Ann. Inst. H. Poincaré
Anal. Non Linéaire 40 (2023), 631–716
DOI 10.4171/AIHPC/47

On the trend to global equilibrium for Kuramoto oscillators

Javier Morales and David Poyato

Abstract. In this paper we study convergence to the stable equilibrium for Kuramoto oscillators. Specifically, we derive estimates on the rate of convergence to the global equilibrium for solutions of the Kuramoto–Sakaguchi equation departing from generic initial data in a large coupling strength regime. As a by-product, using the stability of the equation in the Wasserstein distance, we quantify the rate at which discrete Kuramoto oscillators concentrate around the global equilibrium. In doing this, we achieve a quantitative estimate in which the probability that oscillators concentrate at the given rate tends to 1 as the number of oscillators increases. Among the essential steps in our proof are (1) an entropy production estimate inspired by the formal Riemannian structure of the space of probability measures, first introduced by Otto (2001); (2) a new quantitative estimate on the instability of equilibria with antipodal oscillators based on the dynamics of norms of the solution in sets evolving by the continuity equation; (3) the use of generalized local logarithmic Sobolev- and Talagrand-type inequalities, similar to those derived by Otto and Villani (2000); (4) the study of a system of coupled differential inequalities by a treatment inspired by Desvillettes and Villani (2005). Since the Kuramoto–Sakaguchi equation is not a gradient flow with respect to the Wasserstein distance, we derive such inequalities under a suitable fibered transportation distance.

Contents

2020 Mathematics Subject Classification. Primary 34C15; Secondary 34D06, 35B40, 35Q70, 35Q83, 70F99, 92B20, 92B25.
Keywords. Kuramoto model, synchronization, Wasserstein distance, order parameters, gradient flow, entropy production, logarithmic Sobolev inequality, Talagrand inequality.

1. Introduction

In this paper, we quantify the rate of convergence to the global equilibrium for C^1 solutions to the Kuramoto–Sakaguchi equation departing from generic initial data in the large coupling strength regime, providing a first quantitative result in this context. As a by-product, we derive a quantitative statistical estimate, on the rate of concentration for the original agent-based Kuramoto model. Such a model was introduced by Kuramoto several decades ago [36, 37] and is one of the paradigms to study collective synchronization phenomena in biological and mechanical systems in nature. It has gained extensive attention from the physics and mathematics communities; see [1, 3, 5, 9, 13, 16, 23, 27, 32, 33, 41, 48, 50, 58].

The main motivation to perform our study on the Kuramoto–Sakaguchi equation is threefold. First, this model has become a starting point for a broad family of models in collective dynamics. Historically, many of the central analytical techniques developed to study such models were first applied to the Kuramoto model and later generalized to the rest of the field. Second, the Kuramoto model provides a concrete example of a gradient flow structure in which the energy functional is not convex. This lack of convexity generates challenges to use the theory of gradient flows to derive rates of convergence. Third, we are interested in quantifying the relaxation time of a non-deterministic event. Specifically, in a large coupling strength regime, one expects relaxation of the particle system to the global equilibrium with almost sure probability. However, examples can be constructed such that relaxation fails for some well-prepared initial data.

As anticipated, in the case of identical oscillators, the Kuramoto–Sakaguchi equation exhibits a gradient flow structure in the space of probability measures under the Wasserstein distance. Nowadays, it is well known that transportation distances between measures can be successfully used to study evolutionary equations. More precisely, one of the most surprising achievements of [35, 45, 46] has been that many evolutionary equations of the form

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\nabla \rho + \rho \nabla V + \rho(\nabla W * \rho))$$

can be seen as gradient flows of some entropy functional in the spaces of probability measures with respect to the Wasserstein distance

$$W_2(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\gamma \right)^{\frac{1}{2}},$$

where the infimum ranges over all the possible transference plans, i.e.,

$$\Pi(\mu, \nu) := \left\{ \gamma \in \mathbb{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi_{1\#}\gamma = \mu \text{ and } \pi_{2\#}\gamma = \nu \right\}.$$

When such entropy functionals are convex with respect to the Wasserstein distance, this interpretation allows entropy estimates and functional inequalities to be proved (see [57] for more details on this area). Such tools, in turn, can be used to obtain convergence rates and stability estimates of the corresponding equations.

There are two main difficulties when one tries to use such a theory in the Kuramoto–Sakaguchi equation. First, even in the identical case, as for the Kuramoto model, the entropy functional associated with the equation does not satisfy the necessary convexity hypothesis. Second, in the non-identical case, the Wasserstein gradient flow structure of the equation is not available. On the other hand, the Kuramoto–Sakaguchi equation has the virtue that the broad family of unstable equilibria is characterized easily. Thus, it provides an ideal setting in which to develop techniques to attack the lack of convexity.

In this article, we show how we can circumvent the aforementioned two difficulties. On the other hand, we adapt the entropy method developed by Desvillettes and Villani [19] and we derive quantitative convergence rates for the particular non-convex case of the Kuramoto–Sakaguchi equation (see [10–12, 17, 18, 25, 53] for further applications to the Boltzmann equation and other related models). On the other hand, for such an equation, we find appropriate local logarithmic Sobolev- and Talagrand-type inequalities that are reminiscent of those obtained by Otto and Villani [47]. We hope that this provides insight into how to attack those difficulties in more general situations.

In this section we will first introduce the model. Then we will recall the current state of the art regarding the asymptotics of the model in a strong coupling strength regime. Finally, we will state our main result, the proof of which will be the object of the rest of the paper.

1.1. The Kuramoto model

The Kuramoto model governs the synchronization dynamics of N oscillators – each identified by its phase and natural frequency pair $(\theta_i(t), \omega_i)$ in $\mathbb{T} \times \mathbb{R}$. Such dynamics is given by the system

$$\begin{cases} \dot{\theta}_i = \omega_i + \dfrac{K}{N} \displaystyle\sum_{j=1}^{N} \sin(\theta_j - \theta_i), \\ \theta_i(0) = \theta_{i,0}, \end{cases} \tag{1.1}$$

for $i = 1, \dots, N$. The large crowd dynamics, $N \to \infty$, is captured by the kinetic description, given by the Kuramoto–Sakaguchi equation, which governs the probability distribution of oscillators $f(t, \theta, \omega)$ at $(t, \theta, \omega) \in \mathbb{R}^+ \times \mathbb{T} \times \mathbb{R}$:

$$\begin{cases} \dfrac{\partial f}{\partial t} + \dfrac{\partial}{\partial \theta}(v[f]f) = 0, & (\theta, \omega) \in \mathbb{T} \times \mathbb{R},\ t \geq 0, \\ f(0, \theta, \omega) = f_0(\theta, \omega), & (\theta, \omega) \in \mathbb{T} \times \mathbb{R}. \end{cases} \tag{1.2}$$

We denote the velocity field by $v[f]$, that is,

$$v[f](t, \theta, \omega) := \omega + K \int_{\mathbb{T}} \sin(\theta' - \theta)\rho(t, \theta')\, d\theta', \tag{1.3}$$

and we define

$$\rho(t, \theta) := \int_{\mathbb{R}} f(t, \theta, \omega)\, d\omega, \quad g(\omega) := \int_{\mathbb{T}} f(t, \theta, \omega)\, d\theta = \int_{\mathbb{T}} f_0(\theta, \omega)\, d\theta.$$

Here, K is the positive coupling strength and measures the degree of the interaction between oscillators, and ρ and g respectively describe the macroscopic phase density and natural frequency distribution. The rigorous derivation from (1.1) to (1.2) was done by Lancellotti [38] using Neunzert's method [44].

1.2. The gradient flow structure and stationary solutions

The Kuramoto model in $\mathbb{T}^N$ can be lifted to a dynamical system in $\mathbb{R}^N$. Van Hemmen and Wreszinki [54] observed that by doing this the Kuramoto model can be formulated as a gradient flow of the energy

$$V(\Theta) = -\frac{1}{N}\sum_{j=1}^{N}\omega_j\theta_j + \frac{K}{2N^2}\sum_{k,j=1}^{N}(1-\cos(\theta_j-\theta_k)), \tag{1.4}$$

under the metric of $\mathbb{R}^N$ induced by the scaled inner product

$$\langle v, w\rangle_N = \frac{v\cdot w}{N}. \tag{1.5}$$

Here, $\Theta = (\theta_1,\ldots,\theta_N)$, v, and w belong to $\mathbb{R}^N$. Specifically, (1.1) solves the gradient flow problem

$$\begin{cases} \dot{\Theta}(t) = -\nabla_N V(\Theta(t)), \\ \Theta(0) = \Theta_0, \end{cases} \tag{1.6}$$

where ∇_N denotes the gradient with respect to the scaled inner product. Let us also recall that if we define the order parameters $\Theta \mapsto r(\Theta), \phi(\Theta)$ by the relation

$$r(\Theta)e^{i\phi(\Theta)} := \frac{1}{N}\sum_{k=1}^{N}e^{i\theta_k}, \tag{1.7}$$

then we have that the potential reads

$$V(\Theta) = -\frac{1}{N}\sum_{j=1}^{N}\omega_j\theta_j + \frac{K}{2}(1-r^2(\Theta)), \tag{1.8}$$

and the gradient slope takes the form

$$|\nabla_N V(\Theta)|_N^2 = \frac{1}{N}\sum_{j=1}^{N}|\omega_j - Kr(\Theta)\sin(\theta_j-\phi(\Theta))|^2. \tag{1.9}$$

The main interest of the order parameter is that $r(\Theta)$ represents a measure of coherence for the ensemble of oscillators. Specifically, when $r(\Theta)$ is close to 1, then all the phases θ_i within Θ tend to be synchronized around the same phase value. Moreover, using them we can rewrite system (1.1) as

$$\dot{\theta}_i = \omega_i - Kr(\Theta)\sin(\theta_i - \phi(\Theta)),$$

for every $i = 1, \dots, N$. Without loss of generality we may assume that the natural frequencies are centered, i.e.,

$$\frac{1}{N}\sum_{i=1}^{N}\omega_i = 0. \tag{1.10}$$

We observe that such a condition is not restrictive because we can always perform a linear change of the reference frame to guarantee it. However, this condition is necessary to show the existence of stationary states and we will assume it throughout the paper. In the sequel we recall the conditions for a stationary state $\Theta_\infty = (\theta_{1,\infty}, \dots, \theta_{N,\infty})$ with given order parameters $r_\infty := r(\Theta_\infty) > 0$ and $\phi_\infty := \phi(\Theta_\infty)$ to exist. Specifically, for such a Θ_∞ we must have $\nabla V(\Theta_\infty) = 0$. Using (1.9), we readily obtain the necessary condition

$$\max_{1\le j\le N} |\omega_j| \le K r_\infty. \tag{1.11}$$

By solving the algebraic equation on $\theta_{j,\infty}$ according to the two branches of the sine function, we find that there must exist a partition $\{1, \dots, N\} = I^+ \cup I^-$ into two disjoint sets of indices I^+ and I^- such that

$$\theta_{j,\infty} = \begin{cases} \phi_\infty + \arcsin\Big(\dfrac{\omega_j}{K r_\infty}\Big) & \text{if } j \in I^+, \\ \phi_\infty + \pi - \arcsin\Big(\dfrac{\omega_j}{K r_\infty}\Big) & \text{if } j \in I^-. \end{cases} \tag{1.12}$$

Note that I^+ describes phases around the order parameter ϕ_∞, while I^- stands for phases near the antipode. Finally, in agreement with (1.7), r_∞ and ϕ_∞ cannot be arbitrary. Indeed, since we assume the centering condition (1.10) then by direct computation we find the compatibility condition

$$\frac{1}{N}\sum_{j\in I^+}\sqrt{1-\Big(\frac{\omega_j}{K r_\infty}\Big)^2} - \frac{1}{N}\sum_{j\in I^-}\sqrt{1-\Big(\frac{\omega_j}{K r_\infty}\Big)^2} - r_\infty = 0. \tag{1.13}$$

Altogether, (1.11) and (1.13) provide the necessary and sufficient conditions for r_∞ and ϕ_∞ to admit an equilibrium, whose explicit form is given by (1.12). Note that the implicit equation (1.13) does not always need to admit a solution and special configurations of parameters are needed (e.g., $I^- = \emptyset$ and large enough K; see [1, 16, 27]).

In the same spirit, the Hessian operator of the potential V is given by

$$\langle D_N^2 V(\Theta)v, v\rangle_N = \frac{K}{N}\sum_{j=1}^{N} r\cos(\theta_j - \phi)|v_j|^2 - K\left|\frac{1}{N}\sum_{j=1}^{N} v_j e^{i\theta_j}\right|^2, \tag{1.14}$$

where $D_N^2 V$ denotes the Hessian operator with respect to the scaled inner product (1.5) and $v = (v_1, \dots, v_N)$ is contained in $\mathbb{R}^N$. From this, after accounting for the rotational invariance of the model, we deduce that the stable equilibrium corresponds to $I^- = \emptyset$ where there is no antipodal mass, so that, for every $j = 1, \dots, N$,

$$\theta_{j,\infty} = \phi + \arcsin\Big(\frac{\omega_j}{K r_\infty}\Big).$$

Remark 1.1. When $r = 0$ there are plenty more equilibria. In the identical case it can be shown that they are non-isolated even after accounting for rotational invariance.

For the Kuramoto–Sakaguchi equation, in the case of identical oscillators, the equation enjoys a Wasserstein gradient flow structure (we refer the reader to [31, Appendix A]). In the non-identical case, this structure is not strictly available. Nonetheless, in our analysis, we use several techniques and objects inspired by the theory of gradient flows in the space of probability measures. Similarly, if we consider the continuous version of the order parameters

$$Re^{i\phi} := \int_{\mathbb{T}\times\mathbb{R}} e^{i\theta} f(t,\theta,\omega)\, d\theta\, d\omega, \tag{1.15}$$

equation (1.2) can be restated as

$$\begin{cases} \dfrac{\partial f}{\partial t} + \dfrac{\partial}{\partial \theta}(\omega f - KR\sin(\theta-\phi) f) = 0, & (\theta,\omega)\in\mathbb{T}\times\mathbb{R},\ t\geq 0,\\ f(0,\theta,\omega) = f_0(\theta,\omega), & (\theta,\omega)\in\mathbb{T}\times\mathbb{R}. \end{cases}$$

Without loss of generality, we can assume that g is centered as well, i.e.,

$$\int_{\mathbb{R}} \omega g(\omega)\, d\omega = 0. \tag{1.16}$$

Again, this is a necessary condition for equilibria to exist and we will assume it throughout the paper. As for the discrete system, we obtain analogous conditions like (1.11) and (1.13) for an equilibrium f_∞ with given order parameters R_∞ and ϕ_∞ to exist. Specifically, if $\operatorname{supp} g \subset [-W, W]$, then we have the necessary condition

$$W \leq KR_\infty. \tag{1.17}$$

In addition, there must exist a decomposition $g = g^+ + g^-$ with non-negative g^+ and g^- such that

$$\begin{aligned} f_\infty(\theta,\omega) &= g^+(\omega)\otimes\delta_{\vartheta^+(\omega)}(\theta) + g^-(\omega)\otimes\delta_{\vartheta^-(\omega)}(\theta), \\ \vartheta^+(\omega) &:= \phi_\infty + \arcsin\Big(\frac{\omega}{KR_\infty}\Big), \\ \vartheta^-(\omega) &:= \phi_\infty + \pi - \arcsin\Big(\frac{\omega}{KR_\infty}\Big). \end{aligned} \tag{1.18}$$

Recall that R_∞ and ϕ_∞ are not arbitrary, but the following continuous version of the compatibility condition (1.13) must hold true (see [42] and also [22, Section 1(b)]):

$$\int_{\mathbb{R}} \sqrt{1-\Big(\frac{\omega}{KR_\infty}\Big)^2}\,(g^+(\omega) - g^-(\omega))\, d\omega - R_\infty = 0. \tag{1.19}$$

Once again, (1.17) and (1.19) provide the necessary and sufficient conditions for R_∞ and ϕ_∞ to admit an equilibrium, whose explicit form is given by (1.18). As will become apparent later in the paper (see also [13]), the stable equilibria with $R_\infty > 0$ correspond to the case $g^- = 0$ where there is no antipodal mass.

1.3. Statement of the problem and main results

By direct inspection of the Hessian of the energy (1.14), one can see that, in a large coupling strength regime, out of all of the possible equilibria up to rotations, there is only one that is stable. That is the equilibrium in which the Hessian operator is strictly positive on the subspace orthogonal to rotations. One expects that with probability 1, system (1.1) should converge to such equilibria if the coupling strength is sufficiently large. This phenomenon has been widely observed in numerical simulations, but it remained an open problem for a long time. After some previous partial attempts, the most satisfying answer was recently obtained in [34].

There have been many approaches in the literature to show the convergence of the system to the critical points of (1.14) in the large coupling strength regime. Since stable equilibria have oscillators contained within an interval of size less than π, convergence results have been mainly addressed in the particular case where initial data is originally confined to such a basin of attraction, namely a half-circle. Specifically, in [16, 27] a system of differential inequalities was found for the phase and frequency diameter, which yields the convergence of the system to a phase-locked state. Recall that (1.1) is a gradient flow (1.6) governed by a potential energy (1.4). In [33, 39] the authors derived the convergence to phase-locked states using Łojasiewicz's gradient inequality for analytic potentials [40] and it was used to obtain convergence rates (after some unquantified initial time) in some particular cases where the Łojasiewicz exponent can be explicitly computed. For general initial data along the whole circle, the literature is rare. One of the main difficulties when trying to use standard theory from dynamical systems to show this is the fact that critical points of (1.14) are not isolated (see Remark 1.1). The main contributions in that direction are [28] and [34], but explicit convergence rates are not available. On the one hand, in [28] the authors quantified a lower bound on the coupling strength guaranteeing convergence to the stable equilibrium departing from generic initial configurations. Unfortunately, bounds were not sharp as they depended on the number N of particles. Recently, those results were improved in [34] and N-independent bounds were derived.

In the continuum case (1.2), accumulation of oscillators in the opposite hemisphere of the order parameter was excluded in [31]. Specifically, the authors found a mass concentration phenomenon around the order parameter by obtaining invariant sets containing a sufficiently large portion of the mass. The method of proof is related in spirit to the above result [34] for the discrete model. However, convergence towards a stationary solution was not yet established for generic initial data. See [13] for a particular proof when the phase diameter is smaller than π. Additionally, see [5] for a description of the equilibrium in the kinetic case, where a conditional convergence result is presented, without rates. To date, regarding generic initial data, there are only arguments based on compactness that do not give any bound on the rate of convergence. We remark that while [34] provides N-independent estimates, that is not enough to lift the convergence to equilibrium to the continuous equation. The main obstruction is that the only existing results in the literature on mean field limits for the Kuramoto model are either local in time [38] or uniform in time for restricted initial data [30]. In addition, convergence rates in [34] are not explicit.

Our goal here is precisely to investigate the long-time relaxations of solutions to the global equilibrium. We are interested in the study of rates of convergence for the Kuramoto–Sakaguchi equation towards the stable equilibria from generic initial data in the large coupling strength regime. Additionally, we wish to derive constructive bounds for this convergence and use them to obtain quantitative information about the convergence of the particle system to the global equilibria as well. There are several reasons why one may be interested in explicit bounds on the rate of convergence. In particular, one may look for the qualitative properties of solutions. More importantly, only after getting convergence rates can we use the dynamics of the kinetic equations to deduce quantitative statistical information about the particle system.

The first thing that one might be tempted to do is to apply linearization techniques around the equilibria. This analysis has been done in [20–23], and is connected with the methods in Landau damping. However, there is a fundamental reason not to be content with that analysis, which has to do with the nature of linearization. Quoting Desvillettes and Villani [19],

> *This technique is likely to provide excellent estimates of convergence only after the solution has entered a narrow neighborhood of the equilibrium state, narrow enough that only linear terms are prevailing in the equation. But by nature, it cannot say anything about the time needed to enter such a neighborhood; the latter has to be estimated by techniques which would be well-adapted to the non-linear equation.*

Here is where our contribution takes places, and this is why we will not rely on linearization techniques. Instead, we will stick as close as possible to the physical mechanism of entropy production. Our main result is here:

Theorem 1.1. *Let f_0 be contained in $C^1(\mathbb{T}\times\mathbb{R})$ and let g be compactly supported in $[-W,W]$. Consider the unique global-in-time classical solution $f=f(t,\theta,\omega)$ to (1.2). Then there exists a universal constant C such that if*

$$\frac{W}{K}\leq CR_0^3, \tag{1.20}$$

we can find a time T_0 with the property that

$$T_0\lesssim\frac{1}{KR_0^2}\log\Big(1+W^{1/2}\|f_0\|_2+\frac{1}{R_0}\Big), \tag{1.21}$$

and

$$W_2(f,f_\infty)\lesssim e^{-\frac{1}{40}K(t-T_0)}$$

for every t in $[T_0,\infty)$. Here, f_∞ is the unique global equilibrium of the Kuramoto–Sakaguchi equation up to rotations (see Proposition 3.3).

In the above theorem and throughout the rest of the paper, given two function h_1 and h_2 involving the different parameters in our system, we say that $h_1 \lesssim h_2$ if there exists a universal constant C such that $h_1 \leq C h_2$. Since our argument is constructive, every time we use such notation, we could compute C explicitly. Additionally, because we often deal with absolutely continuous measures, by abuse of notation, we will sometimes use f to denote the measure $f\,dx$.

As a direct consequence of our main theorem, we obtain the following quantitative concentration estimate for the particle system, which complements the results in [34, Theorems 3.2 and 3.3].

Corollary 1.1. *Let μ_t^N be a sequence of empirical measures associated to solutions of the particle system (1.1) starting at independent and identically distributed random initial data with law f_0 (see Section 6 for further details). Assume that f_0, R_0, K, and W satisfy the hypotheses of Theorem 1.1 and let L be an interval with diameter $2/5$ centered around the phase ϕ_∞ of the global equilibrium f_∞. Then there exists a positive time T_0 satisfying (1.21) and an integer N^* with the property that*

$$\log N^* \lesssim \frac{1}{R_0^2} \log\Big(1 + W^{1/2} \|f_0\|_2 + \frac{1}{R_0}\Big),$$

and for any $N \geq N^*$ and any s contained in the interval

$$\Big[T_0, T_0 + \frac{1}{25K} \log\Big(\frac{N}{N^*}\Big)\Big],$$

we can quantify the probability of mass concentration and diameter contraction of the particle system with N oscillators. Indeed, we have

$$\mathbb{P}\Big(\forall t \geq s,\ \exists\, L_s^N(t) \subset \mathbb{T} : L_s^N(s) = L \text{ and (M)–(D) hold}\Big) \geq 1 - C_1 e^{-C_2 N^{\frac{1}{2}}}.$$

Here, conditions (M) and (D) yield mass concentration and diameter contraction. More precisely, such properties are given by

$$\mu_t^N(L_s^N(t) \times \mathbb{R}) \geq 1 - \frac{1}{5} e^{-\frac{1}{20}K(s-T_0)} \qquad \text{for every } t \text{ in } [s,\infty), \tag{M}$$

$$\operatorname{diam}(L_s^N(t)) \leq \max\Big\{\frac{4}{5} e^{-\frac{K}{20}(t-s)}, 12\frac{W}{K}\Big\} \qquad \text{for every } t \text{ in } [s,\infty). \tag{D}$$

Additionally, C_1 and C_2 are universal positive constants which could be explicitly computed.

Remark 1.2. Throughout the paper, we will consider generic initial datum f_0 in C^1 but large coupling strength K satisfying condition (1.20) in Theorem 1.1. The large coupling strength condition is necessary to guarantee convergence to global equilibrium given that the Kuramoto–Sakaguchi equation exhibits a phase transition at a critical value of coupling strength $K = K_c$; see [15, 20, 22, 23, 37, 50–52].

Notice that although initial data is generic, condition (1.20) depends on the degree of (dis)order of the initial data so that the closer R_0 is to zero, the larger K is. This has proven a necessary condition in the particle system too; see Remark 3.4.

Moreover, since none of the constants in Theorem 1.1 depend on norms of the derivative of the solution, they cannot be recovered via scaling arguments and near equilibrium analysis from works such as [20–23].

In fact, when our result is applied to the case of identical oscillators, that is, all the oscillators have the same natural frequency, condition (1.20) can be completely removed and the estimates on T_0 and N^* remain valid when one replaces the term $W^{\frac{1}{2}}\|f_0\|_2$ with the L^2 norm of the initial distribution of phases.

Finally, notice that when R_0 is close to zero then bound (1.21) in Theorem 1.1 predicts an infinitely large transient of time T_0, which is required for the solution to enter the concentration regime. This is not an artifact of the proof, but it is actually observed in the particle system too; see Remark 3.5.

1.4. Ingredients of the proof

The proof of Theorem 1.1 is the first quantitative proof for the relaxation problem of the solutions of the Kuramoto–Sakaguchi equation (1.2) with generic initial data in the large coupling strength regime. It is intricate but rests on a few well-identified principles. Such principles apply with a lot of generalities to many variants of the Kuramoto model. The proof builds upon the following ingredients.

- A quantitative entropy production estimate inspired by the formal Riemannian calculus of probability measures under the Wasserstein distance, which we address in Sections 3.3 and 5. See [46] and also [31, Appendix A] for an overview in the context of Kuramoto–Sakaguchi with identical oscillators. One form of this estimate was originally presented in [31, Lemma 6.5], but we use a more refined version in this work.
- A fibered Wasserstein distance $W_{2,g}$ presented independently in [43] and [49]. This distance is well adapted to the non-linear problem. By using this distance, in Section 3.1 we will derive new logarithmic Sobolev- and Talagrand-type inequalities associated with it; see [47].
- A quantitative instability estimate excluding the equilibria with mass in the opposite pole of the order parameter, which we derive in Section 4. One form of this estimate was originally presented in [31, Corollary 6.2], but we use a more refined version in this work.
- A quantitative emergence of attractor sets, which we derive in Section 4.1. Those sets consist of arcs around the order parameter containing a significant portion of mass which stays together through the evolution. This result is the natural continuous counterpart of the recent result in [34] for the discrete system.

- A new estimate on the norms of the solution on sets evolving by the flow of the continuity equation that allows us to propagate information along the different parts of the system. We discuss these estimates in Section 4 and we use them in Section 4.2 to control the L^2 norm outside the attractors.

The above ingredients are directed at two main objectives, which will settle the basis of the proof of Theorem 1.1 in Section 3.2. First, the second item above pursues a new transportation–dissipation inequality adapted to the Kuramoto–Sakaguchi equation, which involves the novel fibered distance $W_{2,g}$ and a generalized dissipation functional intrinsic to the system. Conditioned to a concentration regime of phases around the order parameter, the solution enters a "convexity area" and the functional inequality implies the convergence to global equilibrium. Second, all other items will be used to precisely quantify the phase concentration mechanism of the Kuramoto–Sakaguchi equation in the large coupling strength regime, which will take the following explicit form.

Lemma 1.1. *Assume that f_0, R_0, K, and W satisfy the hypotheses of Theorem 1.1, consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2), and set $\beta = \pi/3$. Then there exists a positive time T_0 verifying (1.21) such that*

$$R(t) \geq \frac{3}{5} \quad \text{and} \quad \rho(\mathbb{T} \setminus L_\beta^+(t)) \leq e^{-\frac{1}{20}K(t-T_0)} \tag{1.22}$$

for every t in $[T_0, \infty)$.

As discussed in Section 1.3, a weak version of this result was obtained in [31] without an explicit control of the time T_0, and without explicit rates. Since Lemma 1.1 is intricate and requires introducing all the above machinery first, its proof will comprise the core of this paper and we postpone it to Section 5.2.

For pedagogical reasons, before entering into the details of the proof, first we will provide a summary of the strategy and this will be the objective of the next section.

2. Strategy

In this section we will describe the plan of the proof of Theorem 1.1, and the system of differential inequalities upon which our estimates of convergence are based.

Two of the main features of our proof are the fact that it follows the intuition derived from the mechanism of entropy production, and it is systematic. Additionally, it capitalizes on the behavior observed in numerical simulations under a large coupling strength regime.

We will overcome three main difficulties. First, the order parameter R defined in (1.15) is not monotonic and when it vanishes so does the mean-field force between particles. Additionally, our description of the equilibria is only valid when it is positive. This difficulty also plays an essential role in the particle system; see [28, 34]. The second difficulty is the fact that the Kuramoto–Sakaguchi equation tends to concentrate the density, which produces exponential growth of the global L^p norms for $p > 1$. The third difficulty, related to the second one, is that a large family of equilibria with mass in the opposite hemisphere of the order parameter appears in which the entropy production vanishes.

In the particle system (1.1), the potential function V plays the role of the entropy. Consequently, since the particle system is a gradient flow (see (1.6)), we have

$$\frac{d}{dt}V(\Theta(t)) = -|\nabla_N V(\Theta(t))|_N^2.$$

Thus, we can see from this expression that when the particle system slope $|\nabla_N V(\Theta(t))|_N^2$ is large, then the potential function $V(\Theta(t))$ should decrease locally. To quantify the rate of increase of the slope, the starting point is the Hessian operator (1.14) of the energy functional for the particle system. Such an expression implies that $D_N^2 V(\Theta(t))$ is bounded from above (as a quadratic form) by $Kr(\Theta(t))$, that is,

$$\langle D_N^2 V(\Theta(t))v, v\rangle_N \leq Kr(\Theta(t))|v|_N^2$$

for any $(v_1, \ldots, v_N)$ in $\mathbb{R}^N$, which implies the differential inequality

$$-2Kr(\Theta(t))|\nabla_N V(\Theta(t))|_N^2 \leq \frac{d}{dt}|\nabla_N V(\Theta(t))|_N^2 \leq 2K|\nabla_N V(\Theta(t))|^2,$$

along solutions of the Kuramoto model (1.1). Notice that by (1.8),

$$\frac{V(\Theta(t))}{K} \quad \text{and} \quad 1 - r^2(\Theta(t))$$

are related up to lower-order terms that can be neglected thanks to condition (1.20). Similarly, considering the time derivative of the above quantities, we have that the two expressions

$$\frac{|\nabla_N V(\Theta(t))|_N^2}{K} \quad \text{and} \quad \frac{d}{dt}r^2(\Theta(t))$$

should also differ by a lower-order term that, again, can be controlled using (1.20). This justifies that, in the large coupling strength regime, we indistinctly call $\frac{d}{dt}r^2(\Theta(t))$ and $|\nabla_N V(\Theta(t))|_N^2$ the dissipation.

In the continuous case, those objects were extended to the setting of the Kuramoto–Sakaguchi equation (1.2) with identical oscillators using the Riemannian structure for the space of probability measures; see [31, Appendix A]. However, in the non-identical case the Kuramoto–Sakaguchi equation (1.2) is not a Wasserstein gradient flow and this presents an obstacle to try to use the above objects. By analogy, let us define the continuum analog of the particles' slope (1.9) given by

$$\mathcal{I}[f] := \int_{\mathbb{T}\times\mathbb{R}} (\omega - KR\sin(\theta - \phi))^2 f \, d\theta \, d\omega. \tag{2.1}$$

We will again call this quantity the dissipation. Indeed, notice that taking derivatives in (1.15), one clearly obtains the following dynamics of the order parameters:

$$\begin{aligned} \dot{R} &= -\int_{\mathbb{T}\times\mathbb{R}} \sin(\theta - \phi)(\omega - KR\sin(\theta - \phi)) f \, d\theta \, d\omega, \\ \dot{\phi} &= \frac{1}{R}\int_{\mathbb{T}\times\mathbb{R}} \cos(\theta - \phi)(\omega - KR\sin(\theta - \phi)) f \, d\theta \, d\omega. \end{aligned} \tag{2.2}$$

Using it, we will show, in Lemma 3.3, that dissipation and the time derivative of the order parameter are again related up to lower-order terms that can be controlled by condition (1.20), i.e.,

$$\mathcal{I}[f] - W^2 \leq K \frac{d}{dt}(R^2) \leq 3\,\mathcal{I}[f] + W^2.$$

Indeed, in Corollary 3.1 we show that we can again control the growth of the dissipation in the continuous description in a similar way, namely,

$$-2KR\,\mathcal{I}[f] \leq \frac{d}{dt}\mathcal{I}[f] \leq 2K\,\mathcal{I}[f].$$

In Section 2.2, we will describe how this relationship, along with the principle of entropy production discussed below in Section 2.1, can be used to provide a universal lower bound on $R(t)$ of the form λR_0, for some λ in $(0, 1)$. In fact, we will show that by making K sufficiently large we can make λ as close to 1 as needed.

2.1. Displacement concavity and entropy production

We start by describing informally the entropy production principle in our context. Roughly speaking, it will quantify the following fundamental fact:

> *If at some time t the system is far from the family of equilibria with positive order parameter, then the order parameter will increase significantly in the next few instants of time.*

Before making it rigorous, we set some necessary notation that we will systematically use throughout the paper. We define a dynamic neighborhood of the order parameter ϕ and its antipode as follows.

Definition 2.1. Given an angle α in $(0, \frac{\pi}{2})$, we denote by $L_\alpha^+(t)$ the interval (arc) in $\mathbb{T}$ that is centered around $\phi(t)$, and has a diameter $\pi - 2\alpha$, that is,

$$L_\alpha^+(t) := \left(\phi(t) - \frac{\pi}{2} + \alpha, \phi(t) + \frac{\pi}{2} - \alpha\right).$$

Similarly, we denote by $L_\alpha^-(t)$ the interval (arc) in $\mathbb{T}$ of the same diameter that is centered around the antipode $\phi(t) + \pi$, that is,

$$L_\alpha^-(t) := \left(\phi(t) + \frac{\pi}{2} + \alpha, \phi(t) + \frac{3\pi}{2} - \alpha\right).$$

In this way, $L_\alpha^+(t) \cup L_\alpha^-(t)$ is a neighborhood of the average phase and its antipode.

Also, here and throughout the rest of the paper, given a measurable set $B \subset \mathbb{T}$ we define

$$\rho_t(B) := \int_B \rho(t, \theta)\, d\theta,$$

and more generally,

$$\rho(A(t)) := \int_A \rho(t,\theta)\,d\theta$$

for any time-dependent family of measurable sets $t \to A_t$.

Notice that by the explicit formula (1.18), all the possible equilibria in our analysis have phase support confined to small arcs centered around ϕ and its antipode $\phi+\pi$. Indeed, by our assumption (1.20) on $\frac{W}{K}$, the diameter of those arcs can be made arbitrarily small. Therefore, we may fix any small value of α as the size of the neighborhood $L_\alpha^+(t)\cup L_\alpha^-(t)$, containing the phase support of equilibria. For simplicity, we will set $\alpha := \frac{\pi}{6}$ throughout the paper.

To appropriately formulate our entropy production principle, let us depart from our dissipation functional (2.1). As for the particle system (1.9), notice that $\mathcal{I}[f]$ vanishes if, and only if, f is an equilibrium. Hence, $\mathcal{I}[f]$ could be thought of as a natural measure of how close a given f is to the family of equilibria (1.18). The starting point of the principle is the following control of the lateral mass outside the time-dependent neighborhood $L_\alpha^+(t)\cup L_\alpha^-(t)$ by the dissipation functional (equivalently $\frac{d}{dt}R^2$):

$$\rho(\mathbb{T}\setminus L_\alpha^+(t)\cup L_\alpha^-(t)) \leq \frac{1}{KR^2\cos^2(\alpha)}\frac{d}{dt}R^2 + \frac{W^2}{K^2R^2\cos^2(\alpha)}. \tag{2.3}$$

Such an inequality will be proven in Lemma 3.5 and provides relevant information on the system. Specifically, notice that when f is sufficiently far from the family of equilibria (i.e., it has enough mass outside $L_\alpha^+(t)\cup L_\alpha^-(t)$), then the dissipation $\mathcal{I}[f]$ is sufficiently large. Consequently, the time derivative of R^2 is large in this case as well, and this produces an entropy production of the system. The rigorous entropy production principle will be obtained in Lemma 3.2 and will quantify the exact gain in the order parameter. In a nutshell, if at some time $t_0\geq 0$ and for some $\lambda\in(0,1)$ we have

$$R(t_0)\geq \lambda R_0,\qquad \dot R(t_0)\geq \frac{K}{4}\lambda^3R_0^3\cos^2(\alpha),$$

then there must exist some $0<d\leq \frac{1}{3KR_0\log 10}$ so that

$$R^2(t_0+d)-R^2(t_0)\geq \frac{1}{40}\lambda^4R_0^3. \tag{2.4}$$

The growth estimate (2.4) was partially anticipated in the previous work [31, Lemma 6.5]. However, our strengthened version in Lemma 3.2 better fits the approach in this paper.

2.2. Small dissipation regime and lower bounds on the order parameter

When the dissipation is large, the above entropy production principle quantifies the gain of the order parameter in the next few instants of time. Regarding the reverse regime with small dissipation, Lemma 3.5 will show that when $\dot{R}$ is below a critical threshold, we achieve the following differential inequality:

$$\frac{d}{dt}R^2 > \frac{K}{2}\Big(-R^3 + \Big[\lambda R_0 + \frac{3}{5}(1-\lambda)R_0\Big]R^2 - \frac{3}{5}(1-\lambda)\lambda^2 R_0^3\Big), \tag{2.5}$$

which holds in any time interval $[t_1, t_2]$ such that

$$\dot{R}(t) \leq \frac{K}{4}\lambda^3 R_0^3 \cos^2(\alpha) \quad \text{in } [t_1, t_2].$$

Estimate (2.3) will be crucial to derive the lemma. Additionally, note that the right-hand side of (2.5) vanishes when $R = \lambda R_0$. In Corollary 3.6, we will combine this inequality with the above entropy production in Lemma 3.2 to quantify a universal lower bound $R(t) \geq \lambda R_0$ on the order parameter.

2.3. Instability of the antipodal equilibria

The main obstacle to use the above entropy production estimate to show the convergence to the global equilibrium is the fact that it does not exclude the possibility that $\dot{R}$ may vanish or alternate signs over long periods. To overcome such a difficulty, we need to quantify the instability of the antipodal equilibrium, which roughly speaking states the following:

> *If the system is eventually close enough to a critical point and such a critical point has mass in the opposite hemisphere of the order parameter, then the system would depart from such equilibria and mass will leave the opposite hemisphere exponentially fast.*

To quantify this instability, let us first introduce some necessary notation. We consider a smooth regularization of the characteristic function of $L_\alpha^-(t)$ as

$$\chi^-_{\alpha,\delta_0}(\theta) = \xi_{\alpha,\delta_0}(\theta - \phi - \pi),$$

where $\delta_0 > 0$ is a small fixed parameter and ξ_{α,δ_0} is a smooth regularization of the characteristic function of $[-(\frac{\pi}{2} - \alpha), (\frac{\pi}{2} - \alpha)]$, namely,

$$\xi_{\alpha,\delta_0}(r) := \begin{cases} 1 & \text{if } |r| \leq \dfrac{\pi}{2} - \alpha, \\ \dfrac{1}{1 + \exp\Big(\frac{2|r| - (\pi - 2\alpha + \delta_0)}{(\frac{\pi}{2} - \alpha + \delta_0 - |r|)(|r| - \frac{\pi}{2} + \alpha)}\Big)} & \text{if } \dfrac{\pi}{2} - \alpha \leq |r| \leq \dfrac{\pi}{2} - \alpha + \delta_0, \\ 0 & \text{if } |r| \geq \dfrac{\pi}{2} - \alpha + \delta_0. \end{cases} \tag{2.6}$$

As for α, we can take δ_0 as small as desired (e.g., $\delta_0 = 1/2$). For notational simplicity we will set

$$\xi_\alpha := \xi_{\alpha,1/2} \quad \text{and} \quad \chi^-_\alpha := \chi^-_{\alpha,1/2}.$$

Additionally, we will use the notation

$$f_t^2(B) := \int_A f^2(t,\theta,\omega)\,d\theta\,d\omega$$

for any measurable set $B \subset \mathbb{T}$ and, more generally,

$$f^2(\varphi) := \int \varphi(t,\theta,\omega) f^2(t,\theta,\omega)\,d\theta\,d\omega$$

for any function $\varphi\colon \mathbb{R}^+ \times \mathbb{T} \times \mathbb{R} \to \mathbb{R}$. Bearing all the above notation in mind, the main inequality quantifying the instability of equilibria with antipodal mass reads

$$\begin{aligned}\frac{d}{dt} f^2(\chi_\alpha^-(t)) &\leq -KR\sin(\alpha) f^2(\chi_\alpha^-(t))\\ &\quad + 4K f_t^2(\mathbb{T})\Big[\frac{W}{K} + \sqrt{\frac{2\dot{R}}{KR} + \frac{1}{R^2}\frac{W^2}{K^2}} - R\cos(\alpha)\Big]^+.\end{aligned}$$

Although this inequality is a variant of an estimate previously introduced in [31, Corollary 6.2], we prove it in Proposition 4.1 because it better fits the approach in this paper.

Notice that when the system is close enough to an equilibrium so that the dissipation is below a critical threshold, the second term of this inequality vanishes and, indeed, it establishes the instability of equilibria with antipodal mass. However, when one tries to use such an inequality to quantify the convergence rates, but the dissipation is not sufficiently small, one sees that the term $f_t^2(\mathbb{T})$ represents an obstacle. Specifically, it stands to reason that one can produce examples in which $f_t^2(\mathbb{T})$ grows exponentially fast because the Kuramoto–Sakaguchi equation concentrates mass. We solve this difficulty by adopting a Lagrangian viewpoint in which we analyze norms of the solution along sets evolving according to the continuity equation. That is the content of the next section.

2.4. Sliding norms

The key ingredient that allows us to relate the different functionals appearing in our estimates is the notion of sliding norms along the flow of the continuity equation. For this purpose, let $\mathbb{X}_{t_0,t}(\theta,\omega) = (\Theta_{t_0,t}(\theta,\omega),\omega)$ denote the forward flow map, that is,

$$\begin{cases}\dfrac{d}{dt}\mathbb{X}_{t_0,t}(\theta,\omega) = (v[f],0),\\ \mathbb{X}_{t_0,t_0}(\theta,\omega) = (\theta,\omega),\end{cases}$$

associated to the continuity equation (1.2) for any $t, t_0 \geq 0$.

For any measurable set $A \subset \mathbb{T}\times\mathbb{R}$, we will denote the image $\mathbb{X}_{t_0,t}(A)$ by $A_{t_0,t}$. For simplicity, when considering a time-dependent set $A(t)$, we will use the notation $A(t_0)_t$ to denote $A(t_0)_{t_0,t}$. Additionally, given a measurable set $B \subset \mathbb{T}$, we will use $B_{t_0,t}$ to denote the projection of $(B \times [-W,W])_{t_0,t}$ into $\mathbb{T}$. Again, if $B(t)$ is a time-dependent set in $\mathbb{T}$, we will use $B(t_0)_t$ to denote $B(t_0)_{t_0,t}$.

Now we are a position to state our sliding norm estimate which is given by

$$\frac{d}{dt}f^2(A_{t_0,t}) \leq KR\Big(\sup_{(\theta,\omega)\in A_{t_0,t}} \cos(\theta-\phi(t))\Big) f^2(A_{t_0,t}),$$

and holds for any measurable set $A \subset \mathbb{T}\times\mathbb{R}$. We prove this inequality in Lemma 4.1. To use this inequality effectively, one must obtain a control on the dynamics of sets evolving according to the characteristic flow, both in the large and small dissipation regimes. We perform this analysis in Sections 4.1 and 4.2.

2.5. The system of differential inequalities

All the above-mentioned bounds lead to a system of coupled differential inequalities and functional inequalities. For convenience, let us recast it explicitly here:

$$\frac{d}{dt}f^2(A_{t_0,t}) \leq KR\Big(\sup_{(\theta,\omega)\in A_{t_0,t}} \cos(\theta-\phi(t))\Big) f^2(A_{t_0,t}), \tag{2.7}$$

$$-2KR\,\mathcal{I}[f] \leq \frac{d}{dt}\mathcal{I}[f] \leq 2K\,\mathcal{I}[f], \tag{2.8}$$

$$\mathcal{I}[f] - W^2 \leq K\frac{d}{dt}R^2 \leq 3\,\mathcal{I}[f] + W^2, \tag{2.9}$$

$$\begin{aligned}\frac{d}{dt}f^2(\chi_\alpha^-(t)) \leq{}& -KR\sin(\alpha) f^2(\chi_\alpha^-(t))\\ &+ 4Kf_t^2(\mathbb{T})\Big[\frac{W}{K} + \sqrt{\frac{2\dot{R}}{KR} + \frac{1}{R^2}\frac{W^2}{K^2}} - R\cos(\alpha)\Big]^+,\end{aligned} \tag{2.10}$$

$$\frac{d}{dt}R^2 > K\Big(-R^3 + \Big[\lambda R_0 + \frac{3}{5}(1-\lambda)R_0\Big]R^2 - \frac{3}{5}(1-\lambda)\lambda^2 R_0^3\Big), \tag{2.11}$$

where the first inequality holds for any measurable set $A \subset \mathbb{T}\times\mathbb{R}$, the last inequality holds in any interval $[t_1, t_2]$ satisfying the hypotheses of Lemma 3.5, and all of the other inequalities above hold for every t in $[0, \infty)$.

The goal of such a system is to prove our main Lemma 1.1 by providing explicit bounds on the time T_0 and explicit rates of phase concentration. As we discuss next in Section 2.6, this information is crucial to prove our main Theorem 1.1. To achieve this, we use two main components. On the one hand, we study the dynamics of sets along the characteristic flow in Section 4. On the other hand, we recover the general entropy method developed in [19] and adapt it to our setting; see also [10–12, 17, 18, 25, 53] for applications to other models. The argument is described in detail in Section 5 and it consists in performing a subdivision into time intervals subordinated to different scales of values of the order parameter. Such intervals are classified into intervals where the dissipation is above and below a certain threshold. If the dissipation is large on an interval, we use the lower bound (2.8) in the form of our entropy production estimate to quantify the increase of the order parameter. Conversely, if the dissipation is small, we use (2.10) to quantify the departure of the system from the family of equilibria with antipodal mass. To do this effectively, we communicate information between the different regimes using inequality (2.7) and our analysis on the dynamics of sets from Sections 4.1 and 4.2.

2.6. Local displacement convexity and Talagrand-type inequalities

This part comprises the second main component which, together with Lemma 1.1, settles the basis for the proof of our main Theorem 1.1. Note that at the particle level, we see that the Hessian operator (1.14) is positive definite in the subspace orthogonal to rotations whenever the oscillators are strictly contained in a suitable interval. As mentioned in Section 1, the classical theory of gradient flows allows convergence rates towards equilibrium to be derived when the energy is strictly convex. Thus, once the mass enters exponentially fast to the region of convexity after T_0 according to Lemma 1.1, one may hope to recover such a convergence result for our system. Indeed, inspired by the arguments in [47] where the authors prove the logarithmic Sobolev and Talagrand inequalities, we derive analogous inequalities that yield the exponential convergence result and uniqueness of the global equilibrium. Since our system is not a Wasserstein gradient flow, we derive such inequalities for a fibered transportation distance $W_{2,g}$, which is well adapted to the non-linear problem. The proofs of these inequalities are the content of the next section.

3. Functional inequalities and a fibered Wasserstein distance

As discussed before, the proof of Theorem 1.1 will be split into two distinguished parts that capture two qualitatively different features of the dynamics of the Kuramoto–Sakaguchi equation (1.2). First, recall that from many preceding works (see, e.g., [5, 13, 31]) it is apparent that the entropy functional of the equation does not satisfy the necessary convexity properties for the classical theory of gradient flows to work and show convergence towards the global equilibrium. Thus, we need to prove, using different tools, that the dynamics of the equation itself drives the system towards an appropriate "convexity area" exponentially fast after some quantified time $T_0 > 0$. This is the content of Lemma 1.1, where such a convexity area is described by a dynamic neighborhood of the order parameter ϕ. The proof of this result is postponed to forthcoming sections and becomes the cornerstone of this paper.

We devote this part to studying the other main feature of the dynamics, assuming that Lemma 1.1 holds. Specifically, we show that although the system is not a Wasserstein gradient flow, the generalized dissipation functional that was introduced in (2.1) satisfies an appropriate Hessian-type inequality after the solution has entered into the concentration regime quantified in Lemma 1.1. The final step is inspired in [47] by the derivation of the logarithmic Sobolev and Talagrand inequalities for gradient flows in Wasserstein space. Indeed, we will show that despite the fact that our system is not a Wasserstein gradient flow due to the presence of heterogeneities introduced by ω, some dissipation–transportation inequality still can be achieved for an adequate distance on the space of probability measures. Such an inequality, along with the exponential decay of the dissipation, guarantees the exponential convergence to the global equilibrium in Theorem 1.1.

To start, we first study the dynamics of the dissipation functional (2.1) along the flow of the Kuramoto–Sakaguchi equation.

Theorem 3.1. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W,W]$. Consider the unique global-in-time classical solution $f=f(t,\theta,\omega)$ to (1.2). Then*

$$\frac{d}{dt}\mathcal{I}[f] = -K\int_{\mathbb{T}^2\times\mathbb{R}^2}\big((\omega - KR\sin(\theta-\phi)) - (\omega' - KR\sin(\theta'-\phi))\big)^2 \\ \times\cos(\theta-\theta')f(t,\theta,\omega)f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'.$$

Proof. Taking derivatives in the dissipation functional yields the decomposition

$$\frac{d}{dt}\mathcal{I}[f] = I_1 + I_2,$$

where each of the terms takes the form

$$I_1 := 2\int_{\mathbb{T}\times\mathbb{R}}(\omega - KR\sin(\theta-\phi))(-K\dot{R}\sin(\theta-\phi) + KR\cos(\theta-\phi)\dot{\phi})f\,d\theta\,d\omega,$$
$$I_2 := \int_{\mathbb{T}\times\mathbb{R}}(\omega - KR\sin(\theta-\phi))^2\partial_t f\,d\theta\,d\omega.$$

Let us use (2.2) and substitute the formulas for $\dot{R}$ and $\dot{\phi}$ in each term. By doing this, we get

$$\begin{aligned} I_1 &= 2K\int_{\mathbb{T}^2\times\mathbb{R}^2}(\omega - KR\sin(\theta-\phi))(\omega' - KR\sin(\theta'-\phi)) \\ &\qquad\times(\sin(\theta-\phi)\sin(\theta'-\phi) - \cos(\theta-\phi)\cos(\theta'-\phi)) \\ &\qquad\times f(t,\theta,\omega)f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega' \\ &= 2K\int_{\mathbb{T}^2\times\mathbb{R}^2}(\omega - KR\sin(\theta-\phi))(\omega' - KR\sin(\theta'-\phi))\cos(\theta-\theta') \\ &\qquad\times f(t,\theta,\omega)f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega' \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} I_2 &= \int_{\mathbb{T}\times\mathbb{R}}\partial_\theta[(\omega - KR\sin(\theta-\phi))^2](\omega - KR\sin(\theta-\phi))f\,d\theta\,d\omega \\ &= -2K\int_{\mathbb{T}\times\mathbb{R}}(\omega - KR\sin(\theta-\phi))^2R\cos(\theta-\phi)f\,d\theta\,d\omega, \end{aligned}$$

where we have used the Kuramoto–Sakaguchi equation (1.2) and integration by parts. Notice that by definition of the order parameter (1.15), we obtain

$$R\cos(\theta-\phi) = \int_{\mathbb{T}\times\mathbb{R}}\cos(\theta-\theta')f(t,\theta',\omega')\,d\theta'\,d\omega'. \tag{3.2}$$

Using this identity in the above formula for I_2 implies

$$I_2 = -2K \int_{\mathbb{T}^2\times\mathbb{R}^2} (\omega - KR\sin(\theta-\phi))^2 \cos(\theta-\theta') \times f(t,\theta,\omega) f(t,\theta',\omega')\, d\theta\, d\theta'\, d\omega\, d\omega'. \tag{3.3}$$

Let us now change variables (θ,ω) with (θ',ω') in (3.3) and take the mean value of both expressions for I_2. Since the cosine is an even function, we equivalently write

$$I_2 = -K \int_{\mathbb{T}^2\times\mathbb{R}^2} ((\omega - KR\sin(\theta-\phi))^2 + (\omega' - KR\sin(\theta'-\phi))^2) \times \cos(\theta-\theta') f(t,\theta,\omega) f(t,\theta',\omega')\, d\theta\, d\theta'\, d\omega\, d\omega'. \tag{3.4}$$

Finally, putting (3.1) and (3.4) together and completing the square yields the desired result. ■

As a consequence of the previous theorem, we obtain the following quantitative behavior of the dissipation.

Corollary 3.1. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W,W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Then*

$$-2KR\,\mathcal{I}[f] \le \frac{d}{dt}\mathcal{I}[f] \le 2K\,\mathcal{I}[f] \tag{3.5}$$

for all $t \ge 0$. In particular,

$$\mathcal{I}[f](t_0) e^{-2K\int_{t_0}^t R(s)\,ds} \le \mathcal{I}[f](t) \le \mathcal{I}[f](t_0) e^{2K(t-t_0)}$$

for all $t \ge t_0 \ge 0$.

Proof. Note that the second chain of inequalities follows from integration of (3.5) with respect to time. Then we focus on the proof of (3.5), which we divide into two steps associated with the upper bound and lower bound respectively.

Step 1: Upper bound. Using Theorem 3.1 and bounding $\cos(\theta-\theta')$ by 1, we achieve the following upper bound for the derivative of the dissipation functional along f:

$$\begin{aligned}\frac{d}{dt}\mathcal{I}[f] &\le \int_{\mathbb{T}^2\times\mathbb{R}^2} \big(\omega - KR\sin(\theta-\phi) - (\omega' - KR\sin(\theta'-\phi))\big)^2 \\ &\qquad \times f(t,\theta,\omega) f(t,\theta',\omega')\, d\theta\, d\theta'\, d\omega\, d\omega' \\ &= 2K\int_{\mathbb{T}\times\mathbb{R}} (\omega - KR\sin(\theta-\phi))^2 f\, d\theta\, d\omega \\ &\quad - 2K\Big(\int_{\mathbb{T}\times\mathbb{R}} (\omega - KR\sin(\theta-\phi)) f\, d\theta\, d\omega\Big)^2.\end{aligned}$$

Using definition (1.15) of R and ϕ along with assumption (1.16), we clearly obtain that the second term vanishes and we conclude the upper bound.

Step 2: Lower bound. Again, we will use Theorem 3.1 and expand the square to obtain

$$\begin{aligned}
\frac{d}{dt}\mathcal{I}[f] = -2K\int_{\mathbb{T}^2\times\mathbb{R}^2} &(\omega - KR\sin(\theta-\phi))^2\cos(\theta-\theta') \\
&\times f(t,\theta,\omega)f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega' \\
+ 2K\int_{\mathbb{T}^2\times\mathbb{R}^2} &(\omega - KR\sin(\theta-\phi))(\omega' - KR\sin(\theta'-\phi))\cos(\theta-\theta') \\
&\times f(t,\theta,\omega)f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega' \\
= -2KR\int_{\mathbb{T}\times\mathbb{R}} &(\omega - KR\sin(\theta-\phi))^2\cos(\theta-\phi)\,f\,d\theta\,d\omega \\
+ 2K\bigg|\int_{\mathbb{T}\times\mathbb{R}} &(\omega - KR\sin(\theta-\phi))e^{i(\theta-\phi)}\,f\,d\theta\,d\omega\bigg|^2 \\
\geq -2KR\int_{\mathbb{T}\times\mathbb{R}} &(\omega - KR\sin(\theta-\phi))^2\,f\,d\theta\,d\omega,
\end{aligned}$$

where in the second identity we have used (3.2) while in the last inequality we have bounded $\cos(\theta-\theta')$ by 1 and we have neglected the non-negative term. Hence, the desired result follows. ∎

3.1. A fibered Wasserstein distance and relation to dissipation

In this section we introduce a Wasserstein-type distance in the product space $\mathbb{T}\times\mathbb{R}$ that will play an essential role in the aforementioned dissipation–transportation inequality. This metric is constructed through a gluing procedure of the standard quadratic Wasserstein distance in $\mathbb{T}$ between conditional probabilities at any fiber $\omega\in\mathbb{R}$. Since it behaves in a fiberwise way, we call it the fibered quadratic Wasserstein distance. For the convenience of the reader, we recall it here and introduce some of the main properties that will be used throughout the paper. See also [43, 49] for further details.

Definition 3.1 (Fibered quadratic Wasserstein distance). Consider any probability measure $g\in\mathbb{P}(\mathbb{R})$ and let us define the closed subset of those probability measures $\mathbb{T}\times\mathbb{R}$ whose ω-marginal agrees with g, i.e.,

$$\mathbb{P}_g(\mathbb{T}\times\mathbb{R}) := \big\{\mu\in\mathbb{P}(\mathbb{T}\times\mathbb{R}) : (\pi_\theta)_{\#}\mu = g\big\}.$$

We define the fibered quadratic Wasserstein distance on $\mathbb{P}_g(\mathbb{T}\times\mathbb{R})$ as

$$W_{2,g}(\mu,\nu) := \left(\int_{\mathbb{R}} W_2(\mu(\cdot|\omega),\nu(\cdot|\omega))^2\,d_\omega g\right)^{1/2} \tag{3.6}$$

for any $\mu,\nu\in\mathbb{P}_g(\mathbb{T}\times\mathbb{R})$. Here, we denote the family of conditional probabilities (or disintegrations) of μ with respect to the fiber $\omega\in\mathbb{R}$ as

$$\omega\in\mathbb{R}\mapsto\mu(\cdot|\omega)\in\mathbb{P}(\mathbb{T}),$$

which is a Borel-measurable function defined by the formula

$$\int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta,\omega)\, d_{(\theta,\omega)}\mu = \int_{\mathbb{R}} \left(\int_{\mathbb{T}} \varphi(\theta,\omega)\, d_\theta \mu(\cdot|\omega) \right) d_\omega g$$

for any test function $\varphi \in C_b(\mathbb{T}\times\mathbb{R})$.

As for the classical quadratic Wasserstein distance, this distance also admits an equivalent Benamou–Brenier representation (see [4]), which can be obtained by gluing the corresponding representations at any fiber.

Proposition 3.1. *Consider $g \in \mathbb{P}(\mathbb{R})$ and let $f^1, f^2 \in \mathbb{P}_g(\mathbb{T}\times\mathbb{R})$. For g-a.e. value of $\omega \in \mathbb{R}$, let us consider some Wasserstein geodesic $\tau \in [0,1] \mapsto h_\tau(\cdot|\omega) \in \mathbb{P}(\mathbb{T})$ that joins the conditional probabilities with respect to ω, that is,*

$$h_{\tau=0}(\theta) = f^1(\cdot|\omega) \quad \textit{and} \quad h_{\tau=1}(\theta) = f^2(\cdot|\omega).$$

This is an absolutely continuous family with respect to the Wasserstein distance on $\mathbb{T}$ and it has an associated family of potentials $\tau \in [0,1] \mapsto \psi_\tau(\cdot,\omega)$ so that

$$\begin{cases} \dfrac{\partial}{\partial\tau} h_\tau(\cdot|\omega) + \operatorname{div}_\theta(\nabla_\theta \psi_\tau(\cdot,\omega) h_\tau(\cdot|\omega)) = 0, \\ \dfrac{\partial}{\partial\tau}\psi_\tau(\cdot,\omega) + \dfrac{1}{2}|\nabla_\theta \psi_\tau(\cdot,\omega)|^2 = 0, \ \psi_{\tau=0}(\cdot,\omega) = \psi_0(\cdot,\omega), \end{cases} \tag{3.7}$$

for some $\frac{d^2}{2}$-concave function $-\psi_0$ with respect to θ, in the distributional/viscosity sense. Then the following identity holds true:

$$W_{2,g}(f^1,f^2)^2 = \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} |\nabla_\theta \psi_\tau|^2\, dh_\tau\, d\tau, \tag{3.8}$$

where we denote by h_τ the measure that can be recovered from the conditional probabilities $h_\tau(\cdot|\omega)$ with marginal g, that is, for any test function $\varphi \in C_b(\mathbb{T}\times\mathbb{R})$ the following disintegration formula holds:

$$\int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta,\omega)\, dh_\tau = \int_{\mathbb{R}} \left(\int_{\mathbb{T}} \varphi(\theta,\omega)\, d_\theta h_\tau(\cdot|\omega) \right) d_\omega g.$$

Since the proof is a simple gluing procedure applied to the classical result for the quadratic Wasserstein distance, we skip it. The interested reader may want to get further details from [4] and the textbooks [2] and [57, Chapter 13].

Remark 3.1. The second equation in (3.7) is called the Hamilton–Jacobi equation, and using it we observe that (3.8) can be restated as

$$W_{2,g}(f^1,f^2)^2 = \int_{\mathbb{T}\times\mathbb{R}} |\nabla_\theta \psi_\tau|^2 h_\tau\, d\theta\, d\omega \tag{3.9}$$

for every $\tau \in [0,1]$. This suggests that such Wasserstein geodesics have constant speed.

An interesting fact is that this new fibered quadratic Wasserstein distance and the classical quadratic Wasserstein distances in $\mathbb{P}_2(\mathbb{T} \times \mathbb{R})$ are appropriately ordered. Before we state the relation, let us remark on the following fact.

Remark 3.2. The classical quadratic Wasserstein distance W_2 in $\mathbb{P}_2(\mathbb{T} \times \mathbb{R})$ is defined as the transportation cost associated with the standard Riemannian distance in the product space $\mathbb{T} \times \mathbb{R}$. That is, W_2 is defined by

$$W_2(\mu_0^N, f_0) = \left(\inf_{\gamma \in \Pi(\mu_0^N, f_0)} \int_{\mathbb{T}^2 \times \mathbb{R}^2} (d(\theta, \theta')^2 + (\omega - \omega')^2) \, d\gamma \right)^{1/2}$$

for any $\mu, \nu \in \mathbb{P}_2(\mathbb{T} \times \mathbb{R})$. Here, $d(\theta, \theta')$ denotes the canonical Riemannian distance between any two points θ and θ' in $\mathbb{T}$.

For our purposes, such a distance is not appropriate as it is not dimensionally correct. Indeed, θ and ω have different physical units and considering W_2 causes problems in deriving the asymptotic behavior of solutions.

The above remark suggests considering the following correction of the classical quadratic Wasserstein distance in $\mathbb{P}_2(\mathbb{T} \times \mathbb{R})$.

Definition 3.2 (Scaled quadratic Wasserstein distance). Let us consider the scaled Riemannian distance on the product space $\mathbb{T} \times \mathbb{R}$, i.e.,

$$d_K((\theta, \omega), (\theta', \omega')) = \left(d(\theta, \theta')^2 + \frac{(\omega - \omega')^2}{K^2} \right)^{\frac{1}{2}}.$$

We define the scaled quadratic Wasserstein distance on $\mathbb{P}_2(\mathbb{T} \times \mathbb{R})$ by the transportation cost associated with the above scaled Riemannian distance, that is,

$$\mathrm{SW}_2(\mu_0^N, f_0) = \left(\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{T}^2 \times \mathbb{R}^2} \left(d(\theta, \theta')^2 + \frac{(\omega - \omega')^2}{K^2} \right) d\gamma \right)^{1/2}$$

for any $\mu, \nu \in \mathbb{P}_2(\mathbb{T} \times \mathbb{R})$.

We are now ready to state the relation between SW_2 and W_2.

Proposition 3.2. *Consider $g \in \mathbb{P}_2(\mathbb{T})$. Then we obtain*

$$\mathrm{SW}_2(\mu, \nu) \le W_{2,g}(\mu, \nu)$$

for any $\mu, \nu \in \mathbb{P}_g(\mathbb{T} \times \mathbb{R})$. In particular, we have

$$W_2(\mu, \nu) \le W_{2,g}(\mu, \nu).$$

Proof. Consider for g-a.e. $\omega \in \mathbb{R}$ the optimal coupling $\gamma_{0,\omega} \in \Pi(\mu(\cdot|\omega), \nu(\cdot|\omega))$ between the conditional probabilities $\mu(\cdot|\omega)$ and $\nu(\cdot|\omega)$. Then we can construct the probability measure $\gamma \in \mathbb{P}(\mathbb{T}^2 \times \mathbb{R}^2)$ given by

$$\gamma := \gamma_{0,\omega}(\theta, \theta') \otimes \delta_\omega(\omega') \otimes g(\omega). \tag{3.10}$$

Let us see first that it defines a transference plan, that is, $\gamma \in \Pi(\mu,\nu)$. To this end, consider any test function $\varphi \in C_b(\mathbb{T}\times\mathbb{R})$ and note that

$$\begin{aligned}\int_{\mathbb{T}\times\mathbb{R}} \varphi\, d_{(\theta,\omega)}(\pi_{(\theta,\omega)\#}\gamma) &= \int_{\mathbb{T}^2\times\mathbb{R}^2} \varphi(\theta,\omega)\, d_{(\theta,\theta')}\gamma_{0,\omega}\, d_{\omega'}(\delta_\omega)\, d_\omega g\\ &= \int_{\mathbb{T}^2\times\mathbb{R}} \varphi(\theta,\omega)\, d_{(\theta,\theta')}\gamma_{0,\omega}\, d_\omega g\\ &= \int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta,\omega)\, d_\theta(\pi_{\theta\#}\gamma_{0,\omega})\, d_\omega g\\ &= \int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta,\omega)\, d_\theta\mu(\cdot|\omega)\, d_\omega g = \int_{\mathbb{T}\times\mathbb{R}} \varphi\, d_{(\theta,\omega)}\mu.\end{aligned}$$

Then $\pi_{(\theta,\omega)\#}\gamma = \mu$. Similarly, note that

$$\begin{aligned}\int_{\mathbb{T}\times\mathbb{R}} \varphi\, d_{(\theta',\omega')}(\pi_{(\theta',\omega')\#}\gamma) &= \int_{\mathbb{T}^2\times\mathbb{R}^2} \varphi(\theta',\omega')\, d_{(\theta,\theta')}\gamma_{0,\omega}\, d_{\omega'}(\delta_\omega)\, d_\omega g\\ &= \int_{\mathbb{T}^2\times\mathbb{R}} \varphi(\theta',\omega)\, d_{(\theta,\theta')}\gamma_{0,\omega}\, d_\omega g\\ &= \int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta',\omega)\, d_{\theta'}(\pi_{\theta'\#}\gamma_{0,\omega})\, d_\omega g\\ &= \int_{\mathbb{T}\times\mathbb{R}} \varphi(\theta',\omega)\, d_{\theta'}\nu(\cdot|\omega)\, d_\omega g = \int_{\mathbb{T}\times\mathbb{R}} \varphi\, d_{(\theta',\omega')}\nu.\end{aligned}$$

Then we also recover $\pi_{(\theta',\omega')\#}\gamma = \nu$. Also note that, by definition,

$$\begin{aligned}W_{2,g}(\mu,\nu)^2 &= \int_{\mathbb{R}\times\mathbb{T}^2} d(\theta,\theta')^2\, d_{(\theta,\theta')}\gamma_{0,\omega}\, d_\omega g = \int_{\mathbb{T}^2\times\mathbb{R}^2} d(\theta,\theta')^2\, d_{((\theta,\omega),(\theta',\omega'))}\gamma\\ &= \int_{\mathbb{T}^2\times\mathbb{R}^2} d_K((\theta,\omega),(\theta',\omega'))\, d_{((\theta,\omega),(\theta',\omega'))}\gamma \geq \mathrm{SW}_2(\mu,\nu)^2,\end{aligned}$$

where the extra term that has been added in the second line vanishes because of the presence of $\delta_\omega(\omega')$ in (3.10). ■

Indeed, the scaled and fibered Wasserstein distances are strictly ordered.

Remark 3.3. Consider the empirical measures

$$\mu := \frac{1}{2}(\delta_{(\theta_1,\omega_1)} + \delta_{(\theta_2,\omega_2)}) \quad \text{and} \quad \nu := \frac{1}{2}(\delta_{(\theta_2,\omega_1)} + \delta_{(\theta_1,\omega_2)})$$

for some $\theta_1,\theta_2 \in \mathbb{T}$ and $\omega_1,\omega_2 \in \mathbb{R}$. Notice that

$$\pi_{\omega\#}\mu = \pi_{\omega\#}\nu = \frac{1}{2}(\delta_{\omega_1} + \delta_{\omega_2}) =: g,$$

thus, $\mu,\nu \in \mathbb{P}_g(\mathbb{T}\times\mathbb{R})$. Finally, for $\varepsilon_\theta := d(\theta_1,\theta_2)$ and $\varepsilon_\omega := |\omega_1-\omega_2|$ it is clear that

$$W_{2,g}(\mu,\nu)^2 = \varepsilon_\theta^2 \quad \text{and} \quad \mathrm{SW}_2(\mu,\nu)^2 = \frac{1}{K^2}\min\left\{\varepsilon_\theta^2, \frac{\varepsilon_\omega^2}{K^2}\right\}.$$

Consequently, we obtain

$$\begin{aligned}\mathrm{SW}_2(\mu,\nu) &< W_{2,g}(\mu,\nu) \quad \text{if } \frac{\varepsilon_\omega}{K} < \varepsilon_\theta,\\ \mathrm{SW}_2(\mu,\nu) &= W_{2,g}(\mu,\nu) \quad \text{if } \frac{\varepsilon_\omega}{K} \geq \varepsilon_\theta.\end{aligned}$$

We are now ready to state the main relation between this fibered transportation distance (3.6) and the dissipation functional (2.1).

Lemma 3.1. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W,W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Then*

$$\frac{d}{ds}\frac{1}{2}W_{2,g}(f_t,f_s)^2 \leq \mathcal{I}[f]^{\frac{1}{2}}W_{2,g}(f_t,f_s)$$

for every $t \geq 0$ and almost every $s \geq 0$.

A similar result was explored in [49, Theorem 4.4]. There, the author used the definition of $W_{2,g}$ in (3.6) for general measures that may enjoy atoms eventually. In this result, we sketch a simpler proof that uses the representation formula of the derivative of Wasserstein distance for absolutely continuous measures; see [2, Theorem 8.4.6], [57, Theorem 23.9].

Proof of Lemma 3.1. Since f satisfies the Kuramoto–Sakaguchi equation (1.2), then each conditional probability with respect to $\omega \in \mathbb{T}$ verifies the continuity equation

$$\frac{\partial}{\partial t}f(\theta|\omega) + \mathrm{div}_\theta\big((\omega - KR\sin(\theta-\phi))e^{i\theta}f(\theta|\omega)\big) = 0$$

for all $t \geq 0$ and $\theta \in \mathbb{T}$. That is, the disintegrations themselves are driven by the tangent transport field

$$\theta \in \mathbb{T} \mapsto v_t^\omega(\theta) := (\omega - KR\sin(\theta-\phi))e^{i\theta}.$$

Since f is smooth, it is clear that the family $s \in [0,+\infty) \mapsto f_s(\cdot|\omega)$ is locally absolutely continuous with respect to the quadratic Wasserstein distance on $\mathbb{T}$. This clearly guarantees that the following function is also locally absolutely continuous:

$$s \in [0,+\infty) \mapsto W_2(f_t(\cdot|\omega), f_s(\cdot|\omega))^2$$

for every $\omega \in \operatorname{supp} g$; see [2, Theorem 8.4.6] or [57, Theorem 23.9]. In particular, we can take derivatives almost everywhere and obtain the formula

$$\frac{d}{ds}\frac{1}{2}W_2(f_t(\cdot|\omega), f_s(\cdot|\omega))^2 = -\int_{\mathbb{T}}\langle v_s^\omega(\theta), \nabla\psi_{\tau=0}^{s,t}(\theta,\omega)\rangle f_s(\theta|\omega)\,d\theta \tag{3.11}$$

for almost every $t \geq 0$, where the family $\tau \in [0,1] \mapsto (h_\tau^{s,t}, \psi_\tau^{s,t})$ has been chosen according to (3.7) so that it represents a Wasserstein geodesic joining the conditional probabilities of f_s to those of f_t. By the dominated convergence theorem, we can then show that the following function is also absolutely continuous:

$$s \in [0, +\infty) \mapsto W_{2,g}(f_t, f_s)^2.$$

Integrating by parts and using (3.11) we obtain

$$\begin{aligned}\frac{d}{ds}\frac{1}{2}W_{2,g}(f_t, f_s)^2 &= -\int_{\mathbb{T}\times\mathbb{R}} \langle v_s^\omega(\theta), \nabla\psi_{\tau=0}^{s,t}(\theta,\omega)\rangle f_s(\theta|\omega)g(\omega)\,d\theta\,d\omega\\ &= -\int_{\mathbb{T}\times\mathbb{R}} \langle v_s^\omega(\theta), \nabla\psi_{\tau=0}^{s,t}(\theta,\omega)\rangle f_s(\theta,\omega)\,d\theta\,d\omega. \end{aligned} \tag{3.12}$$

Using the Cauchy–Schwarz inequality in (3.12) along with the definition of the dissipation functional (2.1) and the representation of the fibered quadratic Wasserstein distance in Proposition 3.1 we obtain

$$\frac{d}{ds}\frac{1}{2}W_{2,g}(f_t, f_s)^2 \le \mathcal{I}[f]^{\frac{1}{2}}W_{2,g}(f_t, f_s)$$

for almost every $s \geq 0$. Hence, the desired result follows. ∎

As a direct consequence of the above lemma, we obtain the following dissipation–transportation inequality.

Corollary 3.2. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Then*

$$W_{2,g}(f_t, f_s) \le \int_t^s \mathcal{I}[f_\tau]^{1/2}\,d\tau \quad \textit{for all } s \geq t.$$

3.2. Convergence and uniqueness of the global equilibria

In this section we will prove the main Theorem 1.1 on convergence to the global equilibria in the large coupling strength regime. Before we proceed with the proof, let us first show that such an equilibrium is indeed unique up to phase rotations. We remark that this result is already known in the literature.

On the one hand, the direct classical proof follows from the characterization of equilibria f_∞ with $R_\infty > 0$ in Section 1.2. Specifically, assume $\operatorname{diam}(\operatorname{supp}_\theta f_\infty) < \pi$ and $\operatorname{supp} g \subset [-W, W]$. Then $g^- = 0$ in (1.18), meaning that there is no antipodal mass, and the compatibility conditions (1.17) and (1.19) reduce to

$$F(R_\infty) := K\int_{-1}^1 \sqrt{1-s^2}\,g(KR_\infty s)\,ds - 1 = 0, \quad R_\infty \in \Big[\frac{W}{K}, 1\Big].$$

We note that F is a continuous function and it verifies

$$F(1) \le K\int_{-1}^1 g(Ks)\,ds - 1 = 0, \quad F\Big(\frac{W}{K}\Big) = \frac{K}{W}\int_{-W}^W \sqrt{1-\Big(\frac{\omega}{W}\Big)^2}\,g(\omega)\,d\omega - 1.$$

In addition, we have $\lim_{\frac{W}{K}\to 0^+} F(\frac{W}{K}) = +\infty$. Hence, in the large coupling strength regime there must exist at least one $R_\infty > 0$ solving the above implicit equation, which determines a stationary state f_∞ via (1.18). The uniqueness of the equilibrium (up to phase rotations) is clear by virtue of the strict monotonicity of F. Specifically, by differentiation and integrating by parts we have

$$F'(R_\infty) = -K^2 \int_{\mathbb{R}} \sqrt{1 - \Big(\frac{\omega}{KR_\infty}\Big)^2}\, g(\omega)\, d\omega - \frac{1}{R_\infty^3} \int_{\mathbb{R}} \frac{\omega^2}{\sqrt{1-(\frac{\omega}{KR_\infty})^2}} g(\omega)\, d\omega < 0.$$

On the other hand, an indirect proof follows from [13], where the authors derived a strict contractivity estimate for an appropriately modified Wasserstein distance $\widetilde{W}_p$ in $\mathbb{P}_2([0,2\pi)\times\mathbb{R})$ along any couple of solutions lying in the above region of convexity (oscillators confined to a half-circle). Such an estimate yields uniqueness of equilibria and relaxation to equilibria at once, provided that initial data belong to the basin of attraction and coupling strength is large enough. We remark that the geometry of $\mathbb{T}$ was disregarded in that result when defining $\widetilde{W}_p$. Namely, the geometry of $\mathbb{T}$ decreases the transportation cost of mass between phases separated by distances larger than π (when viewed in the real line), so that $\widetilde{W}_2$ becomes strictly larger than our proposed fibered distance $W_{2,g}$.

For self-consistency of our presentation, we provide an alternative proof that exploits our proposed fibered distance $W_{2,g}$ and is based on a similar convexity property. It does not exploit the explicit structure of equilibria so that the method of proof could be interesting on its own to address other settings with less suitable algebraic structure. We leave the full study of similar strict contractivity of $W_{2,g}$ as in [13] to future works.

Proposition 3.3. *Let f_∞ and f'_∞ be stationary measure-valued solutions to (1.2) and assume that they have the same distribution g of natural frequencies, and* $\operatorname{diam}(\operatorname{supp}_\theta f_\infty)$ *and* $\operatorname{diam}(\operatorname{supp}_\theta f'_\infty)$ *are smaller than $\pi/2$. Then they agree up to phase rotations, that is, there exists a constant $c \in \mathbb{R}$ such that*

$$f'_\infty(\theta,\omega) = f_\infty(\theta - c, \omega).$$

Proof. For any $c \in \mathbb{R}$ we consider the rotation operator in the variable θ,

$$\mathcal{T}_c[f'_\infty](\theta,\omega) := f'_\infty(\theta - c, \omega),$$

and define the optimization problem

$$\min_{c\in\mathbb{R}} W_{2,g}(f_\infty, \mathcal{T}_c[f'_\infty])^2. \tag{3.13}$$

The minimum of (3.13) exists from straightforward arguments and will be achieved at some $c = c_0 \in \mathbb{R}$. Without loss of generality, let us assume that $c_0 = 0$. Indeed, otherwise we can replace f'_∞ with $\mathcal{T}_{c_0}[f'_\infty]$ and it does not change the thesis of this result. On the one hand, let us consider the continuity equation

$$\begin{cases} \dfrac{\partial}{\partial s} f'_s + \operatorname{div}_\theta(e^{i\theta} f'_s) = 0, \\ f'_{s=0} = f'_\infty, \end{cases} \tag{3.14}$$

whose solution clearly describes the above family of phase shifts, namely $f_s' = \mathcal{T}_s[f_\infty']$. Since $W_{2,g}(f_\infty, f_\infty')$ minimizes problem (3.13), we obtain a critical value at $s = 0$, i.e.,

$$\frac{d}{ds}\bigg|_{s=0} W_{2,g}(f_\infty, f_s')^2 = 0. \tag{3.15}$$

Let us write down condition (3.15) more explicitly. Indeed, consider a Wasserstein geodesic that joins the conditional probability $f_\infty'(\cdot|\omega)$ to $f_s'(\cdot|\omega)$ and represents it through a family

$$\tau \in [0,1] \to (h_\tau^s, \psi_\tau^s) \quad \text{with} \quad \begin{aligned} h_{\tau=0}^s(\cdot|\omega) &= f_\infty'(\cdot|\omega), \\ h_{\tau=1}^s(\cdot|\omega) &= f_s'(\cdot|\omega), \end{aligned} \tag{3.16}$$

as in (3.7) in Proposition 3.1. Here, although (3.7) holds only in the viscosity/distribution sense, this fact can be handled nowadays by standard regularization arguments; we refer the reader to [57, Chapter 13]. (In particular, our dissipation functional $\mathcal{I}[f]$ is continuous with respect to $W_{2,g}$ which makes it well behaved with respect to regularizations.)

Now observe that, by construction, $f_s'(\cdot|\omega)$ verifies the continuity equation (3.14) that is driven by the trivial tangent transport field $\theta \in \mathbb{T} \mapsto e^{i\theta}$. Then, using the same ideas as in the proof of Lemma 3.1 (see [2, Theorem 8.4.6] or [57, Theorem 23.9]), we obtain

$$\frac{d}{ds}\bigg|_{s=0} \frac{1}{2} W_2(f_\infty(\cdot|\omega), f_s'(\cdot|\omega))^2 = \int_{\mathbb{T}} \langle e^{i\theta}, \nabla_\theta \psi_{\tau=1}^{s=0}(\theta,\omega)\rangle \, d_\theta f_\infty'(\cdot|\omega)$$

for almost every $s \geq 0$. Taking integrals in ω against g and using (3.15) we obtain

$$\int_{\mathbb{T}\times\mathbb{R}} \langle e^{i\theta}, \nabla_\theta \psi_{\tau=1}^{s=0}\rangle \, d_{(\theta,\omega)} f_\infty' = 0.$$

Indeed, using the equations for $h_\tau^{s=0}$ and $\varphi_\tau^{s=0}$ in (3.7), it is clear that the above implies

$$\int_{\mathbb{T}\times\mathbb{R}} \langle e^{i\theta}, \nabla_\theta \psi_\tau^{s=0}\rangle \, d_{(\theta,\omega)} h_\tau^{s=0} = 0 \tag{3.17}$$

for every $\tau \in [0,1]$. On the other hand, by hypothesis f_∞ and f_∞' verify the (stationary) Kuramoto–Sakaguchi equation (1.2), that is,

$$\begin{aligned} &\frac{\partial}{\partial t} f_\infty + \operatorname{div}_\theta\big((\omega - KR_\infty \sin(\theta - \phi_\infty))e^{i\theta} f_\infty\big) = 0, \\ &\frac{\partial}{\partial t} f_\infty' + \operatorname{div}_\theta\big((\omega - KR_\infty' \sin(\theta - \phi_\infty'))e^{i\theta} f_\infty'\big) = 0. \end{aligned}$$

Since the solutions are stationary, we can use the same ideas as before to arrive at the identity

$$\begin{aligned} 0 &= \frac{d}{dt}\frac{1}{2} W_2(f_\infty(\cdot|\omega), f_\infty'(\cdot|\omega))^2 \\ &= \int_{\mathbb{T}} \langle (\omega - KR_\infty' \sin(\theta - \phi_\infty'))e^{i\theta}, \nabla_\theta \psi_{\tau=1}^{s=0}(\cdot,\omega)\rangle \, d_\theta f_\infty'(\cdot|\omega) \\ &\quad - \int_{\mathbb{T}} \langle (\omega - KR_\infty \sin(\theta - \phi_\infty))e^{i\theta}, \nabla_\theta \psi_{\tau=0}^{s=0}(\cdot,\omega)\rangle \, d_\theta f_\infty(\cdot|\omega). \end{aligned}$$

From here on we will omit the superscripts $s = 0$ of $h_\tau^{s=0}$ and $\psi_\tau^{s=0}$ for simplicity, as they are clear from the context. Then integrating against g and using the fundamental theorem of calculus in τ yields

$$\int_0^1 \frac{d}{d\tau} \int_{\mathbb{T}\times\mathbb{R}} \langle (\omega - KR_\tau \sin(\theta - \phi_\tau))e^{i\theta}, \nabla_\theta \varphi_\tau \rangle \, d_{(\theta,\omega)} h_\tau \, d\tau = 0, \tag{3.18}$$

where R_τ and ϕ_τ are order parameters associated with the displacement interpolation h_τ. Let us now expand the derivative in (3.18) and use the Hamilton–Jacobi equation for ψ_τ and the continuity equation for h_τ in (3.7) (see [57, Chapter 13]). Then we obtain

$$A + B + C = 0,$$

where the terms read

$$\begin{aligned} A &:= \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \Big\langle \nabla_\theta \Big(-\frac{1}{2} |\nabla_\theta \psi_\tau|^2 \Big), (\omega - KR_\tau \sin(\theta - \phi_\tau))e^{i\theta} \Big\rangle \, d_{(\theta,\omega)} h_\tau \, d\tau, \\ B &:= \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \Big\langle \frac{d}{d\tau} [\omega - KR_\tau \sin(\theta - \phi_\tau)] e^{i\theta}, \nabla_\theta \psi_\tau \Big\rangle \, d_{(\theta,\omega)} h_\tau \, d\tau, \\ C &:= \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \big\langle \nabla_\theta \langle \nabla_\theta \psi_\tau, (\omega - KR_\tau \sin(\theta - \phi_\tau))e^{i\theta} \rangle, \nabla_\theta \psi_\tau \big\rangle \, d_{(\theta,\omega)} h_\tau \, d\tau. \end{aligned}$$

On the one hand, taking the sum of A and C we can simplify to

$$\begin{aligned} A + C &= -K \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} R_\tau \cos(\theta - \phi_\tau) |\nabla_\theta \psi_\tau|^2 \, d_{(\theta,\omega)} h_\tau \, d\tau \\ &= -K \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \int_{\mathbb{T}\times\mathbb{R}} \cos(\theta - \theta') |\nabla_\theta \psi_\tau|^2 \, d_{(\theta,\omega)} h_\tau \, d_{(\theta',\omega')} h_\tau \, d\tau \\ &= -\frac{K}{2} \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \int_{\mathbb{T}\times\mathbb{R}} \cos(\theta - \theta') \big(|\nabla_\theta \psi_\tau(\theta, \omega)|^2 + |\nabla_\theta \psi_\tau(\theta', \omega')|^2 \big) \\ &\qquad \times d_{(\theta,\omega)} h_\tau \, d_{(\theta',\omega')} h_\tau \, d\tau, \end{aligned} \tag{3.19}$$

where in the second line we have used the properties of the order parameters R_τ and ϕ_τ of the interpolation h_τ, namely,

$$\begin{aligned} R_\tau &= \int_{\mathbb{T}\times\mathbb{R}} \cos(\theta' - \phi_\tau) \, d_{(\theta',\omega')} h_\tau, \\ 0 &= \int_{\mathbb{T}\times\mathbb{R}} \sin(\theta' - \phi_\tau) \, d_{(\theta',\omega')} h_\tau, \end{aligned}$$

and in the third and fourth lines we have used a clear symmetrization argument. Let us now differentiate with respect to τ and use the continuity equation for h_τ to obtain the formulas

$$\begin{aligned} \frac{dR_\tau}{d\tau} &= -\int_{\mathbb{T}\times\mathbb{R}} \sin(\theta' - \phi_\tau) \langle e^{i\theta'}, \nabla_\theta \psi_\tau(\theta', \omega') \rangle \, d_{(\theta',\omega')} h_\tau, \\ R_\tau \frac{d\phi_\tau}{d\tau} &= \int_{\mathbb{T}\times\mathbb{R}} \cos(\theta' - \phi_\tau) \langle e^{i\theta'}, \nabla_\theta \psi_\tau(\theta', \omega') \rangle \, d_{(\theta',\omega')} h_\tau. \end{aligned}$$

Then the term B can be written as

$$\begin{aligned}B &= \int_0^1 \int_{\mathbb{T}\times\mathbb{R}} \langle e^{i\theta}, \nabla_\theta \psi_\tau\rangle \frac{d}{d\tau}\Big(-K\frac{dR_\tau}{d\tau}\sin(\theta-\phi_\tau) + KR_\tau \frac{d\phi_\tau}{d\tau}\cos(\theta-\phi_\tau)\Big)\\
&\qquad\qquad \times d_{(\theta,\omega)}h_\tau \, d\tau\\
&= K\int_0^1 \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} \cos(\theta-\theta')\langle e^{i\theta}, \nabla_\theta\psi_\tau(\theta,\omega)\rangle\langle e^{i\theta'}, \nabla_\theta\psi_\tau(\theta',\omega')\rangle\\
&\qquad\qquad\qquad \times d_{(\theta,\omega)}h_\tau \, d_{(\theta',\omega')}h_\tau \, d\tau. \end{aligned} \tag{3.20}$$

Putting formulas (3.19) and (3.20) into (3.18) gives

$$\begin{aligned}0 = -\frac{K}{2}\int_0^1 \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} \cos(\theta-\theta')\big(\langle e^{i\theta}, \nabla_\theta\psi_\tau(\theta,\omega)\rangle - \langle e^{i\theta'}, \nabla_\theta\psi_\tau(\theta',\omega')\rangle\big)^2\\
\times d_{(\theta,\omega)}h_\tau \, d_{(\theta',\omega')}h_\tau \, d\tau. \end{aligned} \tag{3.21}$$

Since there exists $0<\delta<\pi/2$ such that

$$\operatorname{diam}(\operatorname{supp}_\theta f_\infty) < \delta \quad \text{and} \quad \operatorname{diam}(\operatorname{supp}_\theta f'_\infty) < \delta,$$

then the same is true for the interpolations h_τ. Indeed, this is a consequence of the monotone rearrangement property of the 1-dimensional transport on each fiber. Hence, we can take upper bounds in (3.21) and obtain

$$\begin{aligned}0 &\leq -\frac{K}{2}\cos(\delta)\int_0^1 \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} \big(\langle e^{i\theta}, \nabla_\theta\psi_\tau(\theta,\omega)\rangle - \langle e^{i\theta'}, \nabla_\theta\psi_\tau(\theta',\omega')\rangle\big)^2\\
&\qquad\qquad\qquad \times d_{(\theta,\omega)}h_\tau \, d_{(\theta',\omega')}h_\tau \, d\tau\\
&= -K\cos(\delta)\int_0^1\int_{\mathbb{T}\times\mathbb{R}} |\nabla_\theta\psi_\tau|^2 \, d_{(\theta,\omega)}h_\tau \, d\tau\\
&\quad + K\cos(\delta)\int_0^1 \left(\int_{\mathbb{T}\times\mathbb{R}} \langle e^{i\theta}, \nabla_\theta\psi_\tau\rangle \, d_{(\theta,\omega)}h_\tau\right)^2 d\tau.\end{aligned}$$

Notice that condition (3.17) allows the second term to be neglected. Also, notice that the cosine has positive sign and hence

$$\nabla_\theta \psi_\tau^{s=0} = 0 \quad \text{for } d\tau\otimes h_\tau^{s=0}\text{-a.e. } (\tau,\theta,\omega)\in[0,1]\times\mathbb{T}\times\mathbb{R}.$$

In particular, the continuity equation for $h_\tau^{s=0}$ implies that

$$f_\infty = h_\tau^{s=0} = f'_\infty \quad \text{for all } \tau\in[0,1],$$

thus ending the proof. ∎

We now come back to the proof of Theorem 1.1. First, we show that once the concentration regime in Lemma 1.1 takes place, Theorem 3.1 guarantees that the dissipation decays exponentially fast.

Corollary 3.3. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W,W]$ and centered (i.e., (1.16)). Consider the unique global-in-time classical solution $f=f(t,\theta,\omega)$ to (1.2). Then the following holds true:*

$$\frac{d\mathcal{I}[f]}{dt}\leq -2K\cos(\beta)\,\mathcal{I}[f]+24K(W+K)^2\rho_t(\mathbb{T}\setminus L_\beta^+(t))$$

for every $t\geq 0$.

Proof. Set $\beta=\frac{\pi}{3}$ and use Theorem 3.1 to split the derivative of the dissipation functional into two parts as

$$\frac{d\mathcal{I}[f]}{dt}=I_1+I_2,$$

where the factors read

$$\begin{aligned}I_1=-K\int_{L_\beta^+(t)\times L_\beta^+(t)\times\mathbb{R}\times\mathbb{R}}&\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\times\cos(\theta-\theta')\,f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega',\\I_2=-K\int_{\substack{((\mathbb{T}\times\mathbb{T})\setminus(L_\beta^+(t)\times L_\beta^+(t)))\\\times\mathbb{R}\times\mathbb{R}}}&\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\times\cos(\theta-\theta')\,f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'.\end{aligned}$$

On the one hand, it is clear that

$$\begin{aligned}I_1\leq\ &-K\cos(\beta)\int_{L_\beta^+(t)\times L_\beta^+(t)\times\mathbb{R}\times\mathbb{R}}\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\qquad\times f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'\\=\ &-K\cos(\beta)\int_{\mathbb{T}^2\times\mathbb{R}^2}\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\qquad\times f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'\\&+K\cos(\beta)\int_{\substack{((\mathbb{T}\times\mathbb{T})\setminus(L_\beta^+(t)\times L_\beta^+(t)))\\\times\mathbb{R}\times\mathbb{R}}}\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\qquad\times f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'\\=:\ &I_{11}+I_{12},\end{aligned}\tag{3.22}$$

where in the second identity we have added and subtracted the second term in order to complete an integral in $\mathbb{T}^2\times\mathbb{R}^2$. Indeed, notice that doing so and using (1.16) we get

$$\begin{aligned}I_{11}&=-K\cos(\beta)\int_{\mathbb{T}^2\times\mathbb{R}^2}\big(\omega-KR\sin(\theta-\phi)-(\omega'-KR\sin(\theta'-\phi))\big)^2\\&\qquad\times f(t,\theta,\omega)\,f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'\\&=-2K\cos(\beta)\int_{\mathbb{T}\times\mathbb{R}}(\omega-KR\sin(\theta-\phi))^2 f\,d\theta\,d\omega=-2K\cos(\beta)\,\mathcal{I}[f].\end{aligned}$$

Here, we have used the cancellation of the crossed term after we expand the square appearing in the first factor. Let us call $I_3 = I_{12} + I_2$ and notice that

$$I_3 \le 2K \int_{\substack{((\mathbb{T}\times\mathbb{T})\setminus(L_\beta^+(t)\times L_\beta^+(t)))\\ \times\mathbb{R}\times\mathbb{R}}} \big(\omega - KR\sin(\theta-\phi) - (\omega' - KR\sin(\theta'-\phi))\big)^2$$
$$\times f(t,\theta,\omega) f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'.$$

In other words, we achieved the estimate

$$\frac{d\,\mathcal{I}[f]}{dt} \le -2K\cos(\beta)\,\mathcal{I}[f] + I_3. \tag{3.23}$$

Our last goal is to estimate the remainder I_3. Define the following time-dependent sets:

$$\begin{aligned}
A_1 &:= L_\beta^+(t) \times (\mathbb{T}\setminus L_\beta^+(t)) \times \mathbb{R}\times\mathbb{R},\\
A_2 &:= (\mathbb{T}\setminus L_\beta^+(t)) \times L_\beta^+(t) \times \mathbb{R}\times\mathbb{R},\\
A_3 &:= (\mathbb{T}\setminus L_\beta^+(t)) \times (\mathbb{T}\setminus L_\beta^+(t)) \times \mathbb{T}\times\mathbb{R}.
\end{aligned}$$

Since we have $((\mathbb{T}\times\mathbb{T})\setminus(L_\beta^+(t)\times L_\beta^+(t)))\times\mathbb{R}\times\mathbb{R} = A_1\cup A_2\cup A_3$, then we can split I_3 as

$$I_3 \le I_{31} + I_{32} + I_{33},$$

where the integrals take the form

$$I_{3i} := 2K \int_{A_i} \big(\omega - KR\sin(\theta-\phi) - (\omega' - KR\sin(\theta'-\phi))\big)$$
$$\times f(t,\theta,\omega) f(t,\theta',\omega')\,d\theta\,d\theta'\,d\omega\,d\omega'$$

for every $i = 1,2,3$. Changing variables we observe that $I_{31} = I_{32}$. Then we can focus on estimating I_{31} and I_{33} only. Notice that the integrand can be bounded as

$$\big((\omega - KR\sin(\theta-\phi)) - (\omega' - KR\sin(\theta'-\phi))\big)^2 \le 4(W+K)^2.$$

Then we obtain

$$I_{31}(t) \le 8K(W+K)^2\rho_t(\mathbb{T}\setminus L_\beta^+(t))$$

for every $t \ge 0$. Exactly the same argument allows I_{33} to be estimated and an identical bound to be obtained. Putting everything together into (3.23) finishes the proof. ■

Now we can apply the phase concentration estimate in Lemma 1.1 and Grönwall's lemma in order to derive the desired quantitative estimate on the decay rate of the dissipation.

Corollary 3.4. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W,W]$ and centered (i.e., (1.16)). Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Then there is a universal constant C such that if*

$$\frac{W}{K} \le CR_0^3,$$

there exists a time T_0 *with the property that*

$$T_0 \lesssim \frac{1}{KR_0^2} \log\left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right),$$

and

$$\mathcal{I}[f_t] \lesssim K^2 e^{-\frac{1}{20}K(t-T_0)}$$

for all t *in* $[T_0, \infty)$.

Proof. Let us adjust C small enough so that we meet the hypotheses of Lemma 1.1. Then there exists such a time T_0 so that

$$\rho_t(\mathbb{T} \setminus L_\alpha^+(t)) \leq M e^{-\frac{1}{20}K(t-T_0)}$$

for every $t \geq T_0$ and some universal constant M. This along with Corollary 3.3 implies

$$\frac{d}{dt}\mathcal{I}[f] \leq -2K\cos(\beta)\,\mathcal{I}[f] + 24K(W+K)^2 M e^{-\frac{1}{20}K(t-T_0)}$$

for any $t \geq T_0$. Integrating the inequality, we obtain

$$\begin{aligned}\mathcal{I}[f_t] &\leq \mathcal{I}[f_{T_0}]e^{-2K\cos(\beta)(t-T_0)} + \frac{24K(W+K)^2 M}{2K\cos(\beta) - \frac{1}{20}K}\left(e^{-\frac{K}{20}(t-T_0)} - e^{-2K\cos(\beta)(t-T_0)}\right)\\ &\lesssim (W+K)^2 e^{-\frac{K}{20}(t-T_0)} \lesssim K^2 e^{-\frac{K}{20}(t-T_0)},\end{aligned}$$

where in the second inequality we have used that

$$\mathcal{I}[f_{T_0}] \leq (W+K)^2,$$

by definition (2.1), and in the last inequality we have used the hypothesis on $\frac{W}{K}$. ∎

Using the transportation–dissipation inequality in Corollary 3.2 and the above exponential decay of the dissipation in Corollary 3.4 we obtain the following result.

Corollary 3.5. *Assume that the hypotheses in Corollary 3.4 hold true. Then*

$$W_{2,g}(f_t, f_s) \lesssim e^{-\frac{1}{40}K(t-T_0)} - e^{-\frac{1}{40}K(s-T_0)}$$

for every $s \geq t \geq T_0$.

We are now ready to conclude the proof of the main theorem of this paper.

Proof of Theorem 1.1.

Step 1: Convergence. By the above Corollary 3.5, the net $(f_t)_{t\geq 0}$ verifies the Cauchy condition in the metric space $(\mathbb{P}_g(\mathbb{T}\times\mathbb{R}), W_{2,g})$. Notice that it is a complete metric space. Consequently, there exists some probability measure $f_\infty \in \mathbb{P}_g(\mathbb{T}\times\mathbb{R})$ such that $W_{2,g}(f_t, f_\infty) \to 0$ as $t\to\infty$. Taking limits in the inequality in Corollary 3.5 as $s\to\infty$ yields

$$W_{2,g}(f_t, f_\infty) \lesssim e^{-\frac{1}{40}K(t-T_0)} \tag{3.24}$$

for every $t \geq T_0$, and using the order relation in Proposition 3.2 between the standard quadratic Wasserstein distance and the fibered quadratic Wasserstein distance concludes the exponential convergence in Theorem 1.1.

Step 2: *Uniqueness of the equilibrium.* Notice that, in particular, f_∞ is an equilibrium of the Kuramoto–Sakaguchi equation (1.2) and the asymptotic concentration estimate in Lemma 1.1 guarantees that

$$\operatorname{diam}(\operatorname{supp}_\theta f_\infty) \leq \beta = \frac{\pi}{3} < \frac{\pi}{2}.$$

Hence, by Proposition 3.3 it is unique up to phase shifts. ∎

Remark 3.4. As we advanced in Remark 1.2, a large coupling strength condition like (1.20) is necessary for the relaxation towards global equilibrium in Theorem 1.1 but also appears in the particle system (1.1). To briefly illustrate it, consider the simpler case of two oscillators with phases θ_1, θ_2 and natural frequencies ω_1, ω_2 with $\omega_1 \neq \omega_2$. Define $\theta := \theta_2 - \theta_1$ and $\omega := \omega_2 - \omega_1$ and use (1.1) to derive the following equation for $P(t) := \cos(\theta(t))$:

$$\frac{dP}{dt} = \sqrt{1-P^2}\big(K\sqrt{1-P^2} - \omega\operatorname{sgn}\bar{\theta}\big) \quad \text{for every } t \text{ in } [0,\infty).$$

Here, $\bar{\theta}$ stands for the representative of θ in $(-\pi,\pi]$ modulo 2π. Let us assume that the solution converges to a phase locked state. By [29, Theorem 3.1], it amounts to saying that there are finitely many collisions along the lifespan of the solution. Hence, after accounting for the last collision time $0 \leq T_0 < \infty$, we obtain the Riccati-type equation

$$\frac{dP}{dt} = K\sqrt{1-P^2}\Big(\sqrt{1-P^2} - \frac{W}{K}\Big) \quad \text{for every } t \text{ in } [T_0,\infty), \tag{3.25}$$

where $W = |\omega|$. Under the assumption $\frac{W}{K} < 1$ guaranteeing existence of equilibria of (1.1), equation (3.25) exhibits four different critical points. By direct inspection, one observes that a necessary condition for convergence towards a phase locked state is that $P(T_0)$ is above the second critical point. In particular, if $0 < r(T_0) < \frac{1}{\sqrt{2}}$ that condition amounts to

$$\frac{W^2}{K^2} < 4r(T_0)^2(1-r(T_0)^2).$$

This shows that a condition along the lines of (1.20) is necessary for relaxation towards the global equilibrium. However, we do not claim optimality in the exponent in (1.20) and it may be improved.

Remark 3.5. Similarly, as discussed in Remark 1.2, the time T_0 in Theorem 1.1 that is required for the solution to enter the concentration regime can become infinitely large for R_0 near zero. This also appears in the particle system (1.1). Again, we illustrate a simple example consisting of two oscillators with phases θ_1, θ_2 and frequencies ω_1, ω_2 at nearly antipodal initial configurations. Borrowing the above notation from Remark 3.4 we find

$$\frac{d\theta}{dt} = \omega - K \sin\theta \quad \text{for every } t \text{ in } [0, \infty).$$

For simplicity, we assume $\omega > 0$ and $\omega < K$ so that there are exactly two equilibria $\theta^+ := \arcsin(\omega/K)$ and $\theta^- := \pi - \arcsin(\omega/K)$. Finally, we set the initial datum $\theta_0 := \theta^- - \varepsilon$ for small $\varepsilon > 0$. Note that θ^+ is stable and θ^- is unstable, so that $\theta(t)$ must converge to the stable equilibrium as $t \to \infty$. However, we claim that the transient of time that is required for the solution to enter the concentration regime can be infinitely large if $\varepsilon \approx 0$.

Step 1*: Transient to the interval* $(\theta^+, \frac{\pi}{2})$. Notice that since $\theta_0 \in (\theta^+, \theta^-)$, then $\theta(t)$ is strictly decreasing. Hence, we have $\theta(t) \in (\theta^+, \theta_0]$ for all $t \geq 0$, and

$$\frac{d\theta}{dt} \leq \omega - K \sin(\theta_0),$$

whenever $\theta(t)$ stays in $[\frac{\pi}{2}, \theta_0)$ by monotonicity of the right-hand side. Therefore, a clear continuity argument shows that $\theta(t_\varepsilon) \in (\theta^+, \frac{\pi}{2})$ precisely at time $t_\varepsilon = (\theta_0 - \frac{\pi}{2})/(K \sin(\theta_0) - \omega)$. We remark that t_ε blows up as $\varepsilon \to 0$, so that the closer the initial configuration to the unstable equilibrium with antipodal mass, the longer the transient.

Step 2*: Exponential concentration.* Since $\theta(t)$ is strictly decreasing, then $\theta(t) \in (\theta^+, \theta(t_\varepsilon)]$ for all $t \geq t_\varepsilon$. Noting that $\theta(t_\varepsilon) < \frac{\pi}{2}$, we obtain

$$\frac{d}{dt}(\theta(t) - \theta^+) = -K\big(\sin(\theta(t)) - \sin(\theta^+)\big) \leq -K \cos(\theta(t_\varepsilon))(\theta(t) - \theta^+)$$

for any $t \geq t_\varepsilon$ (by the mean value theorem), and we conclude by Grönwall's lemma.

3.3. Semiconcavity, entropy production estimate, and lower bounds in the order parameter

In this part we quantify the entropy production principle that was anticipated in Section 2.1. As a by-product, in Corollary 3.6 we will obtain a universal lower bound on the order parameter, as discussed in Section 2.2.

Lemma 3.2 (Semiconcavity and entropy production). *Assume that f_0 is contained in $C^1(\mathbb{T} \times \mathbb{R})$ and that g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t, \theta, \omega)$ to* (1.2)*. Let $\alpha = \pi/6$, t_0 be a positive time, and λ be contained in* $(0, 1)$*. Additionally, suppose that*

$$\sqrt{2} R_0 \geq R(t_0) > \lambda R_0 \quad \textit{and} \quad \dot{R}(t_0) \geq \frac{K}{4} \lambda^3 R_0^3 \cos^2(\alpha).$$

Then there exists a universal constant C such that if

$$\frac{W}{K} \leq C\lambda^2 R_0^2, \tag{3.26}$$

then we can find a time $d \geq 0$ such that

$$\begin{aligned} R^2(t_0+d) - R^2(t_0) &\geq \frac{1}{40}\lambda^4 R_0^3, \\ R &\leq \frac{3}{2}R_0 \quad \text{in } [t_0, t_0+d]. \end{aligned} \tag{3.27}$$

Moreover, we can choose d so that

$$d \leq \frac{1}{3KR_0}\log 10.$$

Before we begin the proof of Lemma 3.2, we will need a relation between the time derivative of the order parameter and the dissipation (2.9). That is the content of the following lemma.

Lemma 3.3. *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and that g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ of (1.2). Then the inequality*

$$\mathcal{I}[f_t] - W^2 \leq K\frac{d}{dt}(R^2) \leq 3\,\mathcal{I}[f_t] + W^2 \tag{3.28}$$

holds for every t in $[0,\infty)$.

Proof. By (2.2) we have

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}KR^2 &= -\int KR\sin(\theta-\phi)(\omega - KR\sin(\theta-\phi))f\,d\theta\,d\omega \\ &= \mathcal{I}[f] - \int \omega(\omega - KR\sin(\theta-\phi))f\,d\theta\,d\omega. \end{aligned}$$

Consequently, by Young's inequality, we obtain

$$\frac{1}{2}\frac{d}{dt}KR^2 \leq \mathcal{I}[f] + \frac{1}{2}\int(\omega - KR\sin(\theta-\phi))^2 f\,d\theta\,d\omega + \frac{1}{2}\int \omega^2 f\,d\theta\,d\omega$$

and

$$\frac{1}{2}\frac{d}{dt}KR^2 \geq \mathcal{I}[f] - \frac{1}{2}\int(\omega - KR\sin(\theta-\phi))^2 f\,d\theta\,d\omega - \frac{1}{2}\int \omega^2 f\,d\theta\,d\omega.$$

Hence, the desired result follows. ∎

Now we are ready to prove our entropy production estimate.

Proof of Lemma 3.2. Without loss of generality, we can assume that

$$R < \frac{3}{2}R_0 \quad \text{in} \left[t_0, t_0 + \frac{1}{3KR_0}\log 10\right]. \tag{3.29}$$

Otherwise, if this condition fails for some s in the above interval, then we set $d = s - t_0$ and (3.27) would follow. Thanks to the inequalities (3.5) and (3.28), we arrive at the estimate

$$\begin{aligned}
\frac{dR^2}{dt} &\geq \frac{\mathcal{I}[f_t]}{K} - \frac{W^2}{K} \\
&\geq \frac{\mathcal{I}[f_{t_0}]e^{-3KR_0(t-t_0)}}{K} - \frac{W^2}{K} \\
&\geq \frac{1}{K}\left(\frac{K}{3}\frac{dR^2}{dt}\Big|_{t=t_0} - \frac{W^2}{3}\right)e^{-3KR_0(t-t_0)} - \frac{W^2}{K} \\
&= \frac{2}{3}R(t_0)\dot{R}(t_0)e^{-3KR_0(t-t_0)} - \frac{4W^2}{3K} \\
&\geq \frac{K}{6}\lambda^3 R_0^3 R(t_0)\cos^2(\alpha)e^{-3KR_0(t-t_0)} - \frac{4W^2}{3K}.
\end{aligned}$$

Let us integrate the above inequality on the interval $[t_0, t_0 + d]$ for some d in $[0, \frac{1}{3KR_0}\log 10)$, which we will choose appropriately after the calculations below. By doing this and using (3.29), we deduce that

$$R^2(t_0 + d) - R^2(t_0) \geq \frac{1}{18}\lambda^4 R_0^3 \cos^2(\alpha)[1 - e^{-3KR_0 d}] - \frac{4}{3}\frac{W^2}{K}d.$$

Thus, by choosing $d = \frac{1}{3KR_0}\log 10$, we obtain

$$R^2(t_0 + d) - R^2(t_0) \geq \frac{1}{20}\lambda^4 R_0^3 \cos^2(\alpha) - \frac{4}{9}\frac{W^2}{K^2 R_0}\log 10.$$

Consequently, by selecting C appropriately in (3.26) we conclude that

$$R^2(t_0 + d) - R^2(t_0) \geq \frac{1}{21}\lambda^4 R_0^3 \cos^2(\alpha).$$

Hence, since $\alpha = \pi/6$ the desired result follows. ■

Before showing the lower bound in the order parameter, we will need control in its angular velocity in the small dissipation regime. We achieve this in the following lemma.

Lemma 3.4. *Assume that f_0 is contained in $C^1(\mathbb{T} \times \mathbb{R})$ and that g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t, \theta, \omega)$ to* (1.2). *Then we have*

$$|\dot{\phi}| \leq \frac{1}{R}\sqrt{K\frac{d}{dt}R^2 + W^2}.$$

Proof. By (2.2), and Jensen's inequality, we have

$$\begin{aligned} R|\dot{\phi}| &\leq \int |\cos(\theta-\phi)(\omega - KR\sin(\theta-\phi))| f\,d\theta\,d\omega \\ &\leq \int |(\omega - KR\sin(\theta-\phi))| f\,d\theta\,d\omega \\ &\leq \left(\int |(\omega - KR\sin(\theta-\phi))|^2 f\,d\theta\,d\omega\right)^{\frac{1}{2}} \\ &= I^{\frac{1}{2}} \leq \sqrt{K\frac{d}{dt}R^2 + W^2}, \end{aligned}$$

where in the last inequality, we have used (3.28). Thus, the desired result follows. ■

We will derive a global lower bound on the order parameter as an application of the entropy production estimate (3.2). To achieve this, we consider the following lemma, which controls the rate at which the order parameter can decrease.

Lemma 3.5 (Rate of decrease and mass monotonicity). *Let λ be contained in $(2/3, 1)$, and assume that f_0 is contained $C^1(\mathbb{T}\times\mathbb{R})$ and that g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Additionally, let γ be a positive number in $(\pi/6, \pi/2)$, and let α be as specified in Section 2. Then we have*

$$\frac{d}{dt}R^2 \geq \frac{KR^2\cos^2(\gamma)}{2}\Big(1 - \frac{2W^2}{K^2R^2\cos^2(\gamma)} - \frac{R}{\sin(\gamma)} - \frac{1+\sin(\gamma)}{\sin(\gamma)} f(\chi_\alpha^-)\Big) \tag{3.30}$$

and

$$\frac{d}{dt} f(\chi_\alpha^-) \leq 4K\left[\frac{W}{K} + \sqrt{\frac{2\dot{R}}{KR} + \frac{1}{R^2}\frac{W^2}{K^2}} - R\cos(\alpha)\right]^+ \tag{3.31}$$

for all $t \geq 0$.

Moreover, suppose that $\dot{R}(t_0) \leq 0$, $R(t_0) \geq R_0$,

$$\dot{R} \leq \frac{K\lambda^3R_0^3\cos^2(\alpha)}{4} \quad \textit{in } [t_0, t_0+d] \quad \textit{and} \quad \cos^2(\gamma) = \frac{1-\lambda}{5}R_0$$

for non-negative numbers d and t_0. Then there exists a universal constant C such that if

$$\frac{W}{K} \leq C(1-\lambda)\lambda^2R_0^2, \tag{3.32}$$

then

$$\frac{d}{dt}R^2 > \frac{K\cos^2(\gamma)}{2\sin(\gamma)}\Big(-R^3 + \Big[\lambda R_0 + \frac{3}{5}(1-\lambda)R_0\Big]R^2 - \frac{3}{5}(1-\lambda)\lambda^2R_0^3\Big) \tag{3.33}$$

in $[t_0, t_0+d]$. Consequently,

$$R \geq \lambda R_0 \quad \textit{in } [t_0, t_0+d).$$

Proof. We divide the proof into the following steps:

Step 1*: Derivation of estimate* (3.31). Recall that $\chi_\alpha^-(\theta) = \xi_\alpha(\theta - \phi - \pi)$, with ξ_α as defined in (2.6). Then, by direct computation, we have

$$\begin{aligned}
\frac{d}{dt} f(\chi_\alpha^-) &= \frac{d}{dt}\int_{\mathbb{T}\times\mathbb{R}} \xi_\alpha(\theta-\phi-\pi) f \, d\theta \, d\omega \\
&= \int_{\mathbb{T}\times\mathbb{R}} \xi_\alpha'(\theta-\phi-\pi)[\omega - KR\sin(\theta-\phi) - \dot{\phi}] f \, d\theta \, d\omega \\
&\leq f(|\xi_\alpha'|)[W + |\dot{\phi}|] + KR\int_{\mathbb{T}\times\mathbb{R}} \xi_\alpha'(\theta-\phi-\pi)\sin(\theta-\phi-\pi) f \, d\theta \, d\omega \\
&\leq f(|\xi_\alpha'|)[W + |\dot{\phi}| - KR\cos(\alpha)] \\
&\leq f(|\xi_\alpha'|)\left[W + \frac{1}{R}\sqrt{2KR\frac{d}{dt}R + W^2} - KR\cos(\alpha)\right].
\end{aligned}$$

Notice that in the last inequality we have used Lemma 3.4 in order to estimate $|\dot{\phi}|$ and the only thing that remains to show is the bound on the second term in the third line. Firstly, the support of $\xi_\alpha'(\theta - \phi - \pi)$ consists of $S^+ \cup S^-$ where the sets stand for

$$\begin{aligned}
S^+ &:= \left[\phi + \frac{3\pi}{2} - \alpha, \phi + \frac{3\pi}{2} - \alpha + \frac{1}{2}\right], \\
S^- &:= \left[\phi + \frac{\pi}{2} + \alpha - \frac{1}{2}, \phi + \frac{\pi}{2} + \alpha\right].
\end{aligned}$$

Since $\xi_\alpha'(\theta - \phi - \pi)$ is non-increasing in S^+ and non-decreasing in S^-, we then obtain

$$\begin{aligned}
\theta \in S^+ &\Rightarrow \xi_\alpha'(\theta-\phi-\pi) \leq 0 \text{ and } \sin(\theta-\phi-\pi) \geq \cos(\alpha), \\
\theta \in S^- &\Rightarrow \xi_\alpha'(\theta-\phi-\pi) \geq 0 \text{ and } \sin(\theta-\phi-\pi) \leq -\cos(\alpha).
\end{aligned}$$

Consequently,

$$\xi_\alpha'(\theta-\phi-\pi)\sin(\theta-\phi-\pi) \leq -|\xi_\alpha'(\theta-\phi-\pi)|\cos(\alpha)$$

for all $\theta \in S^+ \cup S^-$, thus yielding the aforementioned bound. Hence, (3.31) follows.

Step 2*: Derivation of estimate* (3.30). By the first equation in (2.2), we obtain the following lower bound on $\dot{R}$:

$$\begin{aligned}
\frac{K}{2}\frac{d}{dt}R^2 &= -\int_{\mathbb{T}\times\mathbb{R}} KR\sin(\theta-\phi)(\omega - KR\sin(\theta-\phi)) f \, d\theta \, d\omega \\
&\geq \int_{\mathbb{T}\times\mathbb{R}} (KR\sin(\theta-\phi))^2 f \, d\theta \, d\omega - \int \omega(KR\sin(\theta-\phi)) f \, d\theta \, d\omega \\
&\geq \frac{1}{2}\int_{\mathbb{T}\times\mathbb{R}} (KR\sin(\theta-\phi))^2 f \, d\theta \, d\omega - \frac{W^2}{2} \\
&\geq \frac{1}{2}K^2R^2\cos^2(\gamma) f(\mathbb{T}\setminus(L_\gamma^+(t)\cup L_\gamma^-(t))) - \frac{W^2}{2}.
\end{aligned}$$

Then we obtain

$$f(\mathbb{T}\setminus(L_\gamma^+(t)\cup L_\gamma^-(t)))\leq \frac{1}{KR^2\cos^2(\gamma)}\frac{d}{dt}R^2+\frac{W^2}{K^2R^2\cos^2(\gamma)}. \tag{3.34}$$

Additionally, using a similar argument on (1.15), where we split the integral into the sectors L_γ^+, L_γ^-, and $\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-)$, allows the following lower bound to be obtained:

$$\begin{aligned}
R&\geq \sin(\gamma)f(L_\gamma^+)-\sin(\gamma)f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-f(L_\gamma^-)\\
&=\sin(\gamma)\big(1-f(L_\gamma^-)-f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))\big)-\sin(\gamma)f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-f(L_\gamma^-)\\
&=\sin(\gamma)-2\sin(\gamma)f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-(1+\sin(\gamma))f(L_\gamma^-)\\
&\geq \sin(\gamma)-2\sin(\gamma)\Big(\frac{1}{KR^2\cos^2(\gamma)}\frac{d}{dt}R^2+\frac{W^2}{K^2R^2\cos^2(\gamma)}\Big)-(1+\sin(\gamma))f(L_\gamma^-).
\end{aligned}$$

Here, we have used the estimate (3.34) in the last inequality. Then (3.30) follows.

Step 3: Upper bound on $f(L_\gamma^-)$. Let us first achieve a lower bound on $f(L_\gamma^+)$. To this end, we use a similar procedure and reverse the inequalities that we considered in the preceding step. Specifically, notice that a similar split in (1.15) allows us to obtain

$$\begin{aligned}
R&\leq f(L_\gamma^+)+\sin(\gamma)f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-\sin(\gamma)f(L_\gamma^-)\\
&=f(L_\gamma^+)+\sin\gamma f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-\sin(\gamma)(1-f(L_\gamma^+)-f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-)))\\
&=(1+\sin(\gamma))f(L_\gamma^+)+2\sin(\gamma)f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))-\sin(\gamma).
\end{aligned}$$

In particular, we obtain the lower bound

$$f(L_\gamma^+)\geq \frac{R}{1+\sin(\gamma)}-\frac{2\sin(\gamma)}{1+\sin(\gamma)}f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))+\frac{\sin(\gamma)}{1+\sin(\gamma)}.$$

Hence, we obtain the upper bound

$$\begin{aligned}
f(L_\gamma^-)&=1-f(L_\gamma^+)-f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))\\
&\leq 1-\frac{\sin(\gamma)}{1+\sin(\gamma)}-\frac{R}{1+\sin(\gamma)}-\frac{1-\sin(\gamma)}{1+\sin(\gamma)}f(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))\\
&\leq \frac{1}{1+\sin(\gamma)}-\frac{R}{1+\sin(\gamma)}.
\end{aligned} \tag{3.35}$$

Notice that since $\dot{R}(t_0)\leq 0$ we can select C appropriately in (3.32) to guarantee that

$$\begin{aligned}
&\frac{W}{K}+\sqrt{\frac{2\dot{R}(t_0)}{KR(t_0)}+\frac{1}{R(t_0)^2}\frac{W^2}{K^2}}-R(t_0)\cos(\alpha)\\
&\qquad\leq \frac{W}{K}+\sqrt{\frac{1}{R_0^2}\frac{W^2}{K^2}}-\lambda R_0\cos(\alpha)<0.
\end{aligned} \tag{3.36}$$

Then estimate (3.31) implies

$$\frac{d}{dt}\Big|_{t=t_0} f(\chi_\alpha^-)(t) \le 0.$$

By continuity and inequalities (3.31) and (3.36), $f(\chi_\alpha^-)(t)$ remains non-increasing along $[t_0, t_0+\delta]$ for small enough $\delta > 0$. Hence, we obtain

$$\begin{aligned} f(L_\gamma^-)(t) &\le f(\chi_\alpha^-)(t) \\ &\le f(\chi_\alpha^-)(t_0) \\ &\le f(L_\gamma^-)(t_0) + f(\mathbb{T} \setminus (L_\gamma^+ \cup L_\gamma^-))(t_0) \\ &\le \frac{1}{1+\sin(\gamma)} - \frac{R(t_0)}{1+\sin(\gamma)} + \frac{W^2}{K^2 R^2(t_0)\cos^2(\gamma)} \end{aligned} \tag{3.37}$$

for all t in $[t_0, t_0+\delta]$. Here, we have used estimates (3.34) and (3.35) along with the hypothesis $\dot{R}(t_0) \le 0$.

Step 4*: Derivation of* (3.33) *and lower bound on* R *in* $[t_0, t_0+\delta]$. Putting the last estimate (3.37) and (3.30) together, we obtain the differential inequality

$$\begin{aligned} \frac{dR^2}{dt} &\ge \frac{KR^2\cos^2(\gamma)}{2\sin(\gamma)}\Big[R(t_0) - R - (1-\sin(\gamma)) - \frac{2\sin(\gamma)W^2}{K^2\cos^2(\gamma)R^2} - \frac{1+\sin(\gamma)W^2}{K^2\cos^2(\gamma)R^2(t_0)}\Big] \\ &> \frac{K\cos^2(\gamma)}{2\sin(\gamma)}[-R^3 + b(t_0)R^2 - c(t_0)] \end{aligned} \tag{3.38}$$

for all t in $[t_0, t_0+\delta]$. Here, the coefficients read

$$\begin{aligned} b(t_0) &:= R(t_0) - \cos^2(\gamma) - \frac{2W^2}{K^2\cos^2(\gamma)R^2(t_0)}, \\ c(t_0) &:= \frac{2W^2}{K^2\cos^2(\gamma)}. \end{aligned}$$

Notice that in the last inequality in (3.38) we have used

$$1 - \sin(\gamma) < \cos^2(\gamma), \quad \sin(\gamma) < 1, \quad \text{and} \quad 1 + \sin(\gamma) < 2.$$

By making C smaller if necessary in (3.32) we can guarantee that

$$\begin{aligned} b(t_0) &= R(t_0) - \cos^2(\gamma) - \frac{2W^2}{K^2R^2(t_0)\cos^2(\gamma)} \\ &\ge R_0 - \frac{(1-\lambda)}{5}R_0 - 10\frac{W^2}{K^2}\frac{1}{R_0^3(1-\lambda)} \\ &\ge R(t_0) - \frac{2(1-\lambda)}{5}R_0 \\ &= \lambda R_0 + (1-\lambda)R_0 - \frac{2(1-\lambda)}{5}R_0 \\ &= \lambda R_0 + \frac{3}{5}(1-\lambda)R_0. \end{aligned}$$

Arguing in a similar way and making C smaller if necessary in (3.32), we can guarantee that

$$\begin{aligned} c(t_0) &:= \frac{2W^2}{K^2\cos^2(\gamma)} \\ &= \left(\frac{W}{K}\right)^2 \frac{10}{(1-\lambda)R_0} \\ &\leq \frac{3}{5}(1-\lambda)\lambda^2 R_0^3. \end{aligned}$$

Consequently, we have

$$\frac{d}{dt}R^2 > \frac{K\cos^2(\gamma)}{2\sin(\gamma)}\Big[-R^3 + \Big[\lambda R_0 + \frac{3}{5}(1-\lambda)R_0\Big]R^2 - \frac{3}{5}(1-\lambda)\lambda^2 R_0^3\Big]$$

in $[t_0, t_0+\delta]$. Since λR_0 is the biggest root of the polynomial

$$p(r) = -r^3 + \Big[\lambda R_0 + \frac{3}{5}(1-\lambda)R_0\Big]r^2 - \frac{3}{5}(1-\lambda)\lambda^2 R_0^3,$$

we obtain the desired lower bound $R \geq \lambda R_0$ in $[t_0, t_0+\delta]$ by an elementary continuity argument (we can see that λR_0 is the biggest root of p from the inequality $p(0) < 0$ and the fact that λ is contained in $(2/3, 1)$ implies that $p'(\lambda R_0) < 0$)).

Step 5: Propagation of (3.33) *and the lower bound on R in* $[t_0, t_0+d]$. The main idea is supported by a continuity method. We proceed by contradiction. Specifically, define the time

$$t_* := \inf\Big\{t \in (t_0+\delta, t_0+d] : \frac{d}{dt}R^2 < \frac{K\cos^2(\gamma)}{2\sin(\gamma)}p(R)\Big\},$$

and assume that $t^* < t_0 + d$. Notice that, by definition, it implies

$$\frac{d}{dt}R^2 \geq \frac{K\cos^2(\gamma)}{2\sin(\gamma)}p(R) \quad \text{for all } t \in [t_0, t_*].$$

In particular, by the same ideas as in Step 4, we have

$$R(t) \geq \lambda R_0 \quad \text{for all } t \in [t_0, t_*].$$

By (3.31) and the fact that

$$\dot{R} \leq \frac{K\lambda^3 R_0^3\cos^2(\alpha)}{4} \quad \text{in } [t_0, t_0+d],$$

making C smaller in (3.32) if necessary, we can guarantee that

$$\begin{aligned} &\frac{W}{K} + \sqrt{\frac{2\dot{R}(t)}{KR(t)} + \frac{1}{R(t)^2}\frac{W^2}{K^2}} - R(t)\cos(\alpha) \\ &\quad \leq \frac{W}{K} + \sqrt{\frac{\lambda^2 R_0^2\cos^2(\alpha)}{2} + \frac{1}{\lambda^2 R_0^2}\frac{W^2}{K^2}} - \lambda R_0\cos(\alpha) < 0 \end{aligned} \tag{3.39}$$

for all t in $[t_0, t_*]$. In particular, by (3.31) and continuity we have $f(\chi_\alpha^-)$ is non-increasing in $[t_0, t_* + \delta_*]$ and some small enough $\delta_* > 0$. Hence, we can repeat the train of thought in Step 4 to extend the upper bound on $f(\chi_\gamma^-)(t)$ in (3.37) to the larger interval $[t_0, t_* + \delta_*]$. Again, the same ideas as in Step 4 imply that

$$\frac{d}{dt} R^2 > \frac{K \cos^2(\gamma)}{2 \sin(\gamma)} p(R) \quad \text{for all } t \in [t_0, t_* + \delta_*],$$

and it contradicts the definition of t_*. ∎

We close this section by showing that we can obtain a universal lower bound on the order parameter. That is the objective of the following corollary.

Corollary 3.6 (Universal lower bound on R). *Suppose that* $1 - \lambda$ *is contained in* $(0, R_0/120)$. *Then there exists a universal constant* C *such that if*

$$\frac{W}{K} < C\lambda^2(1-\lambda)R_0^2, \tag{3.40}$$

then we have

$$R \geq \lambda R_0$$

for every t *in* $[0, \infty)$.

Proof. We begin by choosing C small enough so that it can be taken simultaneously as the corresponding universal constants in Lemmas 3.2 and 3.5. In addition, we claim that either of the following two conditions holds:

(i) We have $\dot{R} < K\lambda^3 R_0^3 \cos^2(\alpha)/4$ in $[0, \infty)$.

(ii) There exist a time t^* and an increasing and strictly positive universal function h, satisfying $R \geq \lambda R_0$ in $[0, t^*]$ and $R(t^*)^2 \geq R_0^2 + h(R_0)$.

We divide the proof of the corollary into two steps, the second of which is the proof of the claim.

Step 1*: We show how the claim implies the corollary.* To see this, we use the following iterative argument based on the fact that R is bounded and the system is autonomous. If condition (ii) of the claim holds, we use the fact that the system is autonomous in time to translate the initial condition of the system to be the configuration at t^*. Since by assumption the value of the order parameter at t^* is bigger than R_0 we are free to apply the claim again with the same value of C to the corresponding shifted initial condition. We can do this iteratively as many time as needed, provided that condition (ii) still holds after the time translation.

To conclude this step, note that since R is bounded and the function h is positive and increasing, then condition (ii) can hold consecutively after each time translation only a finite number of times. Hence, after finitely many time shifts, condition (i) will hold. Finally, once condition (i) holds, the global lower bound follows by applying Lemma 3.5.

Step 2: We show the claim. For this purpose suppose that (i) does not hold, that is, the set

$$\{t \geq 0 : \dot{R}(t) \geq K \tfrac{\lambda^3 R_0^3 \cos^2(\alpha)}{4}\}$$

is not empty. To show that (ii) holds in this case, let us consider the smallest time t_1 such that $\dot{R}(t_1) \geq K\lambda^3 R_0^3 \cos^2(\alpha)/4$. Now let t_2 denote the biggest time bigger than or equal to t_1, such that $\dot{R} \geq K\lambda^3 R_0^3 \cos^2(\alpha)/4$ in $[t_1, t_2]$. Notice that the finiteness of t_2 follows from the boundedness of R. Now, observe that, by definition of t_1, Lemma 3.5 implies that $R \geq \lambda R_0$ in $[0, t_1]$. Moreover, by construction,

$$\dot{R} \geq K \frac{\lambda^3 R_0^3 \cos^2(\alpha)}{4} \quad \text{in } [t_1, t_2].$$

Consequently, $R \geq R(t_1) \geq \lambda R_0$ in $[t_1, t_2]$. Now we consider two cases:

Case 1: $R(t_2) \leq \sqrt{2} R_0$. In this case, observe that Lemma 3.2 implies that we can find a constant d such that

$$R^2(t_2 + d) - R^2(t_2) = \frac{\lambda^4}{40} R_0^3.$$

Consequently, by our assumptions on λ, we have

$$\begin{aligned}
R(t_2 + d)^2 &= R(t_2)^2 + \frac{\lambda^4}{40} R_0^3 \\
&\geq \lambda^2 R_0^2 + \frac{\lambda^4}{40} R_0^3 \\
&\geq R_0^2 + \frac{\lambda^4}{40} R_0^3 - (1 - \lambda^2) R_0^2 \\
&> R_0^2 + \Big(\frac{5}{240} R_0 - 2(1 - \lambda)\Big) R_0^2 \\
&> R_0^2 + \frac{1}{240} R_0^3.
\end{aligned}$$

Here, on the third line, we have used the fact that $\lambda^4 > 9/10$.

Thus, the desired result follows by setting $t^* = t_2 + d$ and

$$h(r) := \frac{r^3}{240}.$$

Case 2: $R(t_2) > \sqrt{2} R_0$. In this case, we obtain

$$R(t_2)^2 - R_0^2 > R_0^2 > \frac{R_0^3}{240}.$$

Hence, the desired result holds for $t^* = t_2$. ∎

4. Instability of antipodal equilibria and sliding norms

We now start implementing the program outlined in Sections 2.3 and 2.4. To do this, we first derive inequalities (2.7) and (2.10).

Proposition 4.1 (Instability of antipodal equilibria). *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported in $[-W, W]$. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2) and let α be as specified in Section 2. Then we have*

$$\frac{d}{dt} f^2(\chi_\alpha^-(t)) \le -KR\sin(\alpha) f^2(\chi_\alpha^-(t)) + 4Kf^2(\mathbb{T})\left[\frac{W}{K} + \sqrt{\frac{2\dot{R}}{KR} + \frac{1}{R^2}\frac{W^2}{K^2}} - R\cos(\alpha)\right]^+$$

and

$$\frac{d}{dt} f^2(\mathbb{T}) \le KRf^2(\mathbb{T}). \tag{4.1}$$

Moreover, with hypothesis (3.32) and notation from Lemma 3.5, if $[t_1, t_2]$ is a time interval such that

$$\dot{R} \le K\frac{\lambda^3 R_0^3\cos^2(\alpha)}{4} \quad \text{in } [t_1, t_2], \tag{4.2}$$

then we have

$$\frac{d}{dt} f^2(L_\alpha^-(t)) \le -K\lambda R_0\sin(\alpha) f^2(L_\alpha^-(t)) \quad \text{in } [t_1, t_2]. \tag{4.3}$$

Proof. We begin with the first inequality in the proposition. Arguing as in Step 1 of the proof of Lemma 3.5 we obtain

$$\begin{aligned}
&\frac{d}{dt}\int \chi_\alpha^-(\theta-\phi+\pi) f^2\, d\theta\, d\omega \\
&\quad= \int \dot{\phi}\chi_\alpha^{-\prime}(\theta-\phi+\pi) f^2\, d\theta\, d\omega + 2\int \chi_\alpha^-(\theta-\phi+\pi) f\partial_t f\, d\theta\, d\omega \\
&\quad= \int [\dot{\phi} + 2\omega - 2KR\sin(\theta-\phi)]\chi_\alpha^{-\prime}(\theta-\phi+\pi) f^2\, d\theta\, d\omega \\
&\qquad+ 2\int \chi_\alpha^-(\theta-\phi+\pi)[\omega - KR\sin(\theta-\phi)] f\partial_\theta f\, d\theta\, d\omega \\
&\quad\le \int [\dot{\phi} + \omega - KR\sin(\theta-\phi)]\chi_\alpha^{-\prime}(\theta-\phi+\pi) f^2\, d\theta\, d\omega \\
&\qquad- \int \chi_\alpha^-(\theta-\phi+\pi) KR\sin(\alpha) f^2\, d\theta\, d\omega.
\end{aligned} \tag{4.4}$$

The first inequality in the proposition follows from Lemma 3.4 and the same arguments as in Step 1 from Lemma 3.5. Inequality (4.1) follows from similar arguments to those of

(4.4) by replacing χ_α with the constant function that is equal to 1 in $\mathbb{T}$. Finally, to derive inequality (4.3), recalling the notation introduced in Section 2.3, replacing χ_α^- with $\chi_{\alpha,\varepsilon}^-$ in (4.4), and arguing as in Step 1 from Lemma 3.5 we get

$$\frac{d}{dt} f^2(\chi_{\alpha,\varepsilon}^-(t)) \leq -KR\sin(\alpha) f^2(\chi_{\alpha,\varepsilon}^-(t)) + KC_{\varepsilon,\alpha} f^2(\mathbb{T})\left[\frac{W}{K} + \sqrt{\frac{2\dot{R}}{KR} + \frac{1}{R^2}\frac{W^2}{K^2}} - R\cos(\alpha)\right]^+.$$

Now we observe that, as in (3.39), the second term of the above inequality vanishes on the interval $[t_1, t_2]$. Consequently, such a term is independent of ε and thus (4.3) follows by letting $\varepsilon \to 0$. ∎

A form of the above lemma was one of the main tools used to derive the main result in [31]. However, to obtain our convergence rates, we work with a sliding version of the L^2 norm. Such sliding norms allow us to propagate the above estimate along the flow of the continuity equation. This technique turns out to be one of the crucial components in our arguments in Section 5.

Lemma 4.1 (Sliding norms). *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g is compactly supported. Consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Then, for any measurable set A, we have*

$$\frac{d}{dt} f^2(A_{t_0,t}) \leq KR\Big(\sup_{(\theta,\omega)\in A_{t_0,t}} \cos(\theta-\phi(t))\Big) f^2(A_{t_0,t}).$$

Proof. By the change of variable theorem, we have

$$\begin{aligned}
\frac{d}{dt}\frac{1}{2}\int_{A_{t_0,t}} f^2\,d\theta\,d\omega &= \frac{d}{dt}\Big|_{t=t_0}\frac{1}{2}\int_A f_t^2(\Theta_{t_0,t}(\theta,\omega),\omega)\partial_\theta\Theta_{t_0,t}\,d\theta\,d\omega\\
&= \int_A f_t(\Theta_{t_0,t}(\theta,\omega),\omega)\\
&\qquad \times\big[\partial_t f(\Theta_{t_0,t}(\theta,\omega),\omega) + \dot{\Theta}_{t_0,t}(\theta,\omega)\partial_\theta f(\Theta_{t_0,t_0}(\theta,\omega),\omega)\big]\partial_\theta\Theta_{t_0,t}\,d\theta\,d\omega\\
&\quad -\frac{1}{2}KR\int_A \cos(\Theta_t(\theta,\omega)-\phi)\partial_\theta\Theta_{t_0,t} f^2\,d\theta\,d\omega\\
&= \int_A f_t(\Theta_{t_0,t}(\theta,\omega),\omega)\big[-\partial_\theta(\omega f - KR\sin(\Theta_{t_0,t}(\theta,\omega)-\phi)f)\\
&\qquad + (\omega - KR\sin(\Theta_{t_0,t}(\theta,\omega)-\phi))\partial_\theta f(\Theta_{t_0,t}(\theta,\omega),\omega)\big]\partial_\theta\Theta_{t_0,t}\,d\theta\,d\omega\\
&\quad +\frac{1}{2}KR\int_A \cos(\Theta_{t_0,t}(\theta,\omega)-\phi) f^2\partial_\theta\Theta_{t_0,t}\,d\theta\,d\omega\\
&= \frac{1}{2}KR\int_A \cos(\Theta_{t_0,t}(\theta,\omega)-\phi) f_t^2(\theta,\omega)\partial_\theta\Theta_{t_0,t}\,d\theta\,d\omega,
\end{aligned}$$

where for t and each ω, $\partial_\theta\Theta_{t_0,t}(\cdot,\omega)$ denotes the Jacobian of the map $\theta \to \Theta_{t_0,t}(\theta,\omega)$. Hence, the desired result follows. ∎

To make full use of the above control, we need to understand the dynamics of the Lagrangian flow associated with the continuity equation. That is the objective of the next part.

4.1. Emergence of attractors

In this section we will show the emergence of time-dependent sets that will act as attractors along the characteristic flow. Such sets, in combination with our analysis on sliding norms in the previous section, will allow us to propagate information between the different parts of the system.

Before showing the emergence of attractor sets, we state the following lemma, which we will repeatedly use throughout the rest of the paper. Additionally, in this part we will use the notation introduced in Section 2.4.

Lemma 4.2 (Emergence of invariant sets). *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$, g has compact support in $[-W, W]$, and consider the unique global-in-time classical solution $f = f(t,\theta,\omega)$ to (1.2). Let $t_0 \geq 0$ be an initial time in $[0,\infty)$ and $L \subset \mathbb{T}$ be an interval. Now assume that initially we have*

$$\rho_{t_0}(L) \geq m \quad \textit{and} \quad p = \inf_{\theta,\theta'\in L} \cos(\theta-\theta')$$

for some positive numbers m and p in $(0,1)$. Additionally, suppose that

$$mp-(1-m)\geq \sigma \quad \textit{and} \quad \frac{W^2}{K^2} \leq \frac{(1-p)\sigma^2}{4} \tag{4.5}$$

for some $\sigma > 0$. Then, if we set

$$\underline{P}(t) = \inf_{\theta,\theta'\in L_{t_0,t}} \cos(\theta-\theta'),$$

the following bounds hold true:

$$\rho(L_{t_0,t}) \geq m, \tag{4.6}$$

$$\inf_{\theta\in L_{t_0,t}} R\cos(\theta-\phi) \geq m\underline{P}-(1-m), \tag{4.7}$$

and

$$1-\underline{P}(t) \leq \max\left((1-p)e^{-\frac{K\sigma}{4}(t-t_0)}, \frac{4}{\sigma^2}\frac{W^2}{K^2}\right) \tag{4.8}$$

for every t in $[t_0,\infty)$

Proof. The proof of (4.8) is based on a continuity method argument that holds under condition (4.5). The argument is based on inequalities (4.6), (4.7), and

$$\frac{dP}{dt} \geq 2K\sqrt{1-P^2}\Big[R\Big(\inf_{\theta\in L_{t_0,t}} \cos(\theta-\phi)\Big)\sqrt{\frac{1-P}{2}}-\frac{W}{K}\Big] \quad \text{for all } t\geq t_0, \tag{4.9}$$

which hold when

$$P=\cos(\Theta_{s,t}(\theta,\omega)-\Theta_{s,t}(\theta',\omega')) \tag{4.10}$$

for any $s\geq t_0$ such that $t\geq s$, and any couple of points (θ,ω) and (θ',ω') contained in $L_{t_0,s}\times[-W,W]$.

We will first prove inequality (4.8) and will prove the remaining inequalities afterwards. Indeed, let us define t' as the supremum of the set of times $t^*\geq t_0$ such that inequality (4.8) holds, for every t in $[t_0,t^*]$. We begin by noting that, by continuity,

$$1-\underline{P}(t')=\max\Big((1-p)e^{-\frac{K\sigma}{4}(t'-t_0)},\frac{4}{\sigma^2}\frac{W^2}{K^2}\Big).$$

Now we must prove that there exists $\delta>0$ such that (4.8) holds in $[t_0,t'+\delta]$. More precisely, our goal is to show that there exists a uniform time $\delta>0$ such that for any pair of characteristics starting at $L_{t_0,t'}\times[-W,W]$ we have that the corresponding P (given by (4.10)) satisfies that $1-P$ is bounded by the right-hand side of (4.8) in $[t',t'+\delta]$.

To do this, let $s=t'$ in the definition of P. Now observe that by (4.7) and (4.9), when $t=t'$, we have

$$\begin{aligned}\frac{dP}{dt}\Big|_{t=t'}&\geq 2K\sqrt{1-P^2}\Big[R\Big(\inf_{\theta\in L_{t_0,t'}}\cos(\theta-\phi)\Big)\sqrt{\frac{1-P}{2}}-\frac{W}{K}\Big]\\&\geq 2K\sqrt{1-P^2}\Big[[m\underline{P}-(1-m)]\sqrt{\frac{1-P}{2}}-\frac{W}{K}\Big]\\&\geq 2K\sqrt{1+P}\Big[\frac{\sqrt{2}}{2}\sigma(1-P)-\frac{W}{K}\sqrt{1-P}\Big].\end{aligned} \tag{4.11}$$

Here, all the time-dependent expressions are evaluated at $t=t'$. Additionally, in the last inequality, we have used our assumption that (4.8) holds on the interval $[t_0,t']$, which together with (4.5) implies the uniform lower bound $p\leq\underline{P}$. Now let (θ,ω) and (θ',ω') be any couple of points contained in $L_{t_0,t'}\times[-W,W]$ such that the corresponding P satisfies

$$1-P(t')=1-\underline{P}(t')=\max\Big((1-p)e^{-\frac{K\sigma}{4}(t'-t_0)},\frac{4}{\sigma^2}\frac{W^2}{K^2}\Big). \tag{4.12}$$

Note that since L is compact, then $L_{t_0,t'}\times[-W,W]$ is compact as well. Thus, the set of such pairs (θ,ω) and (θ',ω') in $L_{t_0,t'}$ whose corresponding P (obtained via (4.10)) satisfies (4.12) is a compact set as well. We will denote such a set by $\mathscr{P}\subset L_{t_0,t'}\times[-W,W]\times L_{t_0,t'}\times[-W,W]$. To continue our proof observe that by using assumption (4.12), we get

$$\frac{\sigma\sqrt{1-P(t')}}{2}\geq\frac{W}{K}$$

for any couple of characteristics in $\mathscr{P}$ and, consequently, by (4.11) we obtain

$$\begin{aligned}\frac{d}{dt}\Big|_{t=t'}(1-P)&\leq -2K\sqrt{1+P}\Big[\frac{\sqrt{2}}{2}\sigma(1-P)-\frac{\sigma(1-P(t'))}{2}\Big]\\&\leq -\frac{2}{5}K\sigma(1-\underline{P}(t'))\end{aligned}$$

$$
< \begin{cases} -\frac{2}{5}K\sigma(1-p)e^{-\frac{K\sigma}{4}(t'-t_0)} & \text{if } \frac{4}{\sigma^2}\frac{W^2}{K^2} < 1-\underline{P}(t'), \\ 0 & \text{if } \frac{4}{\sigma^2}\frac{W^2}{K^2} = 1-\underline{P}(t'). \end{cases}
$$

Since the right-hand side of the above inequality is uniform in the set of pairs in $\mathscr{P}$ and the set $\mathscr{P}$ is compact, we can find $\varepsilon > 0$ such that if $\mathscr{P}_\varepsilon$ is an ε-neighborhood of $\mathscr{P}$, then we have

$$
\begin{aligned}
&\frac{d}{dt}\Big|_{t=t'}(1-\cos(\Theta_{t',t}(\theta,\omega)-\Theta_{t',t}(\theta',\omega'))) \\
&\quad < -\frac{1}{3}K\sigma((1-\underline{P}(t')) \\
&\quad < \begin{cases} -\frac{K}{4}\sigma(1-p)e^{-\frac{K\sigma}{4}(t'-t_0)} & \text{if } \frac{4}{\sigma^2}\frac{W^2}{K^2} < 1-\underline{P}(t'), \\ 0 & \text{if } \frac{4}{\sigma^2}\frac{W^2}{K^2} = 1-\underline{P}(t') \end{cases}
\end{aligned} \tag{4.13}
$$

for any (θ,ω), (θ',ω') in $\mathscr{P}_\varepsilon$. This implies the existence of δ and thus concludes the continuity method argument. Indeed, for characteristics with initial data in $\mathscr{P}_\varepsilon$ the existence of the time interval $[t', t'+\delta)$ follows by the fact that the inequality in (4.13) is strict and uniform in $\mathscr{P}_\varepsilon$. Similarly, for characteristics in $(L_{t_0,t'} \times [-W,W] \times L_{t_0,t'} \times [-W,W]) \backslash \mathscr{P}_\varepsilon$, the existence of the uniform time δ follows by the fact that the characteristics have uniformly bounded speed and ε provides a uniform separation distance.

(Indeed, by continuity and compactness, we can find a uniform time neighborhood of t', in which the infimum for $\underline{P}$ is attained in $\mathscr{P}_{\varepsilon/2}$, and we have already shown the existence of δ in such a case.)

Hence, to complete the proof of the lemma it suffices to derive inequalities (4.6), (4.7), and (4.9). We achieve this in the following steps:

Step 1*: Proofs of inequalities* (4.6) *and* (4.7). Inequality (4.6) follows from the fact that the continuity equation preserves the mass of sets along the characteristic flow. To derive inequality (4.7) we observe that

$$
\begin{aligned}
\inf_{\theta\in L_{t_0,t}} R\cos(\theta-\phi) &= \inf_{\theta\in L_{t_0,t}} \left\langle e^{i\theta}, \int e^{i\theta'} f'\,d\omega'\,d\theta' \right\rangle \\
&\geq \inf_{\theta\in L_{t_0,t}} \int \cos(\theta-\theta')f'\,d\theta'\,d\omega' \\
&\geq \inf_{\theta\in L_{t_0,t}} \bigg[\int_{(L\times[-W,W])_{t_0,t}} \cos(\theta-\theta')f'\,d\theta'\,d\omega' \\
&\qquad\qquad + \int_{\mathbb{T}\times\mathbb{R}\backslash(L\times[-W,W])_{t_0,t}} \cos(\theta-\theta')f'\,d\theta'\,d\omega'\bigg] \\
&\geq m\underline{P} - (1-m).
\end{aligned} \tag{4.14}
$$

This completes Step 1.

Step 2: Proof of inequality (4.9). To obtain (4.9) let us fix t in $[t_0,\infty)$, and let (θ,ω) and (θ',ω') be contained in $L\times[-W,W]$. Additionally, let us set

$$\Theta(s):=\Theta_{t_0,s}(\theta,\omega)\quad\text{and}\quad\Theta'(s):=\Theta_{t_0,s}(\theta',\omega').$$

Then

$$\begin{aligned}
\frac{d}{ds}\Big|_{s=t}&\cos(\Theta-\Theta')=-\sin(\Theta-\Theta')(\dot\Theta-\dot\Theta')\\
&=-\sin(\Theta-\Theta')\big((\omega-\omega')-KR(\sin(\Theta-\phi)-\sin(\Theta'-\phi))\big)\\
&=-\sin(\Theta-\Theta')\left[(\omega-\omega')-2KR\cos\left(\frac{\Theta+\Theta'}{2}-\phi\right)\sin\left(\frac{\Theta-\Theta'}{2}\right)\right]\\
&=-2\cos\left(\frac{\Theta-\Theta'}{2}\right)\left[(\omega-\omega')\sin\left(\frac{\Theta-\Theta'}{2}\right)\right.\\
&\qquad\qquad\left.-2KR\cos\left(\frac{\Theta+\Theta'}{2}-\phi\right)\sin^2\left(\frac{\Theta-\Theta'}{2}\right)\right]\\
&\geq 4KR\cos\left(\frac{\Theta-\Theta'}{2}\right)\left[\cos\left(\frac{\Theta+\Theta'}{2}-\phi\right)\frac{1-\cos(\Theta-\Theta')}{2}\right.\\
&\qquad\qquad\left.-\frac{W}{KR}\sqrt{\frac{1-\cos(\Theta-\Theta')}{2}}\right],
\end{aligned}\tag{4.15}$$

where we have used several standard trigonometric formulas. Now, notice that

$$\frac{\Theta_{t_0,t}(\theta,\omega)+\Theta_{t_0,t}(\theta',\omega')}{2}\quad\text{is contained in } L_{t_0,t},$$

since it is a convex combination of two points in $L_{t_0,t}$. Thus, when $s=t_0$, (4.9) follows by standard trigonometric identities. In the case when s is contained in $[t_0,t]$ we can easily derive (4.9) by the same argument and the semigroup property of the characteristic flow. ∎

As a first application of the above lemma, we quantify below the first time that the system forms an attractor.

Lemma 4.3 (First invariant set). *Assume that f_0 is contained in $C^1(\mathbb{T}\times\mathbb{R})$ and g has compact support in $[-W,W]$. Consider the unique global-in-time classical solution to* (1.2), *$f=f(t,\theta,\omega)$, and let us set an angle $0<\gamma<\frac{\pi}{2}$ so that*

$$\cos^2(\gamma)=\frac{1}{30}R_0.\tag{4.16}$$

Then we can find a universal constant C such that if

$$\frac{W}{K}\leq CR_0^2,\tag{4.17}$$

there exists then a positive time T_{-1} satisfying that

$$T_{-1}\lesssim\frac{1}{KR_0^3},\tag{4.18}$$

and the bounds

$$\rho(L_\gamma^+(T_{-1})_t) \geq \frac{1+\frac{4}{5}R_0}{2}, \tag{4.19}$$

$$\inf_{\theta \in L_\gamma^+(T_{-1})_t} R\cos(\theta-\phi) \geq \frac{3}{5}R_0, \tag{4.20}$$

and

$$\inf_{\theta,\theta' \in L_\gamma^+(T_{-1})_t} \cos(\theta-\theta') \geq 1-\frac{1}{15}R_0 \tag{4.21}$$

hold true for every t in $[T_{-1},\infty)$.

Proof. Define the time

$$T_{-1} := \inf\Big\{t \geq 0 : \dot{R} \leq \frac{KR_0^3}{4\cdot 30^2}\Big\}, \tag{4.22}$$

and note that by construction, (4.18) follows from the fact that R is bounded by 1 and the fundamental theorem of calculus.

The proof of the remaining parts of the lemma will follow directly from an application of Lemma 4.2 by setting $L = L_\gamma^+(T_{-1})$ and $t_0 = T_{-1}$. To verify the corresponding hypotheses, first, we begin by controlling the mass in $L_\gamma^+(T_{-1})$. Indeed, dividing the integral (1.15) in the definition of R into three parts L_γ^+, L_γ^-, and $\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-)$, we obtain the inequality

$$R \leq (1+\sin(\gamma))\rho(L_\gamma^+) - \sin(\gamma) + 2\sin(\gamma)\rho(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-)). \tag{4.23}$$

Consequently, using (3.34) to control $\rho(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-))$, we deduce that

$$\begin{aligned}
\rho(L_\gamma^+) &\geq \frac{R+\sin(\gamma)}{1+\sin(\gamma)} - \frac{2\sin(\gamma)}{1+\sin(\gamma)}\rho(\mathbb{T}\setminus(L_\gamma^+\cup L_\gamma^-)) \\
&\geq \frac{1}{1+\sin(\gamma)}\Big[R+\sin(\gamma) - 2\Big(\frac{1}{KR^2\cos^2(\gamma)}\frac{d}{dt}R^2 + \frac{W^2}{K^2R^2\cos^2(\gamma)}\Big)\Big] \\
&= \frac{1}{1+\sin(\gamma)}\Big[R+1+(\sin(\gamma)-1) - 2\Big(\frac{2\dot{R}}{KR\cos^2(\gamma)} + \frac{W^2}{K^2R^2\cos^2(\gamma)}\Big)\Big]
\end{aligned}$$

for any $t \geq 0$. Then, evaluating the above expression at $t = T_{-1}$, using the fact that by construction $R(T_{-1}) \geq R_0$, and selecting $C < 1/30$ in (4.17), we deduce that

$$\begin{aligned}
\rho(L_\gamma^+(T_{-1})) &\geq \frac{1}{2}\Big[R_0+1+(\sin(\gamma)-1) - 2\Big(\frac{2\dot{R}(T_{-1})}{KR\cos^2(\gamma)} + \frac{W^2}{K^2R^2\cos^2(\gamma)}\Big)\Big] \\
&\geq \frac{1}{2}\Big[R_0+1-\frac{R_0}{30} - 2\Big(30\frac{2\dot{R}(T_{-1})}{KR_0^2} + \frac{30}{R_0}\frac{W^2}{K^2R_0^2}\Big)\Big] \\
&\geq \frac{1}{2}\Big(1+\frac{4}{5}R_0\Big),
\end{aligned} \tag{4.24}$$

where we have used the fact that $1-\sin(\gamma)\leq 1-\sin(\gamma^2)=\cos^2(\gamma)=\frac{R_0}{30}$. Second, we estimate the infimum of the cosine of the difference of angles in $L_\gamma^+(T_{-1})$, that is,

$$\begin{aligned}\inf_{\theta,\theta'\in L_\gamma^+(T_{-1})}\cos(\theta-\theta') &= \cos(\pi-2\gamma)=\cos\Big(2\Big(\frac{\pi}{2}-\gamma\Big)\Big)\\ &= 2\cos^2\Big(\frac{\pi}{2}-\gamma\Big)-1=2\sin^2(\gamma)-1=1-\frac{1}{15}R_0. \end{aligned}\tag{4.25}$$

Finally, considering Lemma 4.2 with $m=\rho(L_\gamma^+(T_{-1}))$ and $p=\cos(\pi-2\gamma)$, and using the bounds in (4.24) and (4.25), we obtain

$$\begin{aligned} mp-(1-m) &\geq \frac{1+\frac{4R_0}{5}}{2}\Big(1-\frac{1}{15}R_0\Big)+\Big(\frac{\frac{4R_0}{5}-1}{2}\Big)\\ &\geq \frac{1}{2}\Big(\frac{8}{5}R_0-\frac{1}{15}R_0-\frac{4}{75}R_0^2\Big)>\frac{3}{5}R_0. \end{aligned}$$

Thus, the desired result follows by applying Lemma 4.2 with $\sigma=\frac{3}{5}R_0$ and noticing that the hypothesis in (4.5) follows by the assumption (4.17) after we take C small enough. ∎

In the next corollary we will explain in which sense the sets whose formation we have shown above have an attractive property. Before stating that, we will need the following notation:

Definition 4.1. Given positive times $t_0\leq t_1$, we will define the new time-dependent interval in $[t_1,\infty)$, which will be a dynamic neighborhood of $L_\gamma^+(t_0)_{t_1}$, as follows. First, we define

$$(L_\gamma^+(t_0)_{t_1})_\epsilon := \left\{\theta\in\mathbb{T} : \inf_{\theta^*\in L_\gamma^+(t_0)_{t_1}}\cos(\theta-\theta^*)\geq 1-\epsilon\right\}$$

for any ϵ in $[R_0/15,1)$. Second, using the notation in Section 2.4, for any $t>t_1$ we will denote the θ-projection of the image of $(L_\gamma^+(t_0)_{t_1})_\epsilon$ under the characteristic flow, that is, $\Theta_{t_1,t}((L_\gamma^+(t_0)_{t_1})_\epsilon\times[-W,W])$, by $(L_\gamma^+(t_0)_{t_1})_{\epsilon,t}$. When t_0 is clear from the context, we will avoid referring to it in the above notation.

Now we are ready to state the corollary.

Corollary 4.1 (Emergence of attractor sets). *Consider non-negative times $t\geq t_1\geq T_{-1}$ and let $\epsilon=R_0/15$. Then there exists a universal constant C such that if*

$$\frac{W}{K}<CR_0^2, \tag{4.26}$$

then

$$\rho\big((L_\gamma^+(T_{-1})_{t_1})_{\epsilon,t}\big)\geq\frac{1+\frac{4}{5}R_0}{2}, \tag{4.27}$$

$$\inf_{\theta\in(L_\gamma^+(T_{-1})_{t_1})_{\epsilon,t}} R\cos(\theta-\phi)\geq\frac{1}{2}R_0, \tag{4.28}$$

and

$$\inf_{\theta,\theta'\in(L_\gamma^+(T_{-1})_{t_1})_{\epsilon,t}} \cos(\theta-\theta') \geq 1-\frac{1}{3}R_0 \tag{4.29}$$

hold true for every t in $[t_1,\infty)$.

Proof. We will show how to select C appropriately at the end of the proof. For the moment, let us make it small enough so that we can use Lemma 4.3. The proof will follow directly from Lemma 4.2 by setting $L := L_\gamma^+(T_{-1})_{t_1,\epsilon}$. To verify the corresponding hypotheses, first we begin by controlling the mass in L. Indeed, by Lemma 4.3 we have

$$\rho(L_\gamma^+(T_{-1})_{t_1,\epsilon}) \geq \rho(L_\gamma^+(T_{-1})_{t_1}) \geq \frac{1+\frac{4}{5}R_0}{2}. \tag{4.30}$$

Second, we estimate the infimum over the cosine of the difference of the angles in $L_\gamma^+(T_{-1})_{t_1,\epsilon}$. For this purpose let $\bar\theta$ be contained in $L_\gamma^+(T_{-1})_{t_1}$. Then, for any θ and θ' in $(L_\gamma^+(T_{-1})_{t_1})_\epsilon$, we have

$$\begin{aligned}
\cos(\theta-\theta') &= \cos(\theta-\bar\theta+\bar\theta-\theta')\\
&= \cos(\theta-\bar\theta)\cos(\theta'-\bar\theta)-\sin(\theta-\bar\theta)\sin(\theta'-\bar\theta)\\
&\geq \Big[1-\frac{1}{15}R_0\Big]^2-\Big[1-\Big[1-\frac{1}{15}R_0\Big]^2\Big]\\
&\geq 2\Big[1-\frac{1}{15}R_0\Big]^2-1 \geq 1-\frac{1}{3}R_0.
\end{aligned}$$

Thus, since θ and θ' were arbitrary, we deduce that

$$\inf_{\theta,\theta'\in L_\gamma^+(T_{-1})_{t_1,\epsilon}} \cos(\theta-\theta') \geq 1-\frac{1}{3}R_0. \tag{4.31}$$

Finally, considering $m=\frac{1+\frac{4}{5}R_0}{2}$ and $p=1-\frac{1}{3}R_0$ and using the bounds in (4.30) and (4.31), we obtain

$$\begin{aligned}
mp-(1-m) &\geq \frac{1+\frac{4R_0}{5}}{2}\Big(1-\frac{1}{3}R_0\Big)+\Big(\frac{\frac{4R_0}{5}-1}{2}\Big)\\
&\geq \frac{1}{2}\Big(\frac{8}{5}R_0-\frac{1}{3}R_0-\frac{4}{15}R_0^2\Big)\\
&> \frac{1}{2}\Big(\frac{24-5-4}{15}\Big)R_0=\frac{1}{2}R_0.
\end{aligned}$$

Therefore, the desired result follows by choosing C appropriately in (4.26) so that (4.5) holds and applying Lemma 4.2 with $\sigma=\frac{R_0}{2}$. ∎

4.2. Control of L^2 norms outside the attractors

In the next lemma, we derive an estimate that we will use in Section 5. The estimate shows that if the entropy production vanishes over sufficiently long intervals of time, then the L^2 norm of the solution in $\mathbb{T}\backslash(L_\gamma^+(T_{-1})_t)_\epsilon$ will begin to decrease exponentially.

Lemma 4.4. *Let $[t_1, t_2]$ be a time interval in $[T_{-1}, \infty)$, such that*

$$\dot{R} \leq K\frac{\lambda^3 R_0^3 \cos^2(\alpha)}{4} \quad \textit{and} \quad R < 2R_0 \quad \textit{in } [t_1, t_2],$$

with α as specified in Section 2. Assume that $\epsilon = R_0/15$ and λ is contained in $(2/3, 1)$. Then there exists a universal constant C and some $\delta > 0$ such that if

$$\frac{W}{K} < C\lambda^2 R_0^2(1-\lambda) \quad \textit{and} \quad t_2 - t_1 \geq \delta, \tag{4.32}$$

then we have

$$f^2\big(\mathbb{T}\backslash(L_\gamma^+(T_{-1})_t)_\epsilon\big) \leq f^2(L_\alpha^-(t_1))e^{K(2\delta R_0 - \frac{(t-t_1-\delta)R_0\sin(\alpha)}{2})} \quad \textit{in } [t_1+\delta, t_2]. \tag{4.33}$$

Moreover, we can choose δ so that

$$\delta \lesssim \frac{1}{K\lambda R_0\cos^2(\alpha)} + \frac{\sin(\alpha)}{K\lambda R_0}\log\frac{1}{R_0}. \tag{4.34}$$

Proof. We will show how to select C appropriately at the end of the proof. For the moment, let us make it small enough that we can use Lemmas 3.5 and 4.3. The proof is based on Lemma 3.5, Proposition 4.1, Lemma 4.3, and the following differential inequalities:

$$\begin{aligned} &\frac{d}{dt}\underline{P} \geq K\lambda R_0\sqrt{1-\underline{P}^2}\Big(\sqrt{1-\underline{P}^2} - \frac{4\cos(\alpha)}{5}\Big) && \text{in } [t_1,t_2]\cap\big\{|P| \leq \sin(\alpha)\big\}, \\ &\frac{d}{dt}(1-P) \leq -\frac{1}{4}\sin(\alpha)K\lambda R_0(1-P) && \text{in } [s,t_2]\cap\big\{P \leq 1 - \tfrac{R_0}{15}\big\}. \end{aligned} \tag{4.35}$$

These inequalities hold when $\underline{P} = \cos(\underline{\Theta}_{r,t}(\theta,\omega) - \phi)$ for any r and for any θ satisfying $\cos(\theta - \phi(r)) = -\sin(\alpha)$ in $[t_1, t_2]$, and when $P = \cos(\Theta_{r',t}(\theta,\omega) - \Theta_{T_{-1},t}(\theta',\omega'))$ for any r' in $[t_1, t_2]$ and any θ and θ' such that $\cos(\theta - \phi(r')) \geq \sin(\alpha)$ and θ' is contained in $L_\gamma^+(T_{-1})$. Here, ω and ω' are contained in $[-W, W]$.

We claim that the inequalities imply that there exists $\delta > 0$ satisfying (4.34) such that

$$\mathbb{T}\backslash(L_\gamma^+(T_{-1})_s)_\epsilon \subset L_\alpha^-(s-\delta)_s$$

for any s in $[t_1+\delta, t_2]$. Here, we are using the notation introduced in Section 2.4 and in Definition 4.1. We divide the proof into three steps, the second of which will be the proof of the claim:

Step 1: We show that the claim implies (4.33). To achieve this let s be contained in $[t_1 + \delta, t_2]$. Then, using Lemma 3.5 and Proposition 4.1, on the interval $[t_1, s-\delta]$ we obtain

$$f^2(L_\alpha^-(s-\delta)) \leq f^2(L_\alpha^-(t_1))e^{-K(\frac{(s-\delta-t_1)KR_0\sin(\alpha)}{2})}.$$

Consequently, once the claim is proven, the lemma would follow by the above inequality and Lemma 4.1.

Step 2*: We show how the inequalities in* (4.35) *imply the claim.* Consider a time r contained in $[t_1, t_2 - \delta]$. Since we are assuming that $\underline{P}(r) = -\sin(\alpha)$, the first inequality in (4.35) implies that there exists $\underline{\delta} > 0$ such that

$$\frac{d}{dt}\underline{P} \geq \frac{K\lambda R_0 \cos^2(\alpha)}{5} \quad \text{in } [r, r + \underline{\delta}].$$

In particular, we can find $\underline{\delta}$ such that the above property holds, $\underline{P}(r + \underline{\delta}) = \sin\alpha$, and

$$\underline{\delta} \leq \frac{10\alpha}{K\lambda R_0 \cos^2(\alpha)}.$$

By the definition of $\underline{P}$ this implies that

$$\mathbb{T}\backslash L^+_\alpha(s) \subset L^-_\alpha(s - \underline{\delta})_s$$

for any s in $[t_1 + \delta, t_2]$. To derive this implication, we have set $s = r + \underline{\delta}$. Consequently, if we let θ be any element $\mathbb{T}\backslash L^-_\alpha(s - \underline{\delta})_s$ and we set $r' = r + \underline{\delta}$ in the definition of P then, by Lemma 4.3 and construction, we have $P(r') > -1$. Moreover, by integrating the second inequality in (4.35) we have that we can find $\bar{\delta} > 0$ such that $P(s + \underline{\delta} + \bar{\delta}) \geq 1 - \frac{R_0}{15}$ and

$$\bar{\delta} \lesssim \frac{\sin(\alpha)}{K\lambda R_0} \log \frac{1}{R_0}.$$

Thus, by the construction of P we obtain

$$\mathbb{T}\backslash(L^+_\gamma(T_{-1})_{r+\underline{\delta}+\bar{\delta}})_\epsilon \subset \mathbb{T}\backslash L^+_\alpha(r + \bar{\delta})_{r+\underline{\delta}+\bar{\delta}}.$$

Consequently, the claim follows by selecting $s = r + \underline{\delta} + \bar{\delta}$ and $\delta = \underline{\delta} + \bar{\delta}$.

Step 3*: We derive* (4.35). Let us denote

$$\underline{\Theta} = \Theta_{r,t}(\theta, \omega), \quad \Theta = \Theta_{r',t}(\theta, \omega), \quad \text{and} \quad \Theta' = \Theta_{T_{-1},t}(\theta', \omega).$$

To derive the first inequality, observe that thanks to Lemma 3.4 and our assumption on $\dot{R}$, we can select the constant in (4.32) appropriately so that we can guarantee that

$$\begin{aligned}
\frac{d}{dt}\cos(\underline{\Theta} - \phi) &= -\sin(\underline{\Theta} - \phi(t))(\dot{\underline{\Theta}} - \dot{\phi}) \\
&= -\sin(\underline{\Theta} - \phi)(\omega - KR\sin(\underline{\Theta} - \phi) - \dot{\phi}) \\
&\geq -|\sin(\underline{\Theta} - \phi)|\left(\frac{1}{R}\sqrt{K\frac{d}{dt}R^2 + W^2} + W - KR|\sin(\underline{\Theta} - \phi)|\right) \\
&\geq |\sin(\underline{\Theta} - \phi)|\left(KR|\sin(\Theta - \phi)| - \frac{4K\lambda R_0\cos(\alpha)}{5}\right).
\end{aligned}$$

Here, in the third inequality, we have used Lemma 3.4. Consequently, $\underline{P}$ satisfies the inequality

$$\frac{d}{dt}\underline{P} \geq K\lambda R_0\sqrt{1 - \underline{P}^2}\left(\sqrt{1 - \underline{P}^2} - \frac{4\cos(\alpha)}{5}\right).$$

Thus, the first inequality in (4.35) follows. Finally, to derive the second inequality we use the same argument as in the derivation of (4.9) to obtain

$$\frac{dP}{dt} \geq 2K\sqrt{1-P^2}\Big[R\cos\Big(\frac{\Theta+\Theta'}{2}-\phi\Big)\sqrt{\frac{1-P}{2}}-\frac{W}{K}\Big] \quad \text{in } [t_1,t_2].$$

Now, using the same arguments as in the proof of inequality (4.35) and equation (4.20), we obtain

$$\begin{aligned}\cos\Big(\frac{\Theta+\Theta'}{2}-\phi\Big) &= \cos\Big(\frac{(\Theta-\phi)+(\Theta'-\phi)}{2}\Big)\\ &\geq \frac{\cos(\Theta-\phi)+\cos(\Theta'-\phi)}{2}\\ &\geq \frac{\sin(\alpha)+\frac{3}{5}R_0}{2} \geq \frac{\sin(\alpha)}{2}.\end{aligned}$$

Here, we have used the fact that the first inequality in (4.35) implies that $\cos(\Theta-\phi) \geq \sin(\alpha)$ in $[r', t_2]$. Thus, we deduce that whenever $1-P \geq R_0/15$, we have

$$\begin{aligned}\frac{dP}{dt} &\geq 2K\sqrt{1-P^2}\Big[\frac{\lambda R_0\sin(\alpha)}{2}\sqrt{\frac{1-P}{2}}-\frac{W}{K}\Big]\\ &\geq 2K\sqrt{1+P}\Big[\frac{\sqrt{2}}{4}\lambda R_0\sin(\alpha)(1-P)-\frac{W}{K}\sqrt{1-P}\Big].\end{aligned}$$

Consequently, by choosing C appropriately in (4.32) so that

$$\frac{W}{K}\sqrt{1-P} < CR_0^2 < \frac{\lambda R_0\sin(\alpha)}{20}(1-P) \quad \text{whenever } 1-P \geq \frac{R_0}{15},$$

we can guarantee that

$$\frac{d}{dt}P \geq \frac{K\lambda R_0}{4}(1-P) \quad \text{whenever } P \leq 1-\frac{R_0}{15}.$$

Hence, the desired result follows. ∎

We close this section with a lemma that will allow us to control the L^2 norm of the solution in $\mathbb{T}\backslash(L_\gamma^+(T_{-1})_t)_\epsilon$ in the intervals of high entropy production.

Lemma 4.5. *Let $[t_1,t_2]$ be a time interval contained in $[T_{-1},\infty)$ with the property that*

$$R < 2R_0 \quad \text{in } [t_1,t_2].$$

Then we have

$$f^2(L_\alpha^-(t)) \leq f^2\big(\mathbb{T}\backslash(L_\gamma^+(T_{-1})_{t_1})_\epsilon\big)e^{2KR_0(t-t_1)} \quad \text{in } [t_1,t_2].$$

Proof. This lemma follows directly from Lemma 4.1 and Corollary 4.1. ∎

5. Average entropy production via differential inequalities

In this section we analyze the system of inequalities presented in Section 2.5 and derived in Sections 3.3 and 4. We will demonstrate that this system provides enough control to quantify the time T_0 presented in Theorem 1.1. With this control in hand, we are able to conclude the proof of Lemma 1.1, by quantifying the phase concentration phenomenon. We begin by describing a subdivision of the interval $[0, T_0]$ inspired by the treatment in [19].

5.1. The subdivision

Since this section is lengthy, we will first sketch the various steps and we will discuss their role in the final quantification of T_0. For clarity of the presentation, we also describe the use of the main results developed in Sections 3.3 and 4 in each step of the proof.

Step A: Building the subdivision. In Section 5.1.1 we will build the time subdivision of $[0, T_0]$. It will consist of a family of maximal time intervals, which we classify as *good* or *bad* depending on whether the dissipation of the system stays above or below a certain threshold. For technical reasons (which will improve our estimate on T_0), the intervals of our subdivision will be subordinated to an initial division according to the different scales of values of the order parameter. Namely, we will first divide the life span according to the times when R^2 doubles its value, yielding a “dyadic” hierarchy of intervals. Our final subdivision will consist in splitting each element of the dyadic hierarchy into good and bad subintervals. For later use, the threshold will be prescribed by the entropy production principle in Lemma 3.2 and the instability estimate of antipodal equilibria in Proposition 4.1, and it will depend on the specific scale of the order parameter in between the various doubling times.

Step B: Initial time of the subdivision. In Section 5.1.2 we will estimate the size of the first time t_0 of the subdivision. For the machinery in Section 4 to work, we will prescribe t_0 as the first time of formation of an attractor, according to Corollary 4.1. Then it will be crucial to quantify the size of this first time t_0. This will be the main objective of this step.

Step C: Gain vs loss on R. In Section 5.1.3 we will quantify the balance between the gain on the order parameter along any good interval, and the eventual loss of order parameter along any bad interval. To this end, we will exploit the entropy production principle in Lemma 3.2 along good intervals, and the universal lower bound in Corollary 3.6 along bad intervals. This control will be useful in the following Steps D and E.

Step D: Total number of good intervals. In Section 5.1.4 we will estimate an upper bound on the total number of good intervals. We note that, by definition, after any bad interval there must exist a good interval, so that the total number of bad intervals automatically gets controlled by the total number of good intervals. This step requires sharp knowledge of the gain vs loss of order parameter in Step C.

Step E: Total length of good intervals. In Section 5.1.5 we will quantify the total length of all good intervals. To this end, we will use the information that, by the definition of good interval, the slope of the order parameter must stay above the threshold. This step will exploit again the precise balance obtained in Step C.

Step F: Control of $f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon)$. In Section 5.1.6 we will quantify a time-dependent barrier which will serve as an upper bound for the evolution of the L^2 norm outside the attractor set. To this end, we will exploit the quantification in Section 4.2 for the growth/decay of such a quantity along the various intervals of the subdivision. Specifically, along long bad intervals we use the exponential decay estimate in Lemma 4.4. Otherwise, if intervals are good, or bad and short, we use the growth estimate in Lemma 4.5. Note that we do not know a priori the distribution of good and bad (either long or short) intervals. However, it is clear that the worst situation is the one where all the early bad intervals are short, since it does not allow for intermediate fall-off of the barrier.

Once we have all the above information, in Section 5.2 we will conclude the quantification of T_0 and the phase-concentration estimate in Lemma 1.1. Note that T_0 comes as the total length of the time intervals with a growing contribution in the above barrier for $f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon)$. In other words, we require a precise control on t_0, the total length of good intervals, and bad and short intervals, which we know by Steps B, D, and E. After such a T_0, the barrier decreases exponentially fast, which implies exponential phase concentration thanks to Jensen's inequality.

5.1.1. Defining the intervals of the subdivision. Here, we give the precise construction of our subdivision. Before we enter into details, we will introduce further notation that we will use throughout this part.

The dyadic hierarchy. Let us consider an auxiliary time partition into subintervals $[r_k, r_{k+1})$ whose endpoints are enumerated in the sequence $\{r_k\}_{k\in\mathbb{N}}$. The partition will be used in this part and is set according to the dyadic behavior of the square of the order parameter R^2. Namely, the sequence provides the first times at which R^2 doubles its value. To this end, let us set $R_0 = R(0)$ and $r_0 = 0$. Additionally, assume that R_k and r_k are given for certain $k \in \mathbb{N}$ and let us define

$$R_{k+1}^2 := 2R_k^2 \quad \text{and} \quad r_{k+1} := \inf\left\{t \geq r_k : R^2(t) \geq 2R_k^2 = R_{k+1}^2\right\}. \tag{5.1}$$

Since R is bounded by 1, then the sequence consists of finitely many terms

$$0 = r_0 < r_1 < \cdots < r_{k_*} < r_{k_*+1} = \infty.$$

Here and throughout this section, we will assume that

$$\frac{W}{K} \leq C R_0^3 \quad \text{and} \quad 1 - \lambda \leq \frac{\cos^2(\alpha)}{180} R_0, \tag{5.2}$$

with C small enough that all the results in Sections 3.3 and 4 hold (note that our assumption on λ implies the lower bound $\lambda > 179/180$ and thus we can suppress λ from the previous constraints on the universal constant C).

Now, let us set

$$\mu_k := \frac{1}{4}\lambda^3 R_k^3 \cos^2(\alpha), \quad d_k := \frac{1}{3KR_k}\log 10, \quad \text{and} \quad \delta_k := \frac{1}{KR_k}\log\Big(\frac{1}{R_k}\Big). \tag{5.3}$$

Observe that (5.2) implies $\frac{W}{K} \le C\lambda^2(1-\lambda)R_k^2$ for any $k = 0, \dots, k_*$, with the same universal constant C. In particular, we can use Lemma 3.5 to obtain

$$R(t) \ge \lambda R_k \quad \text{for all } t \text{ in } [r_k, r_{k+1}). \tag{5.4}$$

Initial time of the subdivision. Let us use Lemma 4.3 to define the corresponding times of formation of attractors. That is, we set

$$T_{-1}^k := \inf\{t \ge r_k : \tfrac{dR}{dt} \le KQR_k^3\}, \tag{5.5}$$

where $k = 0, \dots, k_*$ and Q is chosen so that we meet condition (4.22) when one applies Lemma 4.3 after translating the system in time. Here, for each k, we select the time translation so that the configuration of the system at time r_k is the new initial condition (recall that, by the definition of r_k, we can use Lemma 4.3 with the same universal constant C). Then we define

$$t_0 := \min\{T_{-1}^k : k = 0, \dots, k_*\} \tag{5.6}$$

and

$$k_0 := \max\{k \in \mathbb{Z}_0^+ : r_k \le t_0\}.$$

Notice that since t_0 is the first time in the subdivision, Lemma 4.3 and Corollary 4.1 will apply at any later step. Thus, we will obtain a controlled behavior of the characteristic flow close to the attractor set $(L_\gamma^+(t_0)_t)_\epsilon$. Here, and throughout the rest of this section, we will choose γ by the condition

$$\cos^2\gamma = \frac{1}{30}R_{k_0}. \tag{5.7}$$

We have done so according to condition (4.16).

The subdivision. Subordinated to the "dyadic" sequence $\{r_k\}_{k=0}^{k_*}$, we will construct the sequence of times $\{t_l\}_{l\in\mathbb{N}}$ describing the subdivision in the following way. We start at the time t_0 specified in Lemma 5.1. Assume that for some l in $\mathbb{N}$ the time t_l is given and let us proceed with the construction of t_{l+1}. First, consider the only $k(l)$ in $\{0, \dots, k_*\}$ such that t_l is contained in $[r_{k(l)}, r_{k(l)+1})$. Then we will distinguish between two different situations:

(1) If $\dot R(t_l) < K\mu_{k(l)}$, then we set

$$t_{l+1} := \sup\{t \in [t_l, r_{k(l)+1}) : \dot R(s) < K\mu_{k(l)} \ \forall s \in [t_l, s)\}. \tag{5.8}$$

(2) If $\dot R(t_l) \ge K\mu_{k(l)}$, then we first compute

$$\tilde t_{l+1} := \sup\{t \in [t_l, r_{k(l)+1}) : \dot R(s) \ge K\mu_{k(l)} \ \forall s \in [t_l, s)\}, \tag{5.9}$$

and set t_{l+1} via the following correction:

$$t_{l+1} = \begin{cases} \tilde{t}_{l+1} + d_{k(l)} & \text{if } \tilde{t}_{l+1} + d_{k(l)} \leq r_{k(l)+1}, \\ r_{k(l)+1} & \text{otherwise.} \end{cases} \tag{5.10}$$

The good and the bad sets. We can think of the intervals $[t_l, t_{l+1})$ obeying the above first item as *bad sets* as they are subject to "small" slope of the order parameter. On the contrary, those sets obeying the second item can be thought of as *good sets*, as they involve "large" slope of the order parameter in comparison with the critical value $K\mu_{k(l)}$. The critical value itself depends on the size of $R^2_{k(l)}$ in the above dyadic hierarchy as depicted in (5.3). For this reason, we will collect all the indices l of good and bad sets associated to the index k of the dyadic hierarchy as follows:

$$\begin{aligned} G_k &:= \left\{ l \in \mathbb{Z}_0^+ : t_l \in [r_k, r_{k+1}) \text{ and } \dot{R}(t_l) \geq K\mu_k \right\}, \\ B_k &:= \left\{ l \in \mathbb{Z}_0^+ : t_l \in [r_k, r_{k+1}) \text{ and } \dot{R}(t_l) < K\mu_k \right\}, \end{aligned} \tag{5.11}$$

for every $k = 0, \ldots, k_*$. Equivalently, we will say that $[t_l, t_{l+1})$ is of *type* G_k if $l \in G_k$ and it is of *type* B_k if $l \in B_k$. For notational purposes, we will denote their sizes as

$$\begin{aligned} g_k &:= \# G_k, \\ b_k &:= \# B_k, \end{aligned} \tag{5.12}$$

for every $k = 0, \ldots, k_*$. Notice that as a consequence of definition (5.11), after any interval of type B_k whose closure is properly contained in $[r_k, r_{k+1})$ there is an interval of type G_k. The reverse statement is not necessarily true. Namely, notice that for any l in G_k, we need first to compute the interval $[t_l, \tilde{t}_{l+1})$ according to (5.9) and later we extend it into the interval of type G_k $[t_l, t_{l+1})$. Unfortunately, the slope $\dot{R}$ can both grow or decrease in $[\tilde{t}_{t+1}, t_{l+1})$ and we then lose the control of what is next: either a G_k or B_k set. Nevertheless, this is enough to show that

$$b_k \leq g_k + 1 \quad \text{for all } k = 0, \ldots, k_*. \tag{5.13}$$

Of course, by definition, $g_0 = \cdots = g_{k_0 - 1} = 0$. The size of g_k for $k = k_0, \ldots, k_*$ will be estimated in Lemma 5.3. Finally, for notational simplicity, we will sometimes enumerate the indices in G_k in an increasing manner, namely,

$$G_k = \left\{ l_m^k : m = 1, \ldots, g_k \right\},$$

where $\{l_m^k\}_{1 \leq m \leq g_k}$ is an increasing sequence for each $k = 0, \ldots, k_*$.

5.1.2. Bound of the size of t_0. By Lemma 4.3 we have that each T_{-1}^k can be estimated via (4.18). However, we will show that our dyadic choice allows us to get a sharper estimate of t_0. More specifically, the cubic exponent for R_0 in (4.18) can be relaxed to a quadratic one. This is the content of the following lemma.

Lemma 5.1 (Bound of t_0). *Let t_0 be defined as above and suppose condition (5.2) holds. Then we have*

$$t_0 \lesssim \frac{1}{KR_0^2}.$$

Proof. By construction, it is clear that $k_0 \le k_*$. By the fundamental theorem of calculus and the definition of t_0, we obtain

$$R(r_{k+1}) - R(r_k) = \int_{r_k}^{r_{k+1}} \dot{R}(t)\,dt \ge KQR_k^3(r_{k+1} - r_k)$$

and

$$R(t_0) - R(r_{k_0}) = \int_{k_0}^{t_0} \dot{R}(t)\,dt \ge KQR_{k_0}^3(t_0 - r_{k_0}),$$

for any $k = 0, \dots, k_0 - 1$. Here, we have used the fact that $r_k \le t_0 \le T_{-1}^k$ for every $k = 0, \dots, k_0$. By estimate (5.6) and the definition of T_{-1}^k in (5.5) we can control the time derivative of the order parameter in the above integrals. Using the dyadic definition of r_k we arrive at the bounds

$$r_{k+1} - r_k \le Q\frac{(R(r_{k+1}) - R(r_k))}{KR_k^3} \le \frac{1}{2}\frac{Q}{KR_k^2} \tag{5.14}$$

and

$$t_0 - r_{k_0} \le \frac{Q(R(t_0) - R(r_{k_0}))}{KR_{k_0}^3} \le \frac{1}{2}\frac{Q}{KR_{k_0}^2}, \tag{5.15}$$

for any $k = 0, \dots, k_0 - 1$. To conclude the proof of the lemma, we represent t_0 via a telescopic sum

$$t_0 = t_0 - r_{k_0} + \sum_{k=0}^{k_0-1}(r_{k+1} - r_k) \le \frac{1}{2}\frac{Q}{KR_{k_0}^2}\sum_{k=0}^{k_0}\left(\frac{1}{2}\right)^k \le \frac{Q}{KR_0^2}. \qquad \blacksquare$$

5.1.3. Gain vs loss. In the forthcoming parts, we compare the growth of the order parameter R along intervals of type G_k with its loss on intervals of type B_k. To do this precisely, for each k in $\{k_0, \dots, k_*\}$, we have to give special consideration to the last interval of the subdivision in each $[r_k, r_{k+1})$. We will denote such terminal intervals by $[t_{l(k)}, t_{l(k)+1})$ in such a way that $t_{l(k)}$ is in $[r_k, r_{k+1})$ and $t_{l(k)+1} = r_{k+1}$. We will use the ideas in Corollary 3.6. In the following lemma, we will see that assumption (5.2) implies that the loss in R^2 is smaller than $4/5$ of the gain (except for possibly the terminal interval $[t_{l(k)}, t_{l(k)+1})$).

Lemma 5.2 (Gain vs loss). *Assume that condition (5.2) holds. Then we have*

$$R^2(t_l) - R^2(t_{l+1}) \le \frac{4}{5}(R^2(t_{l_m^k+1}) - R^2(\tilde{t}_{l_m^k+1})) \le \frac{4}{5}(R^2(t_{l_m^k+1}) - R^2(t_{l_m^k}))$$

for any l in B_k and any l_m^k in $G_k \backslash l(k)$.

Proof. Thanks to Corollary 3.6 and Lemma 3.2 we have

$$R^2(t_l) - R^2(t_{l+1}) \leq (1-\lambda^2)R^2(t_l) \leq 4(1-\lambda)R_k^2$$

and

$$R^2(t_{l_m^k+1}) - R^2(\tilde{t}_{l_m^k+1}) \geq \frac{1}{40}\lambda^4 R_k^3.$$

In particular, our thesis holds true as long as one checks the inequality

$$4(1-\lambda) \leq \frac{1}{50}\lambda^4 R_k.$$

The inequality is true due to our choice of λ. Here we have used the fact that $\alpha = \pi/6$ and condition (5.2) implies that $\lambda > 179/180$. ■

5.1.4. Number of intervals of type G_k. Our objective here is to obtain an estimate of the numbers g_k for $k = k_0, \ldots, k_*$. Recall that due to (5.13), this will yield a control on the number of sets of type B_k.

Lemma 5.3 (Bound on g_k). *Assume that condition* (5.2) *holds. Then we have*

$$\max(b_k, g_k) \lesssim \frac{1}{R_k}.$$

Proof. To prove this, recall that by Lemma 3.2, we have

$$\sum_{l=G_k\setminus l(k)} (R^2(t_{l+1}) - R^2(t_l)) \geq (g_k - \chi_{\{l(k)\in G_k\}})\frac{\lambda^4 R_k^3}{40}. \tag{5.16}$$

Thus, Lemma 5.2 implies

$$\sum_{l\in B_k} (R^2(t_{l+1}) - R^2(t_l)) \geq -g_k\frac{\lambda^4 R_k^3}{50}. \tag{5.17}$$

Taking the sum of both the oscillations at good and bad sets, we recover a telescopic sum involving the evaluation of R^2 at the largest and smallest of the times t_l in $[r_k, r_{k+1})$. Recall that by construction, the oscillation of R^2 in $[t_{l(k)}, t_{l(k)+1})$ is positive, independently of whether $l(k)$ is in B_k or G_k. By doing this, we obtain

$$R_{k+1}^2 - \lambda^2 R_k^2 \geq \frac{g_k}{200}\lambda^4 R_k^3 - \frac{1}{40}\chi_{\{l(k)\in G_k\}}\lambda^4 R_k^3.$$

Hence, we deduce the bound

$$g_k \leq \frac{200(2-\lambda^2)R_k^2}{R_k^3} + 5. \tag{5.18}$$

Here, we have used the fact that assumption (5.2) implies that $\lambda > 179/180$. Hence, the desired result follows. ■

5.1.5. Sum of lengths of intervals of type G_k. In this section we control the total diameter of the intervals in G_k. To do this we will consider the sets $\mathring{G}_k$ and $\mathring{B}_k$. The set $\mathring{G}_k$ is obtained by deleting the biggest element from G_k if the last interval in $[r_k, r_{k+1})$ is of type G_k. Otherwise, we let $\mathring{G}_k = G_k$. On the other hand, the set B_k is obtaining by deleting the last element in B_k in the case where the intervals in $[r_k, r_{k+1})$ do not end with two or more intervals of type G_k. Otherwise, we let $\mathring{B}_k = B_k$. Now, we are ready to state our control.

Lemma 5.4. *The sum of the lengths of the intervals* $[t_{l_m^k}, t_{l_m^k+1}]$ *satisfies*

$$\sum_{m=1}^{g_k} (t_{l_m^k+1} - t_{l_m^k}) \lesssim \frac{1}{KR_k^2}.$$

Proof. Let us first bound the length of each time interval $[t_{l_m^k}, t_{l_m^k+1})$ of type G_k for $m = 1, \ldots, g_k$. Notice that, as defined in (5.10), we have the identity

$$t_{l_m^k+1} - t_{l_m^k} = (\tilde{t}_{l_m^k+1} - t_{l_m^k}) + d_k. \tag{5.19}$$

Our next goal is to estimate the first term. To this end, we will use the idea in Lemma 5.3 and the fundamental theorem of calculus to write

$$R(\tilde{t}_{l_m^k+1}) - R(t_{l_m^k}) = \int_{t_{l_m^k}}^{\tilde{t}_{l_m^k+1}} \dot{R}(t)\, dt \geq \frac{1}{4} K\lambda^3 R_k^3 \cos^2(\alpha)(\tilde{t}_{l_m^k+1} - t_{l_m^k})$$

for all $m = 1, \ldots, g_k$. Here, we have used (5.9) to bound the time derivative of R. Hence, we obtain

$$\tilde{t}_{l_m^k+1} - t_{l_m^k} \leq \frac{4}{K\lambda^3 R_k^3 \cos^2(\alpha)} (R(\tilde{t}_{l_m^k+1}) - R(t_{l_m^k})) \tag{5.20}$$

for all $m = 1, \ldots, g_k$. By summing over all the intervals of type $\mathring{G}_k$ we obtain

$$\begin{aligned} &\sum_{l \in \mathring{G}_k} (\tilde{t}_{l+1} - t_l) \\ &\quad \leq \frac{4}{K\lambda^3 R_k^3 \cos^2(\alpha)} \sum_{l \in \mathring{G}_k} (R(\tilde{t}_{l+1}) - R(t_l)) \\ &\quad = \frac{4}{K\lambda^3 R_k^3 \cos^2(\alpha)} \sum_{l \in \mathring{G}_k} [(R(t_{l+1}) - R(t_{l_m^k})) - (R(t_{l+1}) - R(\tilde{t}_{l+1}))]. \end{aligned} \tag{5.21}$$

Let us add and subtract from the first term in (5.21) the oscillations of R over all the sets of type $\mathring{B}_k$. Notice that after doing so the first term becomes a telescopic sum of evaluations of R at points t_l in $[r_k, r_{k+1})$ and it can be easily bounded by the oscillation of R between the largest and smallest t_l that lie in $[r_k, r_{k+1})$. In turn, it can be easily bounded by $R_{k+1} - \lambda R_k$ due to the definition of r_{k+1} in (5.1) and the lower bound on the order parameter given by (5.4). Then we obtain

$$\begin{aligned}\sum_{l\in\mathring{G}_k}(\tilde{t}_{l+1}-t_l)\\ &\leq \frac{4}{K\lambda^3 R_k^3\cos^2(\alpha)}(R_{k+1}-\lambda R_k)\\ &\quad -\frac{4}{K\lambda^3 R_k^3\cos^2(\alpha)}\bigg[\sum_{l\in\mathring{B}_k}(R(t_l)-R(t_{l+1}))+\sum_{l\in\mathring{G}_k}(R(t_{l+1})-R(\tilde{t}_{l+1}))\bigg].\end{aligned} \tag{5.22}$$

Our goal is to show that the term in the third line is non-positive. Indeed, let us use Lemmas 3.2 and 5.2 in the second term of (5.22) to obtain

$$\begin{aligned}\sum_{l\in\mathring{G}_k}(\tilde{t}_{l+1}-t_l) &\leq \frac{4(2-\lambda)}{K\lambda^3 R_k^2\cos^2(\alpha)}-\frac{4}{5\cos^2(\alpha)K\lambda^3 R_k^3}\sum_{l\in\mathring{G}_k}(R(t_{l+1})-R(\tilde{t}_{l+1}))\\ &\leq \frac{4(2-\lambda)}{K\lambda^3 R_k^2\cos^2(\alpha)}.\end{aligned}$$

Hence, by Lemmas 3.2 and 5.3 and (5.19) we deduce that

$$\sum_{m=1}^{g_k}(t_{l_m^k+1}-t_{l_m^k}) \leq d_k g_k + \tilde{t}_{l_{g_k+1}^k} - t_{l_{g_k}^k} + \sum_{l\in\mathring{G}_k}(\tilde{t}_{l+1}-t_l) \lesssim \frac{1}{KR_k^2},$$

where we have used (5.20) and our usual bound on the oscillation to control the difference $(\tilde{t}_{l_{g_k+1}^k} - t_{l_{g_k}^k})$. Thus, the desired result follows. ■

5.1.6. Growth of $f^2(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon)$. Our goal here is to control the growth of $f^2(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon)$ in each interval $[r_k, r_{k+1})$, where the parameter ϵ of the neighborhood is set once for all as

$$\epsilon := \frac{R_0}{15}.$$

Notice that ϵ has been set so that the attractive property in Corollary 4.1 holds true. To initialize the iterative method, we need to control $f^2(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon)$ at $t = t_0$. Hence, we begin by providing a control of the growth of $f_t^2(\mathbb{T})$ during the transient $[0, t_0]$.

Lemma 5.5. *Assume condition (5.2) holds. Then we have*

$$\|f_{t_0}\|_2^2 \leq \|f_0\|_2^2 e^{\frac{4Q}{R_0}}.$$

Proof. Thanks to Proposition 4.1 we obtain

$$\|f_{t_0}\|_2^2 \leq \|f_0\|_2^2 \exp\bigg(K\int_0^{t_0} R(s)\,ds\bigg).$$

Then the main objective is to estimate the time integral of the order parameter. To that end, observe that

$$\begin{aligned}\int_0^{t_0} R(s)\,ds &= \sum_{k=0}^{k_0-1}\int_{r_k}^{r_{k+1}} R(s)\,ds + \int_{r_{k_0}}^{t_0} R(s)\,ds\\ &\leq \sum_{k=0}^{k_0-1} R_{k+1}(r_{k+1}-r_k) + R_{k_0+1}(t_0-r_{k_0})\\ &\leq Q\sum_{k=0}^{k_0}\frac{R_k}{KR_k^2} = Q\sum_{k=0}^{k_0}\frac{1}{KR_0}\Big(\frac{\sqrt{2}}{2}\Big)^k \leq \frac{4Q}{KR_0}.\end{aligned}$$

Notice that we have used (5.14) and (5.15) to estimate the lengths of the intervals $[r_k, r_{k+1})$. Hence, the desired result follows. ∎

Let us now begin our study of the primary goal of this section. To do this, let us introduce the following notation that we will use in this part. Define the parameters

$$D_k := \max(b_k, g_k)(\delta_k + d_k) + \sum_{l=1}^{g_k}(\tilde{t}_{l_m^k+1} - t_{l_m^k}) \tag{5.23}$$

for any $k = k_0, \ldots, k_*$. Notice that its size can be controlled in the following way due to Lemmas 5.3 and 5.4 and the values in (5.3):

$$D_k \lesssim \frac{1}{KR_k^2} + \frac{1}{R_k}\Big[\frac{1}{KR_k}\log\Big(\frac{1}{R_k}\Big) + \frac{1}{KR_k}\Big] \approx \frac{1}{KR_k^2}\log\Big(1+\frac{1}{R_k}\Big). \tag{5.24}$$

Let us also introduce the following sequence of functions $\{F_k\}_{k=k_0}^{k_*}$. We proceed by induction. For $k = k_0$, we define

$$F_{k_0}(t) := \begin{cases} \|f_0\|_2^2 e^{\frac{4Q}{R_0}} e^{2KR_{k_0}(t-t_0)} & \text{for } t \in [t_0, t_0 + D_{k_0}],\\ \|f_0\|_2^2 e^{\frac{4Q}{R_0}} e^{2KR_{k_0}D_{k_0}} e^{-K\frac{R_{k_0}\sin(\alpha)}{2}(t-t_0-D_{k_0})} & \text{for } t \in [t_0 + D_{k_0}, r_{k_0+1}).\end{cases}$$

Assume that F_{k-1} is given in the interval $[r_{k-1}, r_k)$ and let us define F_k in the interval $[r_k, r_{k+1})$ through the formula

$$F_k(t) := \begin{cases} F_{k-1}(r_k)e^{2KR_k(t-r_k)} & \text{for } t \in [r_k, r_k + D_k],\\ F_{k-1}(r_k)e^{2KR_kD_k}e^{-K\frac{R_k\sin(\alpha)}{2}(t-r_k-D_k)} & \text{for } t \in [r_k + D_k, r_{k+1}).\end{cases}$$

Lemma 5.6. *Assume condition (5.2) holds; then we have*

$$F_k(t) \leq \|f_0\|_2^2 e^{\frac{B}{KR_0}\log(1+\frac{1}{R_0})}, \quad t \in [r_k, r_{k+1}),$$

for some universal constant B and for each $k = k_0, \ldots, k_*$.

Proof. By definition, we note that

$$F_k(t) \leq F_{k-1}(r_k)e^{2KR_kD_k} \quad \text{for all } t \in [r_k, r_{k+1})$$

and for every $k = k_0 + 1 \ldots, k_*$. Also, notice that by construction we have

$$F_{k_0}(r_{k_0+1}) \leq \|f_0\|_2^2 e^{\frac{4Q}{R_0}} e^{2KR_{k_0}D_{k_0}}.$$

Then a simple induction shows that

$$F_k(t) \leq \|f_0\|_2^2 e^{\frac{4Q}{R_0}} \prod_{q=k_0}^{k} e^{2KR_qD_q} = \|f_0\|_2^2 \exp\left(\frac{4Q}{R_0} + \sum_{q=k_0}^{k} 2KR_qD_q\right). \tag{5.25}$$

Finally, let us use the bound (5.24) on the above sum to achieve

$$\sum_{q=k_0}^{k} 2KD_qR_q \lesssim \sum_{q=k_0}^{k} \frac{R_q}{KR_q^2} \log\left(1 + \frac{1}{R_q}\right) \lesssim \frac{1}{KR_0} \log\left(1 + \frac{1}{R_0}\right) \sum_{q=k_0}^{k} \left(\frac{\sqrt{2}}{2}\right)^q.$$

Hence, the desired result follows. ∎

The sequence $\{F_k\}_{k=k_0}^{k_*}$ has been constructed as a barrier in order to control the map $t \to f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon)$ at each interval $[r_k, r_{k+1})$. We achieve this in the following theorem. This theorem is the main result in this section. As a by-product, we will derive Corollary 5.1 below and we will prove Lemma 1.1.

Theorem 5.1. *Assume that condition* (5.2) *holds; then we have*

$$f^2\left(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon\right) \leq F_k(t),\ t \in [r_k, r_{k+1})$$

for each $k = k_0, \ldots, k_*$.

Proof. We proceed by induction:

Step 1*: Base case (*$k = k_0$*).* Notice that the inequality is true at $t = t_0$ thanks to Lemma 5.5. Let us now look at each of the intervals of type G_{k_0} and B_{k_0} and quantify the growth or decay rate of $f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon)$ via Lemmas 4.1, 4.4, and 4.5. Specifically, we will distinguish between three different scenarios for each interval $[t_l, t_{l+1})$ with t_l in $[r_{k_0}, r_{k_0+1})$:

(1) If the interval is of type G_{k_0}, then $\dot{R}(t_l) \geq K\mu_{k_0}$ and Lemma 4.4 cannot be used to quantify a decrease estimate of the L^2 norm. Fortunately, we can at least use Lemma 4.5 on the sliding L^2 norm in combination with Corollary 4.1 to obtain

$$\begin{aligned} f^2(L_\alpha^-(t)) &\leq f^2\left(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_{\epsilon,t}\right) \\ &\leq f^2\left(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_\epsilon\right)e^{2KR_{k_0}(t-t_l)} \\ &\leq f^2\left(\mathbb{T} \setminus \left(L_\gamma^+(t_0)_{t_l}\right)_\epsilon\right)e^{2KR_{k_0}(t_{l+1}-t_l)} \end{aligned}$$

for every t in $[t_l, t_{l+1})$.

(2) If the interval is of type B_{k_0}, then there are two different possibilities: either $[t_l, t_{l+1})$ is small or it is large.

(a) If $[t_l, t_{l+1})$ is small (i.e., $t_{l+1} - t_l \leq \delta_{k_0}$), then Lemma 4.4 cannot be used either. Then we have to rely on a similar argument to that of type G_k, and it implies

$$\begin{aligned} f^2(L_\alpha^-(t)) &\leq f^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_\epsilon\big) e^{2KR_{k_0}(t-t_l)} \\ &\leq f^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_\epsilon\big) e^{2KR_{k_0}\delta_{k_0}} \end{aligned}$$

for every t in $[t_l, t_{l+1})$.

(b) Finally, if $[t_l, t_{l+1})$ is large (i.e., $t_{l+1} - t_l > \delta_{k_0}$) then we can apply Lemma 4.4. However, notice that it can only be applied for t in $[t_l + \delta_{k_0}, t_{l+1})$, and in the remaining part of the interval $[t_l, t_l + \delta_{k_0})$ we can only apply the same argument as before, supported by Lemma 4.1 about the sliding L^2 norm. Specifically, for any t in $[t_l, t_l + \delta_k)$ Lemma 4.1 implies

$$f^2(L_\alpha^-(t)) \leq f^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_\epsilon\big) e^{2K\delta_{k_0} R_{k_0}}.$$

Now, for any t in $[t_l + \delta_k, t_{l+1})$, Lemmas 4.4 and 4.5 yield

$$\begin{aligned} f^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon\big) &\leq f^2(L_\alpha^-(t_l)) e^{K\left(2R_{k_0}\delta_{k_0} - \frac{(t-t_l-\delta_{k_0})R_{k_0}\sin(\alpha)}{2}\right)} \\ &\leq f_{t_l}^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_{t_l})_\epsilon\big) e^{K\left(2R_{k_0}\delta_{k_0} - \frac{(t-t_l-\delta_{k_0})R_0\sin(\alpha)}{2}\right)}. \end{aligned}$$

Bearing all these possibilities in mind, let us now show the inequality for F_{k_0} in (t_0, r_{k_0+1}). Fix any time t in (t_0, r_{k_0+1}) and consider the index

$$p := \max\big\{l \in \mathbb{N} : t_l \leq t\big\}.$$

Then we will repeat the above classification at each $[t_l, t_{l+1})$ with l in $\{0, \ldots, p-1\}$ ending with $[t_p, t)$. Also, let us split the indices of intervals of type B_{k_0} into two parts corresponding to small or large intervals as in the above discussion, namely,

$$\begin{aligned} B_{k_0}^S &:= \big\{l \in B_{k_0} : t_{l+1} - t_l \leq \delta_{k_0}\big\}, \\ B_{k_0}^L &:= \big\{l \in B_{k_0} : t_{l+1} - t_l > \delta_{k_0}\big\}. \end{aligned}$$

Notice that we then have the disjoint union

$$\{0, \ldots, p-1\} = G_{k_0,p} \cup B_{k_0,p}^S \cup B_{k_0,p}^L,$$

where $G_{k_0,p} = G_{k_0} \cap \{0, \ldots, p-1\}$, $B_{k_0,p}^S = B_{k_0}^S \cap \{0, \ldots, p-1\}$, and

$$B_{k_0,p}^L = B_{k_0}^L \cap \{0, \ldots, p-1\}.$$

By applying the above discussion in a recursive way, we obtain

$$\begin{aligned} & f_{t_p}^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_{t_p})_\epsilon\big) \\ & \quad \leq f_{t_0}^2(\mathbb{T})\exp\Bigg\{2R_{k_0}K\Bigg[\sum_{l\in G_{k_0,p}}(t_{l+1}-t_l)+\sum_{l\in B_{k_0,p}^S}\delta_{k_0}\Bigg] \\ & \qquad\qquad + \sum_{l\in B_{k_0,p}^L}\left(2R_{k_0}\delta_{k_0}-\frac{(t_{l+1}-t_l-\delta_{k_0})R_{k_0}\sin(\alpha)}{2}\right)\Bigg\}. \end{aligned} \tag{5.26}$$

Similarly, for any t in $(t_p, t_p+\delta_{k_0})$ we have

$$\begin{aligned} & f^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon\big) \\ & \quad \leq f_{t_p}^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_{t_p})_\epsilon\big) \\ & \qquad \times\exp\big\{2KR_{k_0}\big[(t_{p+1}-t_p)\chi_{\{p\in G_{k_0}\}}+\delta_{k_0}\chi_{\{p\in B_{k_0}^S\}}+\delta_{k_0}\chi_{\{p\in B_{k_0}^L\}}\big]\big\} \\ & \quad \leq f_{t_p}^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_{t_p})_\epsilon\big) \\ & \qquad \times\exp\big\{2KR_{k_0}\big[(t_{p+1}-t_p)\chi_{\{p\in G_{k_0}\}}+\delta_{k_0}\chi_{\{p\in B_{k_0}\}}\big]\big\}. \end{aligned} \tag{5.27}$$

Thus, for any t in $[t_p+\delta_{k_0}, t_{p+1})$ we obtain

$$\begin{aligned} & f^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon\big) \\ & \quad \leq f_{t_p}^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_{t_p})_\epsilon\big) \\ & \qquad \times\exp\Big\{2KR_{k_0}\Big[(t_{p+1}-t_p)\chi_{\{p\in G_{k_0}\}}+\delta_{k_0}\chi_{\{p\in B_{k_0}^S\}} \\ & \qquad\qquad +\Big(\delta_{k_0}-\frac{(t-t_p-\delta_{k_0})\sin(\alpha)}{4}\Big)\chi_{\{p\in B_{k_0}^L\}}\Big]\Big\} \\ & \quad \leq f_{t_p}^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_{t_p})_\epsilon\big) \\ & \qquad \times\exp\Big\{2KR_{k_0}\Big[(t_{p+1}-t_p)\chi_{\{p\in G_{k_0}\}}+\delta_{k_0}\chi_{\{p\in B_{k_0}\}} \\ & \qquad\qquad -\frac{(t-t_p-\delta_{k_0})R_0\sin(\alpha)}{4}\chi_{\{p\in B_{k_0}^L\}}\Big]\Big\}. \end{aligned} \tag{5.28}$$

Putting (5.26), (5.27), and (5.28) together and recalling D_k in (5.23) implies

$$\begin{aligned} f_t^2\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_t)_\epsilon\big) \leq f_{t_0}^2(\mathbb{T})\exp\Bigg\{2KD_{k_0}R_{k_0}-\sum_{l\in B_{k_0,p}}K\frac{(t_{l+1}-t_l)R_{k_0}\sin(\alpha)}{2} \\ -K\frac{(t-t_p)R_{k_0}\sin(\alpha)}{2}\chi_{\{p\in B_{k_0}\}}\Bigg\}, \end{aligned} \tag{5.29}$$

where we have absorbed the δ_{k_0} in the last term into D_{k_0}.

On the other hand, notice that we can recover t from the following telescopic sum:

$$
\begin{aligned}
t &= t - t_p + \sum_{l=0}^{p-1}(t_{l+1} - t_l) + t_0 \\
&= t_0 + (t - t_p)\chi_{\{p\in G_{k_0}\}} + (t - t_p)\chi_{\{p\in B_{k_0}\}} + \sum_{l\in G_{k_0,p}} (t_{l+1} - t_l) + \sum_{l\in B_{k_0,p}} (t_{l+1} - t_l) \\
&\leq t_0 + D_{k_0} + (t - t_p)\chi_{\{p\in B_{k_0}\}} + \sum_{l\in B_{k_0,p}} (t_{l+1} - t_l).
\end{aligned}
$$

Consequently,

$$
-(t - t_p)\chi_{\{p\in B_{k_0}\}} - \sum_{l\in B_{k_0,p}} (t_{l+1} - t_l) \leq -(t - t_0 - D_{k_0}),
$$

which can be used to bound the last two terms in the above exponential of (5.29). Then we obtain

$$
\begin{aligned}
&f_t^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon\big) \\
&\quad\leq \begin{cases} f_{t_0}^2(\mathbb{T})e^{2KD_{k_0}} & \text{for } t \in (t_p, t_p + \delta_{k_0}), \\ f_{t_0}^2(\mathbb{T})e^{2KD_{k_0} - \frac{KR_0\sin(\alpha)}{2}(t - t_0 - D_{k_0})} & \text{for } t \in [t_p + \delta_{k_0}, t_{p+1}). \end{cases}
\end{aligned}
\tag{5.30}
$$

Notice that the worst situation is the one where there is no intermediate fall-off, that is, $B^L_{k_0,p} = \emptyset$. Since this scenario dominates all the other possibilities, we will restrict to it without loss of generality. This amounts to the chain of inequalities

$$
\begin{aligned}
t_p + \delta_{k_0} &= t_0 + \delta_{k_0} + \sum_{l\in G_{k_0,p}} (t_{l+1} - t_l) + \sum_{l\in B^S_{k_0,p}} (t_{l+1} - t_l) + \sum_{l\in B^L_{k_0,p}} (t_{l+1} - t_l) \\
&\leq t_0 + \sum_{l\in G_{k_0}} (t_{l+1} - t_l) + \max(g_k, b_k)\delta_k \leq t_0 + D_{k_0},
\end{aligned}
$$

that is, $t_p + \delta_{k_0} \leq t_0 + D_{k_0}$, which leads to restating (5.30) as

$$
\begin{aligned}
&f_t^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon\big) \\
&\quad\leq \begin{cases} f_{t_0}^2(\mathbb{T})e^{2KD_{k_0}R_{k_0}} & \text{for } t \in (t_0, t_0 + D_{k_0}), \\ f_{t_0}^2(\mathbb{T})e^{2KD_{k_0}R_{k_0} - \frac{KR_{k_0}\sin(\alpha)}{2}(t - t_0 - D_{k_0})} & \text{for } t \in [t_0 + D_{k_0}, r_{k_0+1}). \end{cases}
\end{aligned}
$$

Finally, use Lemma 5.5 to relate the L^2 norm at $t = t_0$ and at $t = 0$. Thus, we have shown the claimed bound.

Step 2: Inductive hypothesis. Let us assume that for certain $k_0 < k < k_*$ we have

$$
f^2\big(\mathbb{T} \setminus (L_\gamma^+(t_0)_t)_\epsilon\big) \leq F_q(t), \quad t \in [r_q, r_{q+1}),
$$

for any $q < k$.

Step 3: *Induction step.* The proof for the index k becomes a simple consequence of the inductive hypothesis where again we need to apply Lemmas 4.1, 4.4, and 4.5 repeatedly in the spirit of Step 1 for the base step. ∎

As a consequence of Theorem 5.1 we obtain the following corollary.

Corollary 5.1. *Suppose assumption* (5.2) *holds. Then we have*

$$r_{k+1}-r_k\lesssim\frac{1}{KR_k}\frac{1}{R_0}\log\Big(1+\frac{1}{R_0}+W^{1/2}\|f_0\|_2\Big)$$

for any $k\leq k_*$.

Proof. Thanks to (5.24), we may assume without loss of generality that $r_{k+1}-r_k\geq D_k$. Now, observe that by Theorem 5.1 and (5.25) we have

$$\begin{aligned}f^2\big(\mathbb{T}\backslash(L_\gamma^+(t_0))_\epsilon\big)&\leq F_k(t)\\&\leq\|f_0\|_2^2e^{\frac{4Q}{R_0}}\Bigg(\prod_{q=k_0}^{k}e^{2KR_qD_q}\Bigg)e^{-K\frac{R_k\sin(\alpha)}{2}(t-r_k-D_k)}\\&\leq\|f_0\|_2^2e^{\frac{Q'}{R_0}\log(1+\frac{1}{R_0})}e^{-K\frac{R_k\sin(\alpha)}{2}(t-r_k-D_k)}\end{aligned}$$

for every t in $[r_k+D_k,r_{k+1})$ and some universal constant Q'. On the other hand, by Jensen's inequality we have

$$\rho\big(\mathbb{T}\backslash(L_\gamma^+(t_0)_t)_\epsilon\big)\leq\sqrt{4\pi Wf^2\big(\mathbb{T}\backslash(L_\gamma^+(t_0)_t)_\epsilon\big)}.$$

Consequently, if we let $m(s)=1-\rho(\mathbb{T}\backslash L_\gamma^+(t_0)_s)_\epsilon)$, using Theorem 5.1 we deduce that

$$1-m(s)\leq2\sqrt{\pi}\|f_0\|_2e^{\frac{Q'}{2KR_0}\log(1+\frac{1}{R_0})}e^{-K\frac{R_k\sin(\alpha)}{4}(s-r_k-D_k)}\tag{5.31}$$

for any s in $[D_k+r_k,r_{k+1}]$. On the other hand, by Lemmas 4.2, 4.3 and Corollary 4.1, if we let

$$P(t)=\inf_{\theta,\theta'\in(L_\gamma^+(t_0)_s)_{\epsilon,t}}\cos(\theta-\theta'),\tag{5.32}$$

we have

$$1-P(t)\leq\max\Big[\frac{1}{3}R_{k_0}e^{-\frac{K}{8}R_{k_0}(t-s)},\frac{16}{R_{k_0}^2}\frac{W^2}{K^2}\Big]$$

for every t in $[s,r_{k+1}]$. Additionally, using Lemmas 4.2, 4.3, and Corollary 4.1, if we let $L=(L_\gamma^+(t_0)_s)_\epsilon$ we have

$$\begin{aligned}R(t)&\geq\inf_{\theta,\theta'\in L_{s,t}}R\cos(\theta-\theta')\\&\geq m(s)P(t)-(1-m(s))\\&=\big(1-(1-m(s))\big)P(t)-(1-m(s))\\&\geq P(t)-2(1-m(s))\\&\geq1-(1-P(t))-4\sqrt{\pi}W^{\frac{1}{2}}\|f_0\|_2e^{\frac{Q'}{2R_0}\log(1+\frac{1}{R_0})}e^{-K\frac{R_k\sin(\alpha)}{4}(s-r_k-D_k)}.\end{aligned}\tag{5.33}$$

Now observe that, by construction,

$$\frac{\sqrt{2}}{2} \geq R \quad \text{in } [r_k, r_{k+1}).$$

Consequently, by (5.31) and (5.32), if we set $t = r_{k+1}$ and $s = r_{k+1} - \frac{8}{KR_{k_0}} \log \frac{1}{10R_{k_0}}$ in (5.33), and make C smaller within the constraints of (5.2) if necessary, we obtain

$$\begin{aligned}
&\frac{1}{3} R_{k_0} e^{-\log \frac{1}{10R_{k_0}}} \\
&\qquad + 4\sqrt{\pi} W^{\frac{1}{2}} \| f_0 \|_2 e^{\frac{Q'}{2R_0} \log(1+\frac{1}{R_0})} e^{-K \frac{R_k \sin(\alpha)}{4} (r_{k+1} - r_k - D_k - \frac{8}{KR_{k_0}} \log \frac{1}{10R_{k_0}})} \\
&\geq 1 - \frac{\sqrt{2}}{2}.
\end{aligned} \tag{5.34}$$

Thus,

$$4\sqrt{\pi} \| f_0 \|_2 W^{1/2} e^{\frac{C_1}{R_0} \log(1+\frac{1}{R_0})} e^{-K \frac{R_k \sin(\alpha)}{4} (r_{k+1} - r_k - D_k)} \geq 1 - \frac{\sqrt{2}}{2} - \frac{1}{30} \geq \frac{1}{10},$$

for some universal constant C_1.

Hence,

$$\frac{4}{KR_k \sin(\alpha)} \log(40\sqrt{\pi} W^{\frac{1}{2}} \| f_0 \|_2) + \frac{4C_1}{KR_0} \frac{1}{R_k \sin(\alpha)} \log\Big(1 + \frac{1}{R_0}\Big) + D_k \geq r_{k+1} - r_k.$$

Consequently, using (5.24) the desired result follows. ■

5.2. Proof of Lemma 1.1

We will prove the lemma by proving that

$$\rho\big(\mathbb{T} \backslash (L_\gamma^+(t_0)_s)_{\epsilon,t}\big) \leq e^{-\frac{1}{10} K \sin(\alpha)(t - T_0)} \tag{5.35}$$

and

$$(L_\gamma^+(t_0)_s)_{\epsilon,t} \subset L_\beta^+(t)$$

for every t in $[T_0, \infty)$. Here,

$$s = t - \frac{8}{KR_{k_*}} \log \frac{1}{40R_{k_*}}.$$

Additionally, recall that γ was chosen in (5.7).

We begin by showing the first equation in (1.22). To do this, we control r_{k_*} via the following telescopic sum and Corollary 5.1:

$$\begin{aligned}
r_{k_*} &= t_0 + \sum_{k=k_0}^{k_*} r_{k+1} - r_k \\
&\lesssim \frac{1}{KR_0^2} + \sum_{k=k_0}^{k_*} \frac{1}{KR_k} \frac{1}{R_0} \log\Big(1 + \frac{1}{R_0} + W^{1/2} \| f_0 \|_2\Big)
\end{aligned}$$

$$\lesssim \frac{1}{KR_0^2} + \sum_{k=k_0}^{k_*} \left(\frac{\sqrt{2}}{2}\right)^k \frac{1}{KR_0^2} \log\left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right)$$
$$\lesssim \frac{1}{KR_0^2} \log\left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right).$$

Consequently, by construction, to guarantee the first equation in (1.22) it suffices to take

$$r_{k_*} \leq T_0 \lesssim \frac{1}{KR_0^2} \log\left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right).$$

Indeed, recall that, by definition $R(r_{k_*}) \geq \sqrt{2}/2$, and consequently, by (5.4), we have

$$R(t) \geq \frac{\sqrt{2}}{2}\lambda \geq \frac{3}{5}$$

for every t in $[r_{k_*}, \infty)$

Now we proceed to show that we can guarantee the second equation in (1.22) by selecting T_0 within the desired constraints. To achieve this, we argue as in equations (5.33) and (5.34) from the proof of Corollary 5.1, with

$$s = t - \frac{8}{KR_{k_*}} \log \frac{1}{40R_{k_*}},$$

to obtain

$$\rho\big(\mathbb{T}\setminus(L_\gamma^+(t_0)_s)_{\epsilon,t}\big) \leq 4\sqrt{\pi}W^{1/2}\|f_0\|_2 e^{\frac{Q'}{2R_0}\log(1+\frac{1}{R_0})} \times e^{-K\frac{R_{k_*}\sin(\alpha)}{4}(t-r_{k_*}-D_{k_*}-\frac{8}{KR_{k_*}}\log\frac{1}{40R_{k_*}})} \tag{5.36}$$

and

$$\inf_{\theta\in(L_\gamma^+(t_0)_s)_{\epsilon,t}} \cos(\theta-\phi) \geq 1 - \frac{1}{3}R_{k_*}e^{-\log\frac{1}{40R_{k_*}}} - 4\sqrt{\pi}W^{1/2}\|f_0\|_2 e^{\frac{Q'}{2R_0}\log(1+\frac{1}{R_0})} \times e^{-K\frac{R_{k_*}\sin(\alpha)}{4}(t-r_{k_*}-D_{k_*}-\frac{8}{KR_{k_*}}\log\frac{1}{40R_{k_*}})}, \tag{5.37}$$

for any t in $[r_{k_*} + D_{k_*} + \frac{8}{KR_{k_*}}\log\frac{1}{40R_{k_*}}, \infty)$.

Thus, since $R_{k_*} \geq \sqrt{2}/2$, we see that choosing T_0 in such a way that

$$\frac{1}{KR_0^2}\log\left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right) \gtrsim T_0 \geq \frac{4}{KR_{k_*}\sin(\alpha)}\left[\log\frac{4\sqrt{\pi}W^{1/2}\|f_0\|_2}{R_{k_*}/120}\right] + \frac{Q'+16}{2KR_0}\log\left(1+\frac{1}{40R_0}\right) + r_{k_*} + D_{k_*}, \tag{5.38}$$

we can guarantee that condition (5.35) holds for every t in $[T_0,\infty)$. Indeed, by (5.36), such a choice of T_0, together with Lemma 4.2 and Corollary 4.1, implies that

$$\inf_{\theta\in(L_\gamma^+(t_0)_s)_{\epsilon,t}} \cos(\theta-\phi)\geq \frac{59}{60} \tag{5.39}$$

and

$$\rho\big(\mathbb{T}\backslash(L_\gamma^+(t_0)_s)_{\epsilon,t}\big)\leq \frac{1}{120}e^{-K\frac{R_{k_*}\sin(\alpha)}{4}(t-T_0)},$$

for every t in $[T_0,\infty)$. Consequently, the desired result follows from the fact that (5.39) implies that $(L_\gamma^+(t_0)_s)_{\epsilon,t}\subset L_\beta^+(t)$.

6. Wasserstein stability and applications to the particle system

The main objective of this section is to prove Corollary 1.1. Before we proceed with the proof, let us introduce some necessary tools and notation. Throughout this section, we will set a probability density f_0 that belongs to C^1 and will assume that g has compact support in $[-W,W]$. Indeed, we will assume that f_0, K, and W satisfy the hypotheses of Theorem 1.1. Also, we will consider the unique global-in-time classical solution $f=f(t,\theta,\omega)$ to (1.2).

Definition 6.1 (The random empirical measures). By the consistency theorem of Kolmogorov (see [56, Theorem 3.5]), let us consider a probability space $(E,\mathcal{F},\mathbb{P})$ and set some sequence of random variables for $k\in\mathbb{N}$,

$$(\theta_k(0),\omega_k(0))\colon E\to\mathbb{T}\times\mathbb{R},$$

that are i.i.d. with law f_0. For every $N\in\mathbb{N}$, let us consider the random variables

$$t\mapsto(\theta_1^N(t),\omega_1(0)),\ldots,(\theta_N^N(t),\omega_N(0))$$

solving the particle system (1.1) issued at the above random initial data. Then we define the associated random empirical measures as

$$\mu_t^N:=\frac{1}{N}\sum_{i=1}^N\delta_{(\theta_i^N(t),\omega_i(0))}(\theta,\omega) \tag{6.1}$$

for every $t\geq 0$.

The proof of Corollary 1.1 gathers three different tools:

- First, we will use our main Theorem 1.1, which quantifies the rate of convergence of the solution $f=f(t,\theta,\omega)$ towards the global equilibrium f_∞ as $t\to\infty$.
- Second, we require a *concentration inequality* to quantify the law of large numbers. More specifically, we need to quantify the rate of convergence in probability $\mathbb{P}$ of μ_0^N towards f_0 as the number of oscillators N tends to infinity.

- Finally, in order to propagate the above quantification for larger times, we require some *stability estimate for the transportation distance* between μ_t^N and f_t.

Those tools will allow us to quantify a time in which a sufficient number of oscillators of the particle system is concentrated around a neighborhood of the support of the global equilibrium f_∞. This, along with Lemma 4.2 (which also holds for the particle system), will guarantee that the concentration property of oscillators propagates for larger times. Additionally, we will derive the contraction of the diameter of the configuration of oscillators. Before beginning the rigorous proof, let us elaborate on the concentration and stability inequalities.

6.1. Wasserstein concentration inequality

It is apparent from the literature that the above random empirical measures μ_0^N in Definition 6.1 approximate the initial datum f_0 as $N \to \infty$. Specifically, by the strong law of large numbers (see [55]) we obtain

$$\mu_0^N \overset{*}{\rightharpoonup} f_0, \quad \mathbb{P}\text{-a.s},$$

in the narrow topology of $\mathbb{P}(\mathbb{T} \times \mathbb{R})$ as $N \to \infty$. Unfortunately, this is not enough for our purposes as we seek quantitative estimates for the rate of convergence. Such a quantitative control is called *concentration inequality* and there have been many approaches to it in the literature. Most of them require some special structure on the initial datum f_0 and the sequence of random empirical measures μ_0^N; see [6–8]. Specifically, some transportation–entropy inequality is required. To the best of our knowledge, the first result with those assumptions removed was recently introduced in [26]. In our particular setting, it reads as follows.

Lemma 6.1. *Let f_0 contained in $\mathbb{P}(\mathbb{T} \times \mathbb{R})$ be any probability measure with a distribution of natural frequencies $g = (\pi_\omega)_\# f_0$ and assume that*

$$\mathcal{E}(g) := \int_{\mathbb{R}} e^{\omega^4}\, dg < \infty. \tag{6.2}$$

Take any sequence $\{(\theta_k(0), \omega_k(0))\}_{k\in\mathbb{N}}$ of i.i.d. random variables with law f_0 and set the random empirical measures μ_0^N according to Definition 6.1. Then

$$\mathbb{P}\big(W_2(\mu_0^N, f_0) \geq \varepsilon\big) \leq C_1 e^{-C_2 N \varepsilon^4}$$

for every $\varepsilon > 0$ and N in $\mathbb{N}$. Here, C_1 and C_2 are two positive constants that do not depend on either ε or on N, but only depend on $\mathcal{E}(g)$.

Proof. Take $d = 2$, $p = 2$, $\gamma = 1$, and $\beta = 4$ in [26, Theorem 2]. ■

In the above result, we used the classical quadratic Wasserstein distance W_2, namely,

$$W_2(\mu_0^N, f_0) = \left(\inf_{\gamma\in\Pi(\mu_0^N, f_0)} \int_{\mathbb{T}^2\times\mathbb{R}^2} (d(\theta,\theta')^2 + (\omega-\omega')^2)\, d\gamma\right)^{1/2}.$$

However, as discussed in Remark 3.2, such a distance is not appropriate for this problem due to the fact that the standard quadratic distance on the product Riemannian manifold $\mathbb{T} \times \mathbb{R}$ provides a cost functional which is not dimensionally correct. Indeed, we corrected this situation by scaling ω. Let us recall the scaled quadratic Wasserstein distance (see Definition 3.2)

$$\mathrm{SW}_2(\mu_0^N, f_0) = \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{T}^2 \times \mathbb{R}^2} \left(d(\theta, \theta')^2 + \frac{(\omega - \omega')^2}{K^2} \right) d\gamma \right)^{1/2}.$$

Let us note that by scaling, we can adapt the above Lemma 6.1 to the right transportation distance SW_2. Specifically, let us consider the dilation with respect to ω,

$$\mathcal{D}_K(\omega) := \frac{\omega}{K} \quad \text{for } \omega \in \mathbb{R}.$$

Then we can define the following scaled objects:

$$f_{0,K} := (\mathrm{Id} \otimes \mathcal{D}_K)_\# f_0 \quad \text{and} \quad \mu_{0,K}^N := (\mathrm{Id} \otimes \mathcal{D}_K)_\# \mu_0^N.$$

Notice that $f_{0,K}$ is contained in $\mathbb{P}(\mathbb{T} \times \mathbb{R})$ and the empirical measures $\mu_{0,K}^N$ are i.i.d. variables with law $f_{0,K}$. Interestingly, we obtain the relation

$$\mathrm{SW}_2(\mu_0^N, f_0) = W_2(\mu_{0,K}^N, f_{0,K}).$$

Then, applying Lemma 6.1 to the scaled objects, we obtain the following result.

Lemma 6.2. *Let f_0 be a probability density in $C^1(\mathbb{T} \times \mathbb{R})$, and assume that the distribution of natural frequencies $g = (\pi_\omega)_\# f_0$ has compact support in $[-W, W]$ and that condition (1.20) in Theorem 1.1 holds true. Take any sequence $\{(\theta_k(0), \omega_k(0))\}_{k \in \mathbb{N}}$ of i.i.d. random variables with law f_0 and set the random empirical measures μ_0^N according to Definition 6.1. Then*

$$\mathbb{P}\big(\mathrm{SW}_2(\mu_0^N, f_0) \geq \varepsilon\big) \leq C_1 \exp(-C_2 N \varepsilon^4) \tag{6.3}$$

for every $\varepsilon > 0$ and N in $\mathbb{N}$. Here, C_1 and C_2 are two positive universal constants.

Remark 6.1. Notice that, according to Lemma 6.1, the above C_1 and C_2 only depend upon $\mathcal{E}(g_K)$ where $g_K := \mathcal{D}_{K\#} g$. Since g has compact support in $[-W, W]$ we obtain

$$1 \leq \mathcal{E}(g_K) \leq e^{\frac{W^4}{K^4}},$$

so that C_1 and C_2 will ultimately depend only on $\frac{W}{K}$. However, under assumptions (1.20) in Theorem 1.1, $\frac{W}{K}$ is smaller than a universal constant. Consequently, $\mathcal{E}(g_K)$ can be made smaller than a universal constant arbitrarily close to 1. This justifies considering C_1 and C_2 to be universal constants.

6.2. Wasserstein stability estimate

The study of *Wasserstein stability estimates* or *Dobrušin-type estimates* for measure-valued solutions to kinetic equations is a classical topic. Depending on the degree of regularity of the interaction kernel, an appropriate transportation distance has to be considered. In particular, the starting works by Dobrušin and Neunzert (see [24, 44]) show that the bounded-Lipschitz distance is appropriate for Lipschitz-continuous interaction kernels. This type of inequality has been generalized to some specific kernels with more limited regularity. In particular, the right transportation distance for gradient flows associated with $-\lambda$-convex is the quadratic Wasserstein distance W_2 (see [14]). Indeed, we do not necessarily need an underlying gradient structure, but only require that the interaction kernel is one-sided Lipschitz-continuous. This was proven in [49, Theorem 4.7] for the Kuramoto model with weakly singular weights, which in our case provides the following stability estimate for W_2:

$$W_2(f_t, \bar{f}_t) \leq e^{(2K+\frac{1}{2})t} W_2(f_0, \bar{f}_0), \tag{6.4}$$

which holds for any two measured-valued solutions to (1.2). Notice that units are not correct in the above inequality, and this is again due to the fact that W_2 is not dimensionally correct in this problem (recall 3.2). Instead, we can replace W_2 with SW_2 (see Definition 3.2) to recover the following result.

Lemma 6.3. *Consider $K > 0$ and let f and $\bar{f}$ be weak measured-valued solutions to* (1.2) *with initial datum f_0 and $\bar{f}_0 \in \mathbb{P}_2(\mathbb{T}\times\mathbb{R})$. Then we have*

$$\mathrm{SW}_2(f_t, \bar{f}_t) \leq e^{\frac{5}{2}Kt}\, \mathrm{SW}_2(f_0, \bar{f}_0)$$

for every $t \geq 0$.

Proof. Consider an optimal transference plan γ_0 joining f_0 to $\bar{f}_0$, that is,

$$\gamma_0 \in \Pi(f_0, \bar{f}_0) := \left\{\gamma \in \mathbb{P}((\mathbb{T}\times\mathbb{R})\times(\mathbb{T}\times\mathbb{R})) : (\pi_1)_{\#}\gamma = f_0 \text{ and } (\pi_2)_{\#}\gamma = \bar{f}_0\right\},$$

such that

$$\mathrm{SW}_2(f_0, \bar{f}_0)^2 = \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} d_K((\theta_1,\omega_1),(\theta_2,\omega_2))^2\, d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0.$$

Here π_1 and π_2 represent the projections

$$\begin{aligned}\pi_1((\theta,\omega),(\theta',\omega')) &= (\theta,\omega),\\ \pi_2((\theta,\omega),(\theta',\omega')) &= (\theta',\omega').\end{aligned}$$

Let us consider the following competitor at time t via push-forward, namely,

$$\gamma_t := (\mathbb{X}_{0,t}\otimes\overline{\mathbb{X}}_{0,t})_{\#}\gamma_0 \in \mathbb{P}((\mathbb{T}\times\mathbb{R})\times(\mathbb{T}\times\mathbb{R})),$$

where $\mathbb{X}_{0,t}(\theta,\omega) = (\Theta_{0,t}(\theta,\omega),\omega)$ and $\overline{\mathbb{X}}_{0,t}(\theta,\omega) = (\overline{\Theta}_{0,t}(\theta,\omega),\omega)$ are the characteristic flows associated with the transport fields $v[f]$. Since $\gamma_t \in \Pi(f_t, \bar{f}_t)$,

$$\begin{aligned}\frac{1}{2}\,\mathrm{SW}_2(f_t, \bar{f}_t)^2 &\leq \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} \frac{1}{2} d_K((\theta_1,\omega_1),(\theta_2,\omega_2))^2\, d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_t \\ &= \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}} \frac{1}{2} d_K(\mathbb{X}_{0,t}(\theta_1,\omega_1),\overline{\mathbb{X}}_{0,t}(\theta_2,\omega_2))^2\, d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0 \\ &=: I(t).\end{aligned}$$

Our final goal is to derive some Grönwall-type inequality for I. Fix $(\theta_1,\omega_1),(\theta_2,\omega_2) \in \mathbb{T}\times\mathbb{R}$ and define the following curves in $\mathbb{T}$:

$$\Theta(t) := \Theta_{0,t}(\theta_1,\omega_1) \quad \text{and} \quad \overline{\Theta}(t) := \overline{\Theta}_{0,t}(\theta_2,\omega_2),$$

and the associated characteristic curves in $\mathbb{T}\times\mathbb{R}$,

$$\begin{aligned}\mathbb{X}(t) &:= \mathbb{X}_{0,t}(\theta_1,\omega_1) = (\Theta(t),\omega_1),\\ \overline{\mathbb{X}}(t) &:= \overline{\mathbb{X}}_{0,t}(\theta_2,\omega_2) = (\overline{\Theta}(t),\omega_2).\end{aligned}$$

Set a minimizing geodesic $x_t\colon [0,1] \to \mathbb{T}\times\mathbb{R}$ joining $\mathbb{X}(t)$ to $\overline{\mathbb{X}}(t)$ for every fixed $t > 0$. Notice that the function

$$t \mapsto \frac{1}{2} d_K^2(\mathbb{X}(t),\overline{\mathbb{X}}(t))$$

is Lipschitz continuous. Then we can take derivatives and show that

$$\frac{d}{dt}\frac{1}{2} d_K^2(\mathbb{X}(t),\overline{\mathbb{X}}(t)) \leq -\big\langle \big(v[f_t](\mathbb{X}(t)),0\big), x_t'(0)\big\rangle - \big\langle \big(v[\bar{f}_t](\overline{\mathbb{X}}(t)),0\big), -x_t'(1)\big\rangle \tag{6.5}$$

for almost every $t \geq 0$. Let us now consider $\theta(t) := \overline{\overline{\Theta}(t) - \Theta(t)}$, the representative of $\overline{\Theta}(t) - \Theta(t)$ modulo 2π that lies in $(-\pi,\pi]$. We find two different cases:

Case 1*:* $\theta(t) \in (-\pi,\pi)$. In this case, the only minimizing geodesic reads

$$x_t(s) = (e^{i(\Theta(t)+s\theta(t))}, \omega_1 + s(\omega_2-\omega_1)), \quad s \in [0,1].$$

Then (6.5) reads

$$\frac{d}{dt}\frac{1}{2} d_K^2(\mathbb{X}(t),\overline{\mathbb{X}}(t)) \leq \big(v[\bar{f}_t](\overline{\Theta}(t),\omega_2) - v[f_t](\Theta(t),\omega_1)\big)\theta(t)$$

for almost every $t \geq 0$.

Case 2*:* $\theta(t) = \pi$. In this second case there are exactly two minimizing geodesics

$$x_{t,\pm}(s) = (e^{i(\Theta(t)\pm\pi s)}, \omega_1 + s(\omega_2-\omega_1)), \quad s \in [0,1].$$

Then we restate (6.5) as

$$\frac{d}{dt}\frac{1}{2} d_K^2(\mathbb{X}(t),\overline{\mathbb{X}}(t)) \leq \big(v[\bar{f}_t](\overline{\Theta}(t),\omega_2) - v[f_t](\Theta(t),\omega_1)\big)(\pm\pi)$$

for almost every $t \geq 0$. To sum up, we achieve the following estimate:

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}d_K^2(\mathbb{X}_{0,t}(\theta_1,\omega_1),\bar{\mathbb{X}}_{0,t}(\theta_2,\omega_2))&\\ \leq \big(v[f_t](\Theta_{0,t}(\theta_1,\omega_1),\omega_1)-v[\bar{f}_t](\bar{\Theta}_{0,t}(\theta_2,\omega_2),\omega_2)\big)&\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)}\end{aligned}$$

for every $\theta_1,\theta_2\in\mathbb{T}$, each $\omega_1,\omega_2\in\mathbb{R}$, and almost every $t\geq 0$. Using the dominated convergence theorem, we show that I is absolutely continuous and taking derivatives under the integral sign implies

$$\begin{aligned}\frac{dI}{dt}\leq \int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}}&\big(v[f_t](\Theta_{0,t}(\theta_1,\omega_1),\omega_1)-v[\bar{f}_t](\bar{\Theta}_{0,t}(\theta_2,\omega_2),\omega_2)\big)\\ &\times\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_1,\omega_2)}\,d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0\end{aligned}\tag{6.6}$$

for almost every $t\geq 0$. Also, note that

$$\begin{aligned}v[f_t](\theta,\omega)&=\omega-K\int_{\mathbb{T}\times\mathbb{R}}\sin(\theta-\Theta_{0,t}(\theta_1',\omega_1'))\,d_{(\theta_1',\omega_1')}f_0,\\ v[\bar{f}_t](\theta,\omega)&=\omega-K\int_{\mathbb{T}\times\mathbb{R}}\sin(\theta-\bar{\Theta}_{0,t}(\theta_2',\omega_2'))\,d_{(\theta_2',\omega_2')}\bar{f}_0.\end{aligned}$$

Since $(\pi_1)_\#\gamma_0=f_0$ and $(\pi_2)_\#\gamma_0=\bar{f}_0$,

$$v[f_t](\theta,\omega)=\omega-K\int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}}\sin(\theta-\Theta_{0,t}(\theta_1',\omega_1'))\,d_{((\theta_1',\omega_1'),(\theta_2',\omega_2'))}\gamma_0,\tag{6.7}$$

$$v[\bar{f}_t](\theta,\omega)=\omega-K\int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}}\sin(\theta-\bar{\Theta}_{0,t}(\theta_2',\omega_2'))\,d_{((\theta_1',\omega_1'),(\theta_2',\omega_2'))}\gamma_0.\tag{6.8}$$

Putting (6.7)–(6.8) into (6.6) amounts to

$$\begin{aligned}\frac{dI}{dt}\leq&\int_{(\mathbb{T}\times\mathbb{R})^4}(\omega_1-\omega_2)\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)}\\ &\qquad\times d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0\,d_{((\theta_1',\omega_1'),(\theta_2',\omega_2'))}\gamma_0\\ &-K\int_{(\mathbb{T}\times\mathbb{R})^4}\big(\sin(\Theta_{0,t}(\theta_1,\omega_1)-\Theta_{0,t}(\theta_1',\omega_1'))-\sin(\bar{\Theta}_{0,t}(\theta_2,\omega_2)-\bar{\Theta}_{0,t}(\theta_2',\omega_2'))\big)\\ &\qquad\times\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)}\,d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0\,d_{((\theta_1',\omega_1'),(\theta_2',\omega_2'))}\gamma_0\end{aligned}\tag{6.9}$$

for almost every $t\geq 0$. By Young's inequality, it is clear that

$$\begin{aligned}(\omega_1-\omega_2)\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)}&\leq\frac{K}{2}\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)}^2+\frac{(\omega_1-\omega_2)^2}{2K}\\ &=\frac{K}{2}d_K(\mathbb{X}_{0,t}(\theta_1,\omega_1),\bar{\mathbb{X}}_{0,t}(\theta_2,\omega_2))^2.\end{aligned}$$

This, along with a clear symmetrization argument in the second term, implies

$$\begin{aligned}\frac{dI}{dt} &\leq KI(t)\\ &\quad - \frac{K}{2}\int_{(\mathbb{T}\times\mathbb{R})^4} (\sin(\Theta_{0,t}(\theta_1,\omega_1)-\Theta_{0,t}(\theta_1',\omega_1')) - \sin(\bar{\Theta}_{0,t}(\theta_2,\omega_2)-\bar{\Theta}_{0,t}(\theta_2',\omega_2')))\\ &\qquad \times \left(\overline{\Theta_{0,t}(\theta_1,\omega_1)-\bar{\Theta}_{0,t}(\theta_2,\omega_2)} - \overline{\Theta_{0,t}(\theta_1',\omega_1')-\bar{\Theta}_{0,t}(\theta_2',\omega_2')}\right)\\ &\qquad \times\, d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0\, d_{((\theta_1',\omega_1'),(\theta_2',\omega_2'))}\gamma_0\end{aligned}$$

for almost every $t\geq 0$. Now, using the Lipschitz property of the sine function we achieve the inequality

$$\frac{dI}{dt}\leq (K+4K)I \qquad \text{for a.e. } t\geq 0.$$

Integrating the inequality and using that

$$I(0)=\int_{\mathbb{T}\times\mathbb{R}}\int_{\mathbb{T}\times\mathbb{R}}\frac{1}{2}d_K((\theta_1,\omega_1),(\theta_2,\omega_2))^2\, d_{((\theta_1,\omega_1),(\theta_2,\omega_2))}\gamma_0=\frac{1}{2}\,\mathrm{SW}_2(f_0,\bar{f}_0)^2,$$

yields the desired result. ∎

6.3. Probability of mass concentration and diameter contraction

Now we are ready to begin the proof of Corollary 1.1. Let L and $L_{1/2}$ be intervals of diameter $2/5$ and $1/5$ centered around the order parameter ϕ_∞ of f_∞. Recall that by formula (1.22) we obtain

$$R_\infty=\lim_{t\to\infty}R(t)\geq 3/5.$$

Looking at the structure of the stable equilibria f_∞ in (1.18) (which corresponds to $g^-=0$, that is, no antipodal mass), we observe that for any (θ,ω) in $\operatorname{supp} f_\infty$ we have the relation

$$\theta=\phi_\infty+\arcsin\Big(\frac{\omega}{KR_\infty}\Big).$$

In particular,

$$|\theta-\phi_\infty|\leq \arcsin\Big(\frac{W}{KR_\infty}\Big)\leq \arcsin\Big(\frac{5}{3}\frac{W}{K}\Big).$$

Then we can select C in (1.20), so that we have

$$\operatorname{supp} f_\infty\subset L_{\frac{1}{2}}\times[-W,W]. \tag{6.10}$$

Notice that the choice of the diameter of L is somehow arbitrary and is subordinated to the size of the universal constant C in Theorem 1.1 (the smaller C, the smaller the diameter of L). For simplicity, we have set it to $2/5$ but it can be generalized to sharper values. We divide the proof into the following steps:

Step a. We control the mass of μ_t^N and f_t in $\mathbb{T}\setminus L$, namely,

$$\mu_t^N((\mathbb{T}\setminus L)\times\mathbb{R}) \leq 25\,\mathrm{SW}_2(\mu_t^N, f_\infty)^2, \tag{6.11}$$

$$\rho_t(\mathbb{T}\setminus L) \leq 25\,\mathrm{SW}_2(f_t, f_\infty)^2, \tag{6.12}$$

for any $t > 0$.

Fix $t > 0$ and let $\gamma_t \in \mathbb{P}((\mathbb{T}\times\mathbb{R})\times(\mathbb{T}\times\mathbb{R}))$ be an optimal transport plan between μ_t^N and f_∞ for the scaled Wasserstein distance SW_2. Then we have

$$\begin{aligned}
&\mathrm{SW}_2(\mu_t^N, f_\infty)^2\\
&\quad= \int_{(\mathbb{T}\times\mathbb{R})^2} d_K((\theta,\omega),(\theta',\omega'))^2\,d\gamma_t\\
&\quad\geq \int_{((\mathbb{T}\setminus L)\times\mathbb{R})\times(L_{1/2}\times\mathbb{R})} d(\theta,\theta')^2\,d\gamma_t\\
&\quad\geq \frac{1}{25}\gamma_t\big(((\mathbb{T}\setminus L)\times\mathbb{R})\times(L_{1/2}\times\mathbb{R})\big)\\
&\quad= \frac{1}{25}\big[\gamma_t\big(((\mathbb{T}\setminus L)\times\mathbb{R})\times(\mathbb{T}\times\mathbb{R})\big) - \gamma_t\big(((\mathbb{T}\setminus L)\times\mathbb{R})\times((\mathbb{T}\setminus L_{1/2})\times\mathbb{R})\big)\big]\\
&\quad\geq \frac{1}{25}\big[\gamma_t\big(((\mathbb{T}\setminus L)\times\mathbb{R})\times(\mathbb{T}\times\mathbb{R})\big) - \gamma_t\big((\mathbb{T}\times\mathbb{R})\times((\mathbb{T}\setminus L_{1/2})\times\mathbb{R})\big)\big]\\
&\quad= \frac{1}{25}\big[\mu_t^N((\mathbb{T}\setminus L)\times\mathbb{R}) - f_\infty((\mathbb{T}\setminus L_{1/2})\times\mathbb{R})\big].
\end{aligned}$$

Thus, using the inclusion (6.10), we observe that the second term in the last line of the above inequality vanishes and we obtain (6.11). Similarly, using the above argument with μ_t^N replaced by f_t, we deduce (6.12).

Step b. We claim that we can select T_0 satisfying

$$T_0 \lesssim \frac{1}{KR_0^2}\log\Big(1 + W^{1/2}\|f_0\|_2 + \frac{1}{R_0}\Big), \tag{6.13}$$

and with the additional property that

$$\mathrm{SW}_2(f_t, f_\infty) \leq \frac{1}{\sqrt{500}}e^{-\frac{1}{40}K(t-T_0)} \tag{6.14}$$

for every t in $[T_0,\infty)$.

To show this, take Q_1 large enough and T_0 verifying

$$T_0 \leq \frac{Q_1}{KR_0^2}\log\Big(1 + W^{1/2}\|f_0\|_2 + \frac{1}{R_0}\Big),$$

so that we meet the constraints in Theorem 1.1. Then, using (3.24) and Proposition 3.2 we obtain

$$\mathrm{SW}_2(f_t, f_\infty) \leq Q_2 e^{-\frac{1}{40}K(t-T_0)} \tag{6.15}$$

for all t in $[T_0,\infty)$ and some universal constant Q_2. Notice that by taking Q_1 large enough, we can make Q_2 arbitrarily small (e.g., $Q_2 = \frac{1}{\sqrt{500}}$). This concludes the proof of the claim.

Step c. We compute N in $\mathbb{N}$ and $d_N > 0$ for each $N \geq N^*$ so that

$$\mathbb{P}\Big(\mathrm{SW}_2(\mu_t^N, f_t) \leq \frac{1}{\sqrt{500}} e^{-\frac{1}{40}K(t-T_0)}\Big) \geq 1 - C_1 e^{-C_2 N^{\frac{1}{2}}} \tag{6.16}$$

for any t in $[T_0, T_0 + d_N]$ and any $N \geq N^*$.

First, for each N in $\mathbb{N}$ let us set the scale

$$\varepsilon_N := N^{-\frac{1}{8}}. \tag{6.17}$$

Now we define N^* as

$$N^* := \min\big\{N \in \mathbb{N} : \varepsilon_N e^{\frac{5K}{2} T_0} \leq \tfrac{1}{\sqrt{500}}\big\}, \tag{6.18}$$

so that, by definition, we get the bound

$$N^* \geq 500^4 e^{20 K T_0}.$$

Fix any $N \geq N^*$. Notice that N^* has been defined in (6.18) so that there exists $d_N > 0$ with the property

$$\varepsilon_N e^{\frac{5K}{2}(T_0 + d_N)} = \frac{1}{\sqrt{500}} e^{-\frac{1}{40} K d_N}. \tag{6.19}$$

Indeed, by dividing (6.19) over (6.18), we can quantify d_N in terms of N^* as

$$\frac{\varepsilon_N}{\varepsilon_{N^*}} e^{\frac{5K}{2} d_N} \geq e^{-\frac{1}{40} K d_N}.$$

Consequently, we have

$$d_N \geq \frac{5}{101 K} \log \frac{N}{N^*}.$$

By construction, letting $\varepsilon = \varepsilon_N$ in the concentration inequality (6.3) of Lemma 6.2, we obtain the following quantification:

$$\mathbb{P}\big(\mathrm{SW}_2(\mu_0^N, f_0) \geq \varepsilon_N\big) \leq C_1 e^{-C_2 N^{\frac{1}{2}}} \tag{6.20}$$

for every $N \in \mathbb{N}$. Thus, by monotonicity of the exponential function, we conclude that for any $t \in [T_0, T_0 + d_N]$ we have

$$\begin{aligned} C_1 e^{-C_2 N^{\frac{1}{2}}} &\geq \mathbb{P}\big(\mathrm{SW}_2(\mu_0^N, f_0) \geq \varepsilon_N\big) \\ &\geq \mathbb{P}\big(\mathrm{SW}_2(\mu_t^N, f_t) \geq \varepsilon_N e^{\frac{5K}{2} t}\big) \\ &\geq \mathbb{P}\big(\mathrm{SW}_2(\mu_t^N, f_t) \geq \varepsilon_N e^{\frac{5K}{2}(T_0 + d_N)}\big) \\ &= \mathbb{P}\Big(\mathrm{SW}_2(\mu_t^N, f_t) \geq \frac{1}{\sqrt{500}} e^{-\frac{1}{40} K d_N}\Big) \\ &\geq \mathbb{P}\Big(\mathrm{SW}_2(\mu_t^N, f_t) \geq \frac{1}{\sqrt{500}} e^{-\frac{1}{40} K (t - T_0)}\Big), \end{aligned}$$

where in the first inequality we have used the concentration inequality (6.20), in the second one we have used the stability estimate in Lemma 6.3, and the remainder follow from our choice of d_N in (6.19) and t in $[T_0, T_0 + d_N]$. That ends the proof of (6.16).

Step d. We quantify the probability of mass concentration of μ_t^N in the interval L, namely,

$$\mathbb{P}\left(\mu_t^N(L\times\mathbb{R})\geq 1-\frac{1}{5}e^{-\frac{1}{20}K(t-T_0)}\right)\geq 1-C_1e^{-C_2N^{\frac{1}{2}}} \tag{6.21}$$

for every t in $[T_0, T_0+d_N)$ and any $N\geq N^*$.

Now, by (6.11), (6.14), and the triangular inequality we have

$$\begin{aligned}\mu_t^N((\mathbb{T}\backslash L)\times\mathbb{R}) &\leq 25\,\mathrm{SW}_2(\mu_t^N, f_\infty)^2\\ &\leq 50\big[\mathrm{SW}_2(\mu_t^N, f_t)^2+\mathrm{SW}_2(f_t, f_\infty)^2\big]\\ &\leq 50\left[\mathrm{SW}_2(\mu_t^N, f_t)^2+\frac{1}{500}e^{-\frac{1}{20}K(t-T_0)}\right]\end{aligned}$$

for every t in $[T_0, T_0+d_N)$. Hence, we obtain

$$\mu_t^N(L\times\mathbb{R})\geq 1-\frac{1}{10}e^{-\frac{3}{10}K(t-T_0)}-50\,\mathrm{SW}_2(\mu_t^N, f_t)^2$$

for each t in $[T_0, T_0+d_N]$. This, along with (6.16), concludes the proof of (6.21).

Step e. We quantify the probability of mass concentration and diameter contraction along the time interval $[s,\infty)$ for any s in $[T_0, T_0+d_N]$.

We are now ready to finish the proof of Corollary 1.1. Let us consider $N\geq N^*$, s in $[T_0, T_0+d_N]$, and any realization of the random empirical measure μ^N (recall Definition 6.1) so that the condition within (6.21) holds. Hence, by construction, we obtain that at such a realization,

$$p:=\inf_{\theta,\theta'\in L}\cos(\theta-\theta')\geq\frac{4}{5}\quad\text{and}\quad m:=\mu_s^N(L\times\mathbb{R})\geq 1-\frac{1}{5}e^{-\frac{1}{20}K(s-T_0)}\geq\frac{4}{5}.$$

Then we obtain the relation

$$mp-(1-m)=\frac{4}{5}\cdot\frac{4}{5}-\left(1-\frac{4}{5}\right)=\frac{11}{25}.$$

In particular, take $\sigma:=2/5$ and notice that the above relations along with assumption (1.20) in Theorem 1.1 guarantee condition (4.5) within the hypotheses of Lemma 4.2. Notice that the result also holds true for the particle system. Consequently, it asserts that for such a realization of μ^N we can consider a time-dependent interval $L_s^N(t)$ with $t\geq s$ so that $L_s^N(s)=L$ and

$$\begin{aligned}\mu_t^N(L_s^N(t)\times\mathbb{R})&\geq 1-\frac{1}{5}e^{-\frac{1}{20}K(s-T_0)},\\ 1-\inf_{\theta,\theta'\in L_s^N(t)}\cos(\theta-\theta')&\leq\max\left\{\frac{1}{5}e^{-\frac{K}{10}(t-s)}, 25\frac{W^2}{K^2}\right\},\end{aligned} \tag{6.22}$$

for any $t\geq s$. Indeed, we have $L_s^N(t)=\pi_\theta(\mathbb{X}_{s,t}^N(L\times[-W,W]))$, where $\mathbb{X}_{s,t}^N$ represents the flow of the particle system, that is, the flow of $v[\mu^N]$. Our final goal is to simplify the last condition in (6.22). To this end, let us consider $D_s^N(t) := \operatorname{diam}(L_s^N(t))$ and notice that the inequality implies

$$2\frac{(D_s^N(t))^2}{5} \leq 1 - \cos(D_s^N(t)) \leq \max\left\{\frac{1}{5}e^{-\frac{K}{10}(t-s)}, 25\frac{W^2}{K^2}\right\} \tag{6.23}$$

for any $t \geq s$. In particular, we obtain (D). Thus, Corollary 1.1 follows.

Acknowledgments. Part of this research was done while the second author was visiting CSCAMM at the University of Maryland, USA. We are grateful for the hospitality of CSCAMM, Prof. P.-E. Jabin, and Prof. E. Tadmor. We also thank the anonymous referee for useful comments and for pointing us to reference [34].

Funding. The work of J. Morales was supported by the NSF grants DMS16-13911, RNMS11-07444 (KI-Net) and ONR grant N00014-1812465. The work of D. Poyato was partially supported by the MECD (Spain) research grant FPU14/06304, the MINECO-Feder (Spain) research grant number RTI2018-098850-B-I00, the Junta de Andalucia (Spain) projects PY18-RT-2422 and A-FQM-311-UGR18, and the European Research Council (ERC) grant agreement no. 639638.

References

[1] J. A. Acebrón, L. L. Bonilla, C. J. P. Pérez-Vicente, F. Ritort, and R. Spigler, The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* **77** (2005), no. 1, 137–185

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*. 2nd edn, Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel, 2008 Zbl 1145.35001 MR 2401600

[3] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Synchronization in complex networks. *Phys. Rep.* **469** (2008), no. 3, 93–153 MR 2477097

[4] J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84** (2000), no. 3, 375–393 Zbl 0968.76069 MR 1738163

[5] D. Benedetto, E. Caglioti, and U. Montemagno, On the complete phase synchronization for the Kuramoto model in the mean-field limit. *Commun. Math. Sci.* **13** (2015), no. 7, 1775–1786 Zbl 1345.37065 MR 3393174

[6] E. Boissard, *Problèmes d'interaction discret-continu et distances de Wasserstein*. Ph.D. thesis, Université de Toulouse III, 2011

[7] E. Boissard, Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. Probab.* **16** (2011), no. 83, 2296–2333 Zbl 1254.60014 MR 2861675

[8] F. Bolley, A. Guillin, and C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields* **137** (2007), no. 3-4, 541–593 Zbl 1113.60093 MR 2280433

[9] M. Breakspear, S. Heitmann, and A. Dafferstshofer, Generative models of cortical oscillations: Neurobiological implications of the Kuramoto model. *Front. Hum. Neurosci.* **4** (2010), 190

[10] M. J. Cáceres, J. A. Carrillo, and T. Goudon, Equilibration rate for the linear inhomogeneous relaxation-time Boltzmann equation for charged particles. *Comm. Partial Differential Equations* **28** (2003), no. 5-6, 969–989 Zbl 1045.35094 MR 1986057

[11] E. A. Carlen and M. C. Carvalho, Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. *J. Statist. Phys.* **67** (1992), no. 3-4, 575–608 Zbl 0899.76317 MR 1171145

[12] E. A. Carlen and M. C. Carvalho, Entropy production estimates for Boltzmann equations with physically realistic collision kernels. *J. Statist. Phys.* **74** (1994), no. 3-4, 743–782 Zbl 0831.76074 MR 1263387

[13] J. A. Carrillo, Y.-P. Choi, S.-Y. Ha, M.-J. Kang, and Y. Kim, Contractivity of transport distances for the kinetic Kuramoto equation. *J. Stat. Phys.* **156** (2014), no. 2, 395–415 Zbl 1303.82016 MR 3215629

[14] J. A. Carrillo, F. James, F. Lagoutière, and N. Vauchelet, The Filippov characteristic flow for the aggregation equation with mildly singular potentials. *J. Differential Equations* **260** (2016), no. 1, 304–338 Zbl 1323.35005 MR 3411674

[15] H. Chiba, A proof of the Kuramoto conjecture for a bifurcation structure of the infinite-dimensional Kuramoto model. *Ergodic Theory Dynam. Systems* **35** (2015), no. 3, 762–834 Zbl 1322.37036 MR 3334903

[16] Y.-P. Choi, S.-Y. Ha, S. Jung, and Y. Kim, Asymptotic formation and orbital stability of phase-locked states for the Kuramoto model. *Phys. D* **241** (2012), no. 7, 735–754 Zbl 1238.34102 MR 2897541

[17] L. Desvillettes and K. Fellner, Exponential decay toward equilibrium via entropy methods for reaction-diffusion equations. *J. Math. Anal. Appl.* **319** (2006), no. 1, 157–176 Zbl 1096.35018 MR 2217853

[18] L. Desvillettes and C. Villani, On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: The linear Fokker-Planck equation. *Comm. Pure Appl. Math.* **54** (2001), no. 1, 1–42 Zbl 1029.82032 MR 1787105

[19] L. Desvillettes and C. Villani, On the trend to global equilibrium for spatially inhomogeneous kinetic systems: The Boltzmann equation. *Invent. Math.* **159** (2005), no. 2, 245–316 Zbl 1162.82316 MR 2116276

[20] H. Dietert, Stability and bifurcation for the Kuramoto model. *J. Math. Pures Appl. (9)* **105** (2016), no. 4, 451–489 Zbl 1339.35318 MR 3471147

[21] H. Dietert, Stability of partially locked states in the Kuramoto model through Landau damping with Sobolev regularity. 2017, arXiv:1707.03475

[22] H. Dietert and B. Fernandez, The mathematics of asymptotic stability in the Kuramoto model. *Proc. A.* **474** (2018), no. 2220, 20180467 Zbl 1425.35206 MR 3910336

[23] H. Dietert, B. Fernandez, and D. Gérard-Varet, Landau damping to partially locked states in the Kuramoto model. *Comm. Pure Appl. Math.* **71** (2018), no. 5, 953–993 Zbl 1392.34034 MR 3794519

[24] R. L. Dobrušin, Vlasov equations. *Funct. Anal. Appl.* **13** (1979), no. 2, 115–123 Zbl 0422.35068 MR 541637

[25] K. Fellner, L. Neumann, and C. Schmeiser, Convergence to global equilibrium for spatially inhomogeneous kinetic models of non-micro-reversible processes. *Monatsh. Math.* **141** (2004), no. 4, 289–299 Zbl 1112.82038 MR 2053654

[26] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (2015), no. 3-4, 707–738 Zbl 1325.60042 MR 3383341

[27] S.-Y. Ha, T. Ha, and J.-H. Kim, On the complete synchronization of the Kuramoto phase model. *Phys. D* **239** (2010), no. 17, 1692–1700 Zbl 1213.34068 MR 2684624

[28] S.-Y. Ha, H. K. Kim, and S. W. Ryoo, Emergence of phase-locked states for the Kuramoto model in a large coupling regime. *Commun. Math. Sci.* **14** (2016), no. 4, 1073–1091 Zbl 1383.92016 MR 3491817

[29] S.-Y. Ha, H. K. Kim, and S. W. Ryoo, On the finiteness of collisions and phase-locked states for the Kuramoto model. *J. Stat. Phys.* **163** (2016), no. 6, 1394–1424 Zbl 1345.70019 MR 3503269

[30] S.-Y. Ha, J. Kim, J. Park, and X. Zhang, Uniform stability and mean-field limit for the augmented Kuramoto model. *Netw. Heterog. Media* **13** (2018), no. 2, 297–322 Zbl 1405.70007 MR 3811564

[31] S.-Y. Ha, Y.-H. Kim, J. Morales, and J. Park, Emergence of phase concentration for the Kuramoto-Sakaguchi equation. *Phys. D* **401** (2020), 132154 Zbl 1453.34075 MR 4034671

[32] S.-Y. Ha, D. Ko, J. Park, and X. Zhang, Collective synchronization of classical and quantum oscillators. *EMS Surv. Math. Sci.* **3** (2016), no. 2, 209–267 Zbl 1403.34032 MR 3576533

[33] S.-Y. Ha, Z. Li, and X. Xue, Formation of phase-locked states in a population of locally interacting Kuramoto oscillators. *J. Differential Equations* **255** (2013), no. 10, 3053–3070 Zbl 1325.34066 MR 3093356

[34] S.-Y. Ha and S. W. Ryoo, Asymptotic phase-locking dynamics and critical coupling strength for the Kuramoto model. *Comm. Math. Phys.* **377** (2020), no. 2, 811–857 Zbl 1447.34040 MR 4115007

[35] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* **29** (1998), no. 1, 1–17 Zbl 0915.35120 MR 1617171

[36] Y. Kuramoto, Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics (Kyoto Univ., Kyoto, 1975)*, pp. 420–422. Lecture Notes in Phys. 39, Springer, Berlin, 1975 Zbl 0335.34021 MR 0676492

[37] Y. Kuramoto, *Chemical oscillations, waves, and turbulence*. Springer Series in Synergetics 19, Springer, Berlin, 1984 Zbl 0558.76051 MR 762432

[38] C. Lancellotti, On the Vlasov limit for systems of nonlinearly coupled oscillators without noise. *Transport Theory Statist. Phys.* **34** (2005), no. 7, 523–535 Zbl 1206.82107 MR 2265477

[39] Z. Li, X. Xue, and D. Yu, On the Łojasiewicz exponent of Kuramoto model. *J. Math. Phys.* **56** (2015), no. 2, 022704 MR 3390886

[40] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels. In *Les équations aux Dérivées Partielles (Paris, 1962)*, pp. 87–89, Éditions du Centre National de la Recherche Scientifique (CNRS), Paris, 1963 MR 0160856

[41] S. C. Manrubia, A. S. Mikhailov, and D. H. Zanette, *Emergence of Dynamical Order. Synchronization Phenomena in Complex Systems*. World Sci. Lect. Notes Complex Syst. 2, World Scientific, Hackensack, NJ, 2004 Zbl 1119.34001

[42] R. Mirollo and S. H. Strogatz, The spectrum of the partially locked state for the Kuramoto model. *J. Nonlinear Sci.* **17** (2007), no. 4, 309–347 Zbl 1119.37051 MR 2335124

[43] J. Morales, *Least action principles with applications to gradient flows and kinetic equations*. Ph.D. thesis, The University of Texas at Austin, 2017

[44] H. Neunzert, An introduction to the nonlinear Boltzmann-Vlasov equation. In *Kinetic theories and the Boltzmann equation (Montecatini, 1981)*, pp. 60–110, Lecture Notes in Math. 1048, Springer, Berlin, 1984 Zbl 0575.76120 MR 740721
[45] F. Otto, Evolution of microstructure in unstable porous media flow: A relaxational approach. *Comm. Pure Appl. Math.* **52** (1999), no. 7, 873–915 Zbl 0929.76136 MR 1682800
[46] F. Otto, The geometry of dissipative evolution equations: The porous medium equation. *Comm. Partial Differential Equations* **26** (2001), no. 1-2, 101–174 Zbl 0984.35089 MR 1842429
[47] F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** (2000), no. 2, 361–400 Zbl 0985.58019 MR 1760620
[48] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization A universal concept in nonlinear sciences*. Camb. Nonlinear Sci. Ser. 12, Cambridge University Press, Cambridge, 2001 Zbl 0993.37002 MR 1869044
[49] D. Poyato, Filippov flows and mean-field limits in the kinetic singular Kuramoto model. 2019, arXiv:1903.01305
[50] S. H. Strogatz, From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Phys. D* **143** (2000), no. 1–2, 1–20 Zbl 0983.34022 MR 1783382
[51] S. H. Strogatz and R. E. Mirollo, Stability of incoherence in a population of coupled oscillators. *J. Statist. Phys.* **63** (1991), no. 3-4, 613–635 MR 1115806
[52] S. H. Strogatz, R. E. Mirollo, and P. C. Matthews, Coupled nonlinear oscillators below the synchronization threshold: Relaxation by generalized Landau damping. *Phys. Rev. Lett.* **68** (1992), no. 18, 2730–2733 Zbl 0969.92501 MR 1160290
[53] G. Toscani and C. Villani, Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation. *Comm. Math. Phys.* **203** (1999), no. 3, 667–706 Zbl 0944.35066 MR 1700142
[54] J. L. van Hemmen and W. F. Wreszinski, Lyapunov function for the Kuramoto model of nonlinearly coupled oscillators. *J. Stat. Phys.* **72** (1993), no. 1–2, 145–166 Zbl 1097.82530
[55] V. S. Varadarajan, On the convergence of sample probability distributions. *Sankhyā* **19** (1958), 23–26 Zbl 0082.34201 MR 94839
[56] S. R. S. Varadhan, *Probability theory*. Courant Lect. Notes Math. 7, American Mathematical Society, Providence, RI, 2001 MR 1852999
[57] C. Villani, *Optimal transport*. Grundlehren Math. Wiss. 338, Springer, Berlin, 2009 MR 2459454
[58] P. Villegas, P. Moretti, and M. A. Muñoz, Frustrated hierarchical synchronization and emergent complexity in the human connectome network. *Sci. Rep.* **4** (2014), 5990

Received 19 June 2020; revised 19 December 2021; accepted 20 December 2021.

Javier Morales
Center for Scientific Computation and Mathematical Modeling (CSCAMM), University of Maryland, College Park, MD 20742-3289, USA; javierm1@cscamm.umd.edu

David Poyato
Institut Camille Jordan (ICJ), UMR 5208 CNRS & Université Claude Bernard Lyon 1, 69100 Villeurbanne, France; and Research Unit “Modeling Nature” (MNat), Universidad de Granada, Granada, 18071, Spain; poyato@math.univ-lyon1.fr