we can write

$$(2.4) \quad \partial_t(r^2) = 2(m \partial_t m + n \partial_t n) = 2r^2 - \frac{2}{3}(m^4 + n^4).$$

Since $\frac{r^4}{2} \leq m^4 + n^4 \leq r^4$, we have

$$2r^2 \left(1 - \frac{r^2}{3}\right) \leq \partial_t(r^2) \leq r^2 \left(2 - \frac{r^2}{3}\right).$$

Hence, either $m(0) = n(0) = 0$ and the trajectory stays put at the origin, or $\liminf_{t \rightarrow +\infty} r^2 \geq 3$ and $\limsup_{t \rightarrow +\infty} r^2 \leq 6$. Since the trajectory either stays put at the origin or never visits it, we can define an angular variable θ such that $m = r \cos \theta$ and $n = r \sin \theta$, and we have

$$(2.5) \quad \partial_t \theta = \frac{m \partial_t n - n \partial_t m}{r^2} = \alpha + \frac{1}{3r^2} mn(m^2 - n^2) = \alpha + \frac{r^2}{12} \sin(4\theta),$$

where we used $mn = \frac{1}{2}r^2 \sin(2\theta)$ and $m^2 - n^2 = r^2 \cos(2\theta)$ to derive the last identity. Since $\alpha \geq 1$ and $r^2 \leq 6$ in the annulus \mathcal{A} , we see that $\partial_t \theta$ never vanishes in \mathcal{A} , and in particular the vector field F_α never vanishes in \mathcal{A} . (We can also note that the system (2.1) turns anticlockwise in \mathcal{A} .) Using this, the previous observation on the liminf and limsup of r^2 , and the Poincaré-Bendixson theorem [10, Theorem 3.7.1], we deduce that the system (2.1) admits a limit cycle in \mathcal{A} . The divergence of the vector field F_α is $\nabla \cdot F_\alpha(m, n) = 2 - r^2$, which is strictly negative everywhere in \mathcal{A} . It thus follows from [10, Theorem 3.4.2] that every limit cycle in \mathcal{A} is stable. If there were two distinct cycles in \mathcal{A} , then there would need to be an unstable cycle in-between (since the Poincaré return map [10, Section 3.4] cannot have two stable fixed points without having an unstable fixed point in-between), which we ruled out. Hence there is exactly one limit cycle for the system (2.1), which we denote by \mathcal{C}_α . The fact that every trajectory except the one staying put at the origin is attracted to \mathcal{C}_α is again a consequence of the Poincaré-Bendixson theorem [10, Theorem 3.7.1]. Recalling that $\partial_t \theta$ is bounded away from zero, we see that \mathcal{C}_α circles around the origin. It remains to see (2.3). Let (m_0, n_0) be a point on \mathcal{C}_α , let $(m_t, n_t)_{t \geq 0}$ be the solution to (2.1), with $(\theta_t, r_t)_{t \geq 0}$ the associated polar coordinates, and let $T_\alpha > 0$ be the minimal period of this motion (recall that F_α never vanishes along \mathcal{C}_α). By the Jordan curve theorem, the set $\mathbb{R}^2 \setminus \mathcal{C}_\alpha$ contains exactly two connected components. On the unbounded component, the winding number of \mathcal{C}_α is null. It jumps by ± 1 unit upon crossing \mathcal{C}_α , and the winding number of \mathcal{C}_α at the origin is $(\theta_{T_\alpha} - \theta_0)/(2\pi)$. By (2.5) and the fact that $\mathcal{C}_\alpha \subseteq \mathcal{A}$, we have that $|\partial_t \theta - \alpha| \leq \frac{1}{2}$, and since $\alpha \geq 1$, the winding number at the origin is 1 and

$$T_\alpha \left(\alpha - \frac{1}{2}\right) \leq \theta_{T_\alpha} - \theta_0 = 2\pi \leq T_\alpha \left(\alpha + \frac{1}{2}\right).$$

This yields (2.3) and thus completes the proof. \square

A key aspect we will exploit later is that T_α can be made arbitrarily small by sending α to infinity, while at the same time keeping the rough location of the

cycle \mathcal{C}_α unchanged, in the sense that it always stays within the fixed annulus \mathcal{A} . Although we will not need this, we mention that one can in fact describe the asymptotic shape of \mathcal{C}_α more precisely in the limit of large α . Indeed, notice from (2.3) that $\partial_t(r^2)$ remains bounded uniformly over $\alpha \geq 1$ and $(m, n) \in \mathcal{A}$. Hence, as α tends to infinity, the radius r along the cycle \mathcal{C}_α cannot change by more than C/α (since $T_\alpha \leq 4\pi/\alpha$). In other words, the cycle \mathcal{C}_α asymptotically becomes a circle as α tends to infinity. Moreover, we have that

$$0 = \int_0^{T_\alpha} \partial_t(r_t^2) dt = \int_0^{T_\alpha} \left(2r_t^2 - \frac{2r_t^4}{3} (\cos^4 \theta_t + \sin^4 \theta_t) \right) dt.$$

Since for large α , the radius r is close to a constant R_α and $\partial_t \theta$ is close to the constant α , and using also that \cos^4 and \sin^4 average to $\frac{3}{8}$, we get from the previous display that $R_\alpha^2 \simeq R_\alpha^4/4$, that is, $R_\alpha \simeq 2$. In other words, the cycle \mathcal{C}_α tends to the circle of radius 2 centered at the origin as α tends to infinity.

Using symmetry, we record information on the first and second moments of m and n along the cycle \mathcal{C}_α .

Proposition 2.2 (square average). *Consider the system (2.1) initialized at a point in the cycle \mathcal{C}_α , and recall that we denote by T_α the minimal period of the resulting trajectory. We have*

$$(2.6) \quad \frac{1}{T_\alpha} \int_0^{T_\alpha} m_t dt = \frac{1}{T_\alpha} \int_0^{T_\alpha} n_t dt = 0$$

and

$$(2.7) \quad \frac{1}{T_\alpha} \int_0^{T_\alpha} m_t^2 dt = \frac{1}{T_\alpha} \int_0^{T_\alpha} n_t^2 dt \geq \frac{3}{2}.$$

Proof. We first observe that the function $t \mapsto (-n_t, m_t)$ is also a solution to (2.1). By uniqueness of the cycle \mathcal{C}_α , we deduce (2.6) and the equality in (2.7). Since the cycle \mathcal{C}_α is contained in the annulus \mathcal{A} , we have $m^2 + n^2 \geq 3$ everywhere in \mathcal{C}_α , and this completes the proof of the proposition. \square

Remark 2.3. As per our previous discussion on the fact that \mathcal{C}_α tends to the circle of radius 2 as α tends to infinity, one could show that $\frac{1}{T_\alpha} \int_0^{T_\alpha} m_t^2 dt$ and $\frac{1}{T_\alpha} \int_0^{T_\alpha} n_t^2 dt$ tend to 2 as α tends to infinity.

We can strengthen the statement of stability of the cycle \mathcal{C}_α as follows. Recall the notation for the vector field F_α in (2.2).

Proposition 2.4 (Lyapunov function). *Let $g(z) := \text{dist}^2(z, \mathcal{C}_\alpha)$ denote the squared distance between $z \in \mathbb{R}^2$ and the limit cycle \mathcal{C}_α . There exist $\delta, c > 0$ such that for every α sufficiently large and $z \in \mathbb{R}^2$, we have*

$$g(z) \leq \delta \quad \implies \quad (F_\alpha \cdot \nabla g)(z) \leq -cg(z).$$

Proof. Since \mathcal{C}_α is a smooth simple closed curve, we can find a neighborhood U of \mathcal{C}_α in which the nearest-point projection $\pi : U \rightarrow \mathcal{C}_\alpha$ is well-defined and

smooth, so that for every $z \in U$, we have

$$g(z) = \inf_{z' \in \mathcal{C}_\alpha} |z - z'|^2 = |z - \pi(z)|^2.$$

The optimality condition yields that $z - \pi(z)$ is orthogonal to the tangent to \mathcal{C}_α at $\pi(z)$, which is oriented along $F_\alpha(\pi(z))$, that is, we have

$$(2.8) \quad (z - \pi(z)) \cdot F_\alpha(\pi(z)) = 0.$$

By the envelope theorem [2, Theorem 2.21], we also see that

$$(2.9) \quad \nabla g(z) = 2(z - \pi(z)),$$

and thus

$$F_\alpha \cdot \nabla g(z) = 2(z - \pi(z)) \cdot F_\alpha(z).$$

Using (2.8), we can rewrite this as

$$\begin{aligned} F_\alpha \cdot \nabla g(z) &= 2(z - \pi(z)) \cdot (F_\alpha(z) - F_\alpha(\pi(z))) \\ &= 2(z - \pi(z)) \cdot \nabla F_\alpha(\pi(z))(z - \pi(z)) + O(|z - \pi(z)|^3), \end{aligned}$$

where $\nabla F_\alpha(u)$ denotes the Jacobian of F_α at $u \in \mathbb{R}^2$, and where we used that the second derivatives of F_α do not depend on α . Denoting $\nu(z) := \frac{z - \pi(z)}{|z - \pi(z)|}$ for a unit normal to \mathcal{C}_α at $\pi(z)$, we can rewrite this as

$$(2.10) \quad F_\alpha \cdot \nabla g(z) = g(z) \left[2\nu(z) \cdot \nabla F_\alpha(\pi(z))\nu(z) + O(\sqrt{g(z)}) \right].$$

Recalling that $\nu(z)$ is orthogonal to $F_\alpha(\pi(z)) = (F_{\alpha,1}(\pi(z)), F_{\alpha,2}(\pi(z)))$, and that the choice of orientation of $\nu(z)$ is immaterial, we can take $\nu(z) = (F_{\alpha,2}(\pi(z)), -F_{\alpha,1}(\pi(z)))/|F_\alpha|$ (recall from Proposition 2.1 that F_α only vanishes at the origin). In view of (2.10), the proof of Proposition 2.4 will be complete provided that we can show that, for α sufficiently large,

$$(2.11) \quad \sup_{z \in \mathcal{A}} \begin{pmatrix} F_{\alpha,2}(z) \\ -F_{\alpha,1}(z) \end{pmatrix} \cdot \nabla F_\alpha(z) \begin{pmatrix} F_{\alpha,2}(z) \\ -F_{\alpha,1}(z) \end{pmatrix} < 0.$$

We have

$$\nabla F_\alpha(m, n) = \begin{pmatrix} 1 - m^2 & -\alpha \\ \alpha & 1 - n^2 \end{pmatrix}.$$

Noting that the statement (2.11) only concerns the symmetric part of ∇F_α , we can rewrite it as

$$(2.12) \quad \inf \{ (m^2 - 1)F_{\alpha,2}(m, n)^2 + (n^2 - 1)F_{\alpha,1}(m, n)^2 : m^2 + n^2 \in [3, 6] \} > 0.$$

We now proceed to prove (2.12). We have

$$(2.13) \quad \begin{aligned} (m^2 - 1)F_{\alpha,2}(m, n)^2 + (n^2 - 1)F_{\alpha,1}(m, n)^2 \\ = \alpha^2(m^4 + n^4 - m^2 - n^2) + R_\alpha(m, n), \end{aligned}$$

where

$$R_\alpha(m, n) := \frac{4\alpha}{3}mn(m^2 - n^2) + (m^2 - 1)\left(n - \frac{n^3}{3}\right)^2 + (n^2 - 1)\left(m - \frac{m^3}{3}\right)^2.$$

We can find a constant $C < +\infty$ such that for every (m, n) with $m^2 + n^2 \leq 6$, we have $|R_\alpha(m, n)| \leq C\alpha$. We now bound the leading-order term from below. Since $m^4 + n^4 \geq \frac{1}{2}(m^2 + n^2)^2$, we have that

$$m^4 + n^4 - m^2 - n^2 \geq \frac{1}{2}r^4 - r^2 = \frac{1}{2}r^2(r^2 - 2).$$

Since $r^2 \geq 3$ in the annulus \mathcal{A} , this quantity is bounded from below by $3/2$. Returning to (2.13), we thus obtain that for every $(m, n) \in \mathcal{A}$,

$$(m^2 - 1)F_{\alpha,2}(m, n)^2 + (n^2 - 1)F_{\alpha,1}(m, n)^2 \geq \frac{3\alpha^2}{2} - C\alpha.$$

For α sufficiently large, the right-hand side is strictly positive. This completes the proof of (2.12), and therefore of the proposition. \square

3. MEAN-FIELD DYNAMICS

We now come back to the study of the mean-field SDE in (1.4), with the notation $m_t = \mathbb{E}[X_t]$ and $n_t = \mathbb{E}[Y_t]$; in other words we now consider the case of $\varepsilon > 0$. An important ingredient of our analysis, extending the fact that the deterministic dynamics studied in the previous section remains in a fixed bounded set uniformly over $\alpha \geq 1$, is that (X_t, Y_t) satisfy moment bounds that are uniform in α . We start with estimates on the second moments.

Proposition 3.1 (Second moment bound). *For every $\alpha \geq 0$, $\varepsilon \in (0, 1]$ and $t \geq 0$, we have*

$$(3.1) \quad \mathbb{E}[X_t^2 + Y_t^2] \leq \max(8, \mathbb{E}[X_0^2 + Y_0^2]).$$

In particular, if $\mathbb{E}[X_0^2 + Y_0^2] \leq 8$, then $|m_t|^2 + |n_t|^2 \leq 8$ for every $t \geq 0$.

Proof. By Itô's formula, we have

$$\frac{1}{2}\partial_t \mathbb{E}[X_t^2] = -\alpha m_t n_t + \mathbb{E}[X_t^2] - \frac{\mathbb{E}[X_t^4]}{3} + \varepsilon,$$

$$\frac{1}{2}\partial_t \mathbb{E}[Y_t^2] = \alpha m_t n_t + \mathbb{E}[Y_t^2] - \frac{\mathbb{E}[Y_t^4]}{3} + \varepsilon.$$

Denoting $\varphi(t) := \mathbb{E}[X_t^2] + \mathbb{E}[Y_t^2]$, and using that

$$\mathbb{E}[X_t^4] + \mathbb{E}[Y_t^4] \geq \mathbb{E}[X_t^2]^2 + \mathbb{E}[Y_t^2]^2 \geq \frac{1}{2}\varphi(t)^2,$$

we obtain that

$$\frac{1}{2}\partial_t \varphi(t) \leq \varphi(t) - \frac{1}{6}\varphi(t)^2 + 2\varepsilon.$$

The quantity on the right side is negative whenever $\varphi(t) \geq 8$, so (3.1) follows by a comparison argument. The last part of the proposition statement is obtained using the Cauchy-Schwarz inequality. \square

We define

$$(3.2) \quad \tilde{X}_t := X_t - m_t, \quad \tilde{Y}_t := Y_t - n_t.$$

Noting that

$$(3.3) \quad \partial_t m_t = -\alpha n_t + m_t - \frac{\mathbb{E}[X_t^3]}{3}, \quad \partial_t n_t = \alpha m_t + n_t - \frac{\mathbb{E}[Y_t^3]}{3},$$

we obtain that

$$(3.4) \quad d\tilde{X}_t = \left[(1 - m_t^2)\tilde{X}_t - m_t(\tilde{X}_t^2 - \mathbb{E}[\tilde{X}_t^2]) - \frac{1}{3}(\tilde{X}_t^3 - \mathbb{E}[\tilde{X}_t^3]) \right] dt + \sqrt{2\varepsilon} dB_t,$$

and similarly for \tilde{Y} .

Proposition 3.2 (Moment bounds). *For every even integer $k \geq 2$ and provided that $\mathbb{E}[X_0^k + Y_0^k]$ is finite, there exists a constant C_k depending only on k and $\mathbb{E}[X_0^k + Y_0^k]$ such that for every $\alpha \geq 0$, $\varepsilon \in (0, 1]$, and $t \geq 0$, we have*

$$\mathbb{E}[\tilde{X}_t^k + \tilde{Y}_t^k] \leq C_k.$$

Proof. By Itô's formula, we have for every integer $k \geq 2$ that

$$(3.5) \quad \begin{aligned} \frac{1}{k} \partial_t \mathbb{E}[\tilde{X}_t^k] &= \mathbb{E} \left[\tilde{X}_t^{k-1} \left((1 - m_t^2)\tilde{X}_t - m_t(\tilde{X}_t^2 - \mathbb{E}[\tilde{X}_t^2]) - \frac{1}{3}(\tilde{X}_t^3 - \mathbb{E}[\tilde{X}_t^3]) \right) \right] \\ &\quad + \varepsilon(k-1)\mathbb{E}[\tilde{X}_t^{k-2}]. \end{aligned}$$

We start with the analysis of the case $k = 4$. Noting that $\frac{1}{3} = \frac{1/2}{2} + \frac{1/2}{6}$, we have by Hölder's inequality that

$$(3.6) \quad \mathbb{E}[|\tilde{X}_t|^3] \leq \mathbb{E}[\tilde{X}_t^2]^{\frac{3}{4}} \mathbb{E}[\tilde{X}_t^6]^{\frac{1}{4}}.$$

By Proposition 3.1, we have that

$$\mathbb{E}[\tilde{X}_t^2] \leq \mathbb{E}[X_t^2] \leq C, \quad |m_t|^2 + |n_t|^2 \leq C,$$

where the constant $C < +\infty$ is allowed to depend on an upper bound on $\mathbb{E}[X_0^2 + Y_0^2]$. We thus find that, possibly after enlarging the constant C ,

$$\frac{1}{4} \partial_t \mathbb{E}[\tilde{X}_t^4] \leq \mathbb{E}[\tilde{X}_t^4] + C \left(\mathbb{E}[|\tilde{X}_t|^5] + \mathbb{E}[|\tilde{X}_t|^3] + \mathbb{E}[\tilde{X}_t^6]^{\frac{1}{2}} + \varepsilon \right) - \frac{1}{3} \mathbb{E}[\tilde{X}_t^6].$$

For some constant $C' < +\infty$, we have for every $x \in \mathbb{R}$ that $C|x|^5 \leq \frac{x^6}{9} + C'$ and $C\sqrt{|x|} \leq \frac{|x|}{9} + C'$, hence

$$\frac{1}{4} \partial_t \mathbb{E}[\tilde{X}_t^4] \leq \mathbb{E}[\tilde{X}_t^4] + C \mathbb{E}[|\tilde{X}_t|^3] - \frac{1}{9} \mathbb{E}[\tilde{X}_t^6] + 2CC' + C\varepsilon.$$

Denoting $\varphi(t) := \mathbb{E}[\tilde{X}_t^4]$, using Jensen's inequality and the assumption $\varepsilon \leq 1$, and reabsorbing everything into a larger constant $C < +\infty$, we have that

$$\frac{1}{4} \partial_t \varphi(t) \leq \varphi(t) + C\varphi(t)^{\frac{3}{4}} - \frac{1}{9}\varphi(t)^{\frac{3}{2}} + C.$$

The right-hand side is negative whenever $\varphi(t)$ is sufficiently large, so we obtain the result for the fourth moment of \tilde{X}_t .

Now that $|\mathbb{E}[\tilde{X}_t^3]|$ is controlled, a similar argument allows one to bound the moments of \tilde{X}_t order k for every even integer k from (3.5). The argument for controlling the moments of \tilde{Y}_t is identical. \square

For convenience, we use the notation $p_t := (m_t, n_t)$. We also recall the notation T_α from Proposition 2.1 for the minimal period of the deterministic dynamics (2.1) along the cycle. Our main result Theorem 1.1 is a direct consequence of the following inductive structure.

Theorem 3.3 (Induction). *There exists a constant $C < +\infty$ such that the following holds for every $k > \frac{3}{2}$, $\alpha < +\infty$ sufficiently large, $\varepsilon > 0$ sufficiently small, initialization X_0, Y_0 with $X_0^2 + Y_0^2 \leq 8$, and $t_0 \geq 0$. If*

$$(3.7) \quad \mathbb{E}[\tilde{X}_{t_0}^2] \leq \alpha^{-k}, \quad \mathbb{E}[\tilde{Y}_{t_0}^2] \leq \alpha^{-k}, \quad \text{dist}(p_{t_0}, \mathcal{C}_\alpha) \leq \alpha^{1-k},$$

then the same inequalities are valid at time $t_0 + T_\alpha$, that is,

$$(3.8) \quad \mathbb{E}[\tilde{X}_{t_0+T_\alpha}^2] \leq \alpha^{-k}, \quad \mathbb{E}[\tilde{Y}_{t_0+T_\alpha}^2] \leq \alpha^{-k}, \quad \text{dist}(p_{t_0+T_\alpha}, \mathcal{C}_\alpha) \leq \alpha^{1-k},$$

and moreover, letting \bar{p}_{t_0} denote the point on \mathcal{C}_α that is closest to p_{t_0} and $(\bar{p}_t)_{t \geq t_0}$ denote the solution to the deterministic system in (2.1), we have for every $t \in [t_0, t_0 + T_\alpha]$ that

$$(3.9) \quad |p_t - \bar{p}_t| \leq C\alpha^{1-k}.$$

Proof. We fix $k > \frac{3}{2}$. Under our assumption that the initial condition must satisfy $X_0^2 + Y_0^2 \leq 8$, we can bound all moments of X_t and Y_t uniformly over $\alpha \geq 0$, $\varepsilon \in (0, 1]$, and $t \geq 0$. In particular, we have $|p_t|^2 = |m_t|^2 + |n_t|^2 \leq 8$ for all $t \geq 0$. From now on, we assume that (3.7) holds.

Step 1. In this step, we show that for α sufficiently large and $\varepsilon > 0$ sufficiently small, we have for every $t \in [t_0, t_0 + T_\alpha]$ that $\mathbb{E}[\tilde{X}_t^2] \leq 2\alpha^{-k}$ and $\mathbb{E}[\tilde{Y}_t^2] \leq 2\alpha^{-k}$. Using that $|m_t|^2 \leq 8$ and the Itô identity (3.5), we have that

$$\frac{1}{2} \partial_t \mathbb{E}[\tilde{X}_t^2] \leq \mathbb{E}[\tilde{X}_t^2] + 3\mathbb{E}[|\tilde{X}_t|^3] + \varepsilon.$$

Since by Proposition 3.2 all moments of \tilde{X}_t are bounded, we can appeal to Hölder's inequality to find a constant $C < +\infty$ such that $\mathbb{E}[|\tilde{X}_t|^3] \leq C\mathbb{E}[\tilde{X}_t^2]^{1-\frac{1}{2k}}$, so that

$$(3.10) \quad \frac{1}{2} \partial_t \mathbb{E}[\tilde{X}_t^2] \leq \mathbb{E}[\tilde{X}_t^2] + C\mathbb{E}[\tilde{X}_t^2]^{1-\frac{1}{2k}} + \varepsilon.$$

Denoting $T := \sup\{t \in [t_0, t_0 + T_\alpha] : \mathbb{E}[\tilde{X}_t^2] \leq 2\alpha^{-k}\}$, using that $T_\alpha \leq 4\pi/\alpha$ and redefining $C < +\infty$ as necessary, we have for every $t \in [t_0, T]$ that

$$\mathbb{E}[\tilde{X}_t^2] \leq \alpha^{-k} + C\alpha^{-1} \left(\alpha^{-k} + \alpha^{\frac{1}{2}-k} + \varepsilon \right).$$

For α sufficiently large and $\varepsilon > 0$ sufficiently small, we can make sure that the right-hand side above is smaller than $2\alpha^{-k}$, and therefore that $T = t_0 + T_\alpha$ as desired. The argument for $\mathbb{E}[\tilde{Y}_t^2]$ is identical.

Step 2. In this step, we show that provided that α is sufficiently large, we have for every $t \in [t_0, t_0 + T_\alpha]$ that

$$(3.11) \quad \text{dist}(p_t, \mathcal{C}_\alpha) \leq \alpha^{1-k}.$$

Recall the notation F_α in (2.2), and the identity (3.3) which we can rewrite as

$$(3.12) \quad \partial_t p_t = F_\alpha(p_t) + R_t,$$

where we have set

$$\begin{aligned} R_t &:= \frac{1}{3}(m_t^3 - \mathbb{E}[X_t^3], n_t^3 - \mathbb{E}[Y_t^3]) \\ &= \left(-m_t \mathbb{E}[\tilde{X}_t^2] - \frac{1}{3} \mathbb{E}[\tilde{X}_t^3], -n_t \mathbb{E}[\tilde{Y}_t^2] - \frac{1}{3} \mathbb{E}[\tilde{Y}_t^3] \right). \end{aligned}$$

Recalling that \tilde{X}_t has bounded moments of every order, and using the result of the previous step, we can argue as in the sentence above (3.10) to find a constant $C < +\infty$ such that for every $t \in [t_0, t_0 + T_\alpha]$,

$$(3.13) \quad |R_t| \leq C\alpha^{\frac{1}{2}-k}.$$

Recall that we write $g(z) = \text{dist}^2(z, \mathcal{C}_\alpha)$. We have

$$\partial_t(g(p_t)) = (F_\alpha \cdot \nabla g)(p_t) + R_t \cdot \nabla g(p_t).$$

Let $\delta, c > 0$ be as given by Proposition 2.4, and let $T := \sup\{t \in [t_0, t_0 + T_\alpha] : g(p_t) \leq \delta\}$. For every $t \in [t_0, T]$, we can appeal to this proposition to bound the first term in the previous display by $-cg(p_t)$. Using that $|\nabla g(z)| = 2\sqrt{g(z)}$, the Cauchy-Schwarz and Young's inequalities, we have

$$|R_t \cdot \nabla g(p_t)| \leq 2\sqrt{g(p_t)}|R_t| \leq \frac{c}{2}g(p_t) + \frac{2|R_t|^2}{c}.$$

Combining these and enlarging the constant $C < +\infty$, we obtain that for every $t \in [t_0, T]$,

$$\partial_t(g(p_t)) \leq -\frac{c}{2}g(p_t) + C\alpha^{1-2k}.$$

Integrating this (and updating C) yields that for every $t \leq T$,

$$g(p_t) \leq e^{-\frac{c}{2}t} \alpha^{2-2k} + C\alpha^{1-2k}(1 - e^{-\frac{c}{2}t}).$$

By taking $\alpha \geq C$, we ensure that the right-hand side of the inequality above remains below α^{2-2k} . In particular, since $k > 1$, we have $T = t_0 + T_\alpha$ and the announced inequality (3.11) is proved.

Step 3. For α sufficiently large, the point in \mathcal{C}_α that is closest to p_{t_0} is well-defined; we denote it by $\bar{p}_{t_0} = (\bar{m}_{t_0}, \bar{n}_{t_0})$, and let $\bar{p}_t = (\bar{m}_t, \bar{n}_t)_{t \geq t_0}$ denote the solution to the system (2.1) with this initialization. In this step, we show that for a sufficiently large constant $C < +\infty$, the inequality (3.9) is valid. We denote the L^∞ norm of ∇F_α on the ball of radius $\sqrt{8}$ by L_α . Recalling that $(p_t)_{t \geq t_0}$ and $(\bar{p}_t)_{t \geq t_0}$ both take values in this ball, as well as (3.12) and (3.13), we have for almost every $t \in [t_0, t_0 + T_\alpha]$ that

$$\partial_t (|p_t - \bar{p}_t|) \leq L_\alpha |p_t - \bar{p}_t| + C\alpha^{\frac{1}{2}-k},$$

with

$$|p_0 - \bar{p}_0| \leq \alpha^{1-k}.$$

Integrating this, we find that for every $t \in [t_0, t_0 + T_\alpha]$,

$$|p_t - \bar{p}_t| \leq e^{L_\alpha T_\alpha} \left(\alpha^{1-k} + C\alpha^{\frac{1}{2}-k} \right).$$

The conclusion follows by noting that $L_\alpha \leq C\alpha$ while $T_\alpha \leq 4\pi/\alpha$.

Step 4. We now show that $\mathbb{E}[\tilde{X}_{t_0+T_\alpha}^2] \leq \alpha^{-k}$ and $\mathbb{E}[\tilde{Y}_{t_0+T_\alpha}^2] \leq \alpha^{-k}$. As preparation for this, we deduce from the previous step and Proposition 2.2 that

$$(3.14) \quad \left| \frac{1}{T_\alpha} \int_{t_0}^{t_0+T_\alpha} m_t dt \right| \leq C\alpha^{1-k}, \quad \frac{1}{T_\alpha} \int_{t_0}^{t_0+T_\alpha} m_t^2 dt \geq \frac{5}{4}.$$

Using (3.5) with $k = 2$, we have

$$\frac{1}{2} \partial_t \mathbb{E}[\tilde{X}_t^2] = (1 - m_t^2) \mathbb{E}[\tilde{X}_t^2] - m_t \mathbb{E}[\tilde{X}_t^3] - \frac{1}{3} \mathbb{E}[\tilde{X}_t^4] + \varepsilon.$$

We recall from (3.10) that $\partial_t \mathbb{E}[\tilde{X}_t^2] \leq C\alpha^{\frac{1}{2}-k} + \varepsilon$, and a small modification of the argument leading to this bound, starting from the previous display, allows one to bound the absolute value of $\partial_t \mathbb{E}[\tilde{X}_t^2]$ by the same expression. For $\varepsilon > 0$ sufficiently small, using again that $T_\alpha \leq 4\pi/\alpha$, we thus have that

$$\int_{t_0}^{t_0+T_\alpha} (1 - m_t^2) (\mathbb{E}[\tilde{X}_t^2] - \mathbb{E}[\tilde{X}_{t_0}^2]) dt \leq C\alpha^{-\frac{3}{2}-k}.$$

Using (3.5) with $k = 3$ and similar arguments as those leading to the bound on $|\partial_t \mathbb{E}[\tilde{X}_t^2]|$, one can show that $|\partial_t \mathbb{E}[\tilde{X}_t^3]| \leq C\alpha^{\frac{1}{2}-k}$, and therefore

$$\int_{t_0}^{t_0+T_\alpha} m_t (\mathbb{E}[\tilde{X}_{t_0}^3] - \mathbb{E}[\tilde{X}_t^3]) dt \leq C\alpha^{-\frac{3}{2}-k}.$$

Combining the last three displays (and using again that we take $\varepsilon > 0$ sufficiently small and adjusting C), we obtain that

$$\begin{aligned} \mathbb{E}[\tilde{X}_{t_0+T_\alpha}^2] &\leq \mathbb{E}[\tilde{X}_{t_0}^2] \left(1 + 2 \int_{t_0}^{t_0+T_\alpha} (1 - m_t^2) dt \right) \\ &\quad - 2\mathbb{E}[\tilde{X}_{t_0}^3] \int_{t_0}^{t_0+T_\alpha} m_t dt + C\alpha^{-\frac{3}{2}-k}. \end{aligned}$$

Using again that $|\mathbb{E}[\tilde{X}_{t_0}^3]| \leq C\alpha^{\frac{1}{2}-k}$ and (3.14) yields that

$$\mathbb{E}[\tilde{X}_{t_0+T_\alpha}^2] \leq \mathbb{E}[\tilde{X}_{t_0}^2] \left(1 - \frac{T_\alpha}{2} \right) + C\alpha^{\frac{1}{2}-2k} + C\alpha^{-\frac{3}{2}-k}.$$

Recalling that $\mathbb{E}[\tilde{X}_{t_0}^2] \leq \alpha^{-k}$, that $T_\alpha \geq 2\pi/\alpha$, and that $k > \frac{3}{2}$, we obtain the desired result by taking α sufficiently large. \square

4. NUMERICAL EXPERIMENTS

In this section, we display some numerical experiments showing the behavior described in the main theorems. For this we compute a particle approximation of the minimax SDE that we study, that is, we take particles $(X_i, Y_i)_{1 \leq i \leq N}$ and solve

$$(4.1) \quad dX_t^i = \left(-\frac{\alpha}{N} \sum_{j=1}^N Y_t^j + X_t^i - \frac{(X_t^i)^3}{3} \right) dt + \sqrt{2\varepsilon} dB_t,$$

$$(4.2) \quad dY_t^i = \left(\frac{\alpha}{N} \sum_{j=1}^N X_t^j + Y_t^i - \frac{(Y_t^i)^3}{3} \right) dt + \sqrt{2\varepsilon} dB'_t,$$

via a standard Euler-Maruyama discretization scheme. In the following we took the parameters $N = 500$, $\alpha = 1.5$ and i.i.d. initializations taken as $X_0^i \sim \mathcal{N}(-0.2, 0.25)$ and $Y_0^i \sim \mathcal{N}(0.4, 0.25)$. We simulate until $T = 20$ to observe a steady behavior. We show this for two different diffusion coefficients: $\varepsilon = 0.25$ and $\varepsilon = 0.5$. In the first case the system appears to stabilize on a periodic motion, whereas in the second case, the system seems to converge to a stationary measure which should be, up to particle approximation, the one depicted in [12, 11].

We show in Figure 1 the empirical mean and variance of the particles as a function of time t . We observe that in the first case, for which $\varepsilon = 0.25$, these quantities stabilize to seemingly periodic functions, whereas in the second case, they converge to constants.

In Figure 2, we show snapshots of the particles at different times for the two different values of ε . We observe that in the first case ($\varepsilon = 0.25$), the particles fluctuate around a periodic cycle, which is close to the limit cycle of the deterministic system (depicted as a black line). On the other hand, in the second case the particles are spread out and seem to be samples of a stationary measure, which resembles a mixture of four Gaussians. Animations of the particle system are available at https://thebiglouloup.github.io/loucaspillaudvivien/minimax_animations.html.

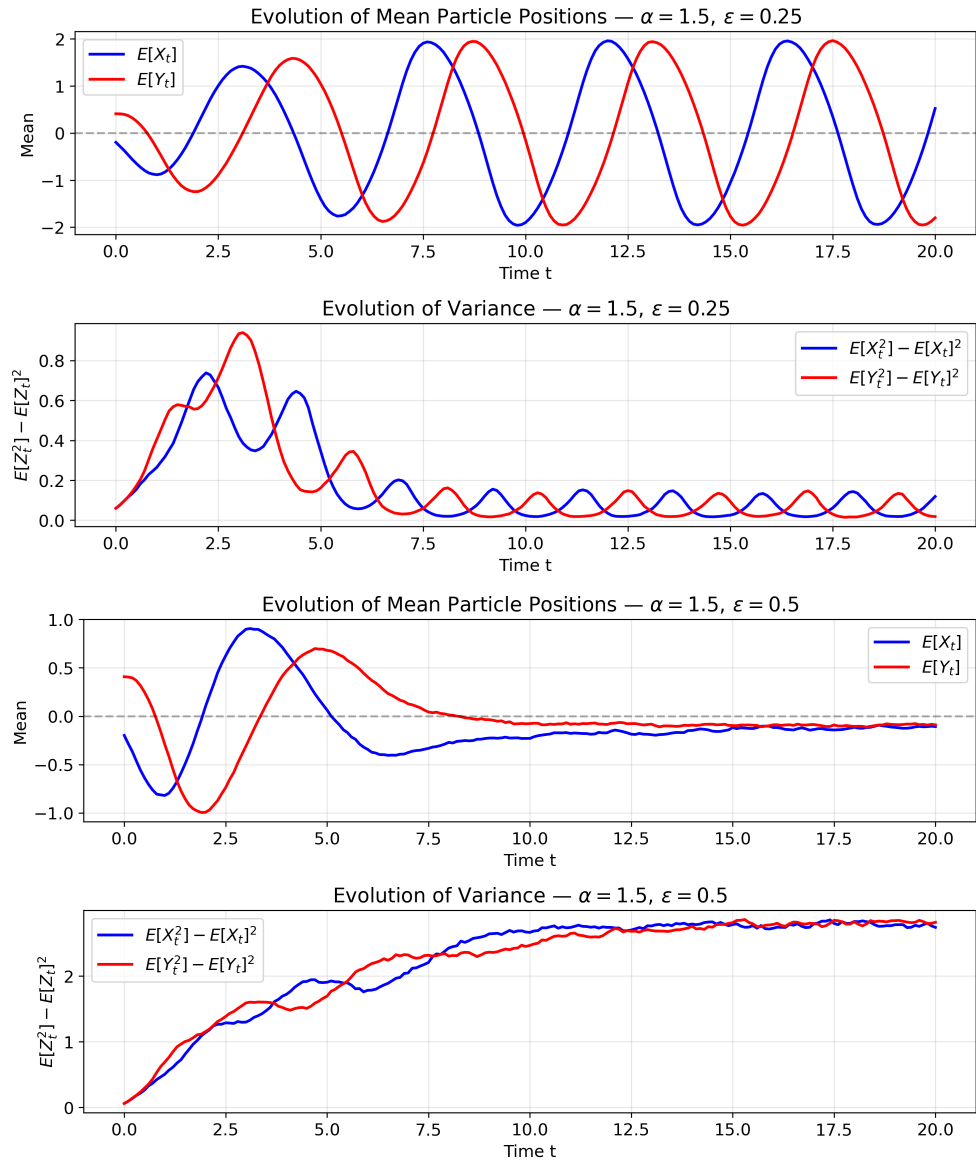


FIGURE 1. Mean and variance of the particles as a function of time for $\epsilon = 0.25$ (two top plots) and $\epsilon = 0.5$ (two bottom plots).

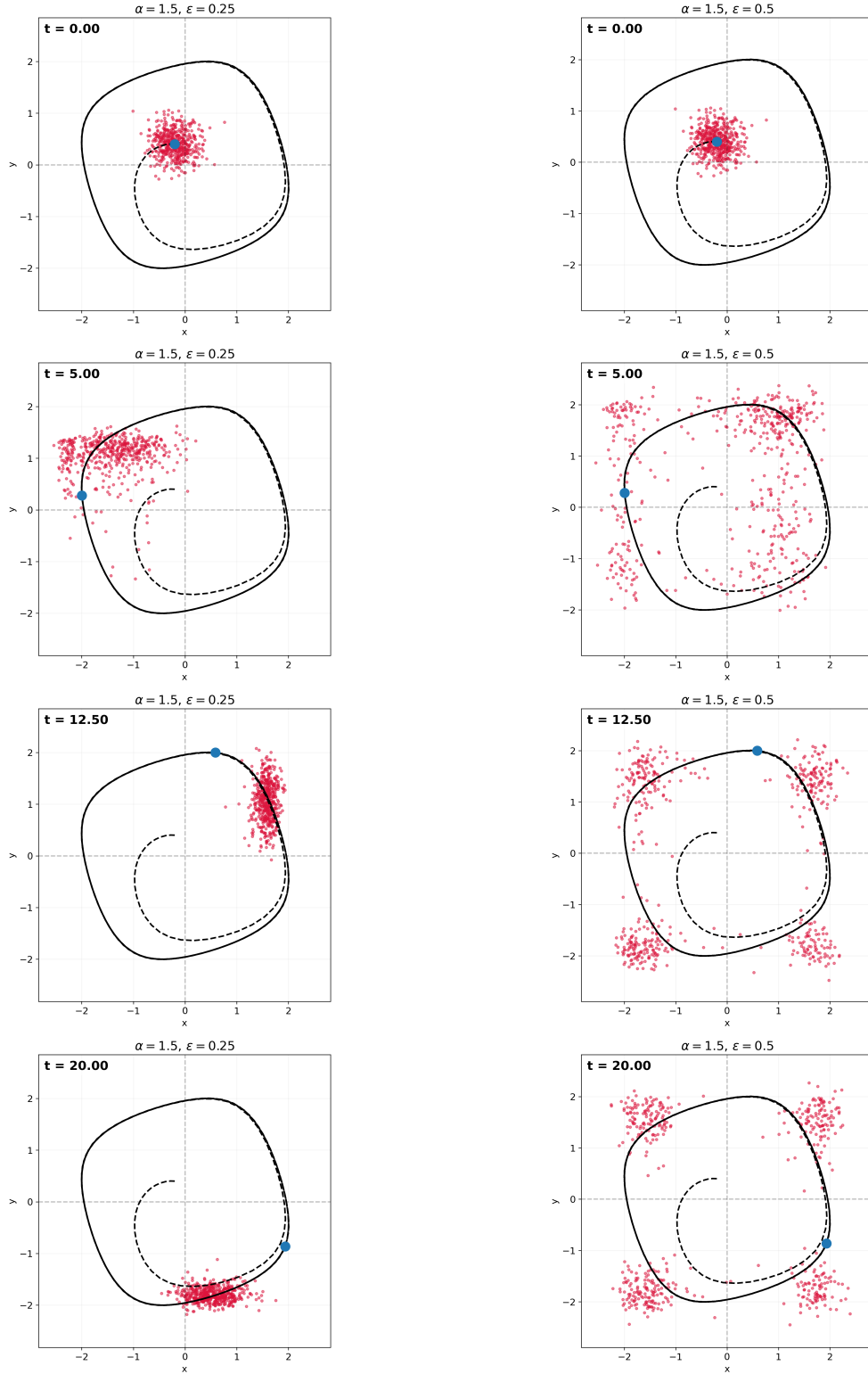


FIGURE 2. Particle snapshots for $\varepsilon = 0.25$ (left) and $\varepsilon = 0.5$ (right) at times $t = 0, 5, 12.5, 20$. The red points represent the positions of the $N = 500$ particles, the blue dot stands for the solution of the deterministic trajectory at time t . Its trajectory is displayed as a black dotted line, which merges with the limit cycle represented as a continuous black curve.

Acknowledgements. JCM acknowledges the support of the ERC MSCA grant SLOHD (101203974), and of the French National Research Agency (ANR) under the France 2030 grant ANR-24-RR11-0002 operated by the Inria Quadrant Program.

REFERENCES

- [1] Carles Domingo-Enrich, Samy Jelassi, Arthur Mensch, Grant Rotskoff, and Joan Bruna. A mean-field analysis of two-player zero-sum games. *Advances in neural information processing systems*, 33:20215–20226, 2020.
- [2] Tomas Dominguez and Jean-Christophe Mourrat. *Statistical mechanics of mean-field disordered systems: a Hamilton-Jacobi approach*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich, 2024.
- [3] Gonçalo dos Reis, William Salkeld, and Julian Tugaut. Freidlin-Wentzell LDP in path space for McKean-Vlasov equations and the functional iterated logarithm law. *Ann. Appl. Probab.*, 29(3):1487–1540, 2019.
- [4] Wei Hong, Shanshan Hu, and Wei Liu. McKean-Vlasov SDE and SPDE with locally monotone coefficients. *Ann. Appl. Probab.*, 34(2):2136–2189, 2024.
- [5] Juno Kim, Kakei Yamamoto, Kazusato Oko, Zhuoran Yang, and Taiji Suzuki. Symmetric mean-field Langevin dynamics for distributional minimax problems. In *International Conference on Learning Representations*, 2024.
- [6] Yulong Lu. Two-scale gradient descent ascent dynamics finds mixed Nash equilibria of continuous games: a mean-field perspective. In *International Conference on Machine Learning*, pages 22790–22811. PMLR, 2023.
- [7] Chao Ma and Lexing Ying. Provably convergent quasistatic dynamics for mean-field two-player zero-sum games. In *International Conference on Learning Representations*, 2022.
- [8] Jean-Christophe Mourrat. PL conditions do not guarantee convergence of gradient descent-ascent dynamics. Preprint, arXiv:2602.16517.
- [9] Felix Otto and Cédric Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.
- [10] Lawrence Perko. *Differential equations and dynamical systems*, volume 7 of *Texts in Applied Mathematics*. Springer-Verlag, New York, third edition, 2001.
- [11] Geuntaek Seo, Minseop Shin, Pierre Monmarché, and Beomjun Choi. Local exponential stability of mean-field langevin descent-ascent in Wasserstein space. Preprint, arXiv:2602.01564.
- [12] Guillaume Wang and Lénaïc Chizat. Open problem: Convergence of single-timescale mean-field Langevin descent-ascent for two-player zero-sum games. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 5345–5350. PMLR, 2024.
- [13] Guillaume Wang and Lénaïc Chizat. Local convergence of mean-field Langevin dynamics: from gradient flows to linearly monotone games. Preprint, arXiv:2602.11999.

(Jean-Christophe Mourrat) DEPARTMENT OF MATHEMATICS, ENS LYON AND CNRS, LYON, FRANCE

(Loucas Pillaud-Vivien) CERMICS, CNRS, ENPC, INSTITUT POLYTECHNIQUE DE PARIS, MARNE-LA-VALLÉE, FRANCE