ON THE TREND TO GLOBAL EQUILIBRIUM FOR KURAMOTO OSCILLATORS

JAVIER MORALES AND DAVID POYATO

Abstract. In this paper, we study the 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 in a large coupling strength regime from generic initial data. 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 the oscillators will concentrate at the given rate tends to one 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 F. Otto in [35]; 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 the ones derived by F. Otto and C. Villani in [36]; 4) The study of a system of coupled differential inequalities, by a treatment inspired by the work of L. Desvillettes and C. Villani [13]. 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

1. Introduction 2
1.1. The Kuramoto model 2
1.2. The gradient flow structure and stationary solutions 4
1.3. Statement of the problem and main results 6
1.4. Ingredients 9
2. Strategy 9
2.1. Displacement concavity and entropy production 11
2.2. Small dissipation regime and lower bounds in the order parameter 12
2.3. Instability of the antipodal equilibria 13
2.4. Sliding norms 14

Date: March 9, 2022.

2010 Mathematics Subject Classification. 34C15, 34D06, 35B40, 35Q70, 35Q83, 70F99, 92B20, 92B25.

Key words and phrases. Kuramoto model, synchronization, Wasserstein distance, order parameters, gradient flow, entropy production, logarithmic Sobolev inequality, Talagrand inequality.

Acknowledgment. The work of J. Morales was supported by NSF grants DMS16-13911, RNMS11-07444 (KI-Net) and ONR grant N00014-1812465. The work of D. Poyato has been partially supported by the MECID (Spain) research grant FPU14/06304, the MINECO-Feder (Spain) research grant number MTM2014-53406-R and the Junta de Andalucia (Spain) Project FQM 954. Part of this research has been 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.
1. Introduction

In the present paper, we quantify the rate of convergence to the global equilibrium for $C^1$ solutions to the Kuramoto-Sakaguchi equation from generic initial data, 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 Y. Kuramoto several decades \cite{26,27} ago 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 community, see \cite{1,3,5,9,10,12,17,20,23,31,37,39,44}.

The main motivation to perform our study on the Kuramoto-Sakaguchi equation is threefold. First, such a 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. Such lack of convexity generates challenges to use theory of gradient flows to derive rates of convergence. Third, we are interested in quantifying the relaxation time of a nondeterministic event. Indeed, in a large coupling strength regime, one expects relaxation to the global equilibrium of the particle system with almost sure probability. However, such relaxation fails for some well prepared initial data.

In the case of identical oscillators, the Kuramoto-Sakaguchi equation exhibits a gradient flow structure in the space of probability measure 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
has been that many evolutionary equations of the form
\[ \frac{\partial \rho}{\partial t} = \text{div} \left( \nabla \rho + \rho \nabla V + \rho (\nabla W * \rho) \right), \]
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.,
\[ \Pi(\mu, \nu) := \{ \gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi_1 \# \gamma = \mu \text{ and } \pi_2 \# \gamma = \nu \}. \]

When such entropy functionals are convex with respect to the Wasserstein distance, such an interpretation allows proving entropy estimates and functional inequalities (see [43] for more details on this area). Such tools, in turns, 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 nonidentical 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 adapt the techniques developed by L. Desvillettes and C. Villani in [13] to derive quantitative convergence rates for a nonconvex gradient flow in the particular context of the Kuramoto-Sakaguchi equation. We hope that this provides insight on how to attack this difficulty in more general situations.

In this section, we shall 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 \( T \times \mathbb{R} \). Such dynamics is given by the system

\[
\begin{cases}
\dot{\theta}_i = \omega_i + \frac{K}{N} \sum_{j=1}^{N} \sin(\theta_j - \theta_i), \\
\theta_i(0) = \theta_{i,0},
\end{cases}
\]

for \( i = 1, \cdots, N \). 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 \( (t, \theta, \omega) \in \mathbb{R}^+ \times \mathbb{T} \times \mathbb{R} \)

\[
\begin{aligned}
\frac{\partial f}{\partial t} + \frac{\partial}{\partial \theta}(v[f]f) &= 0, \quad (\theta, \omega) \in \mathbb{T} \times \mathbb{R}, \ t \geq 0, \\
f(0, \theta, \omega) &= f_0(\theta, \omega), \quad (\theta, \omega) \in \mathbb{T} \times \mathbb{R}.
\end{aligned}
\]

We denote the velocity field by \( v[f] \), that is,

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

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 \( \rho \) 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 [28] using Neunzert’s method [33].

1.2. The gradient flow structure and stationary solutions. The Kuramoto model in \( \mathbb{T}^N \) can be lifted to a dynamical system in \( \mathbb{R}^N \). J. L. van Hemmen and W. F. Wreszinski [40] observed that by doing this the Kuramoto model can be formulated as a gradient flow of the energy

\[
V(\Theta) = -\frac{1}{N} \sum_{j=1}^{N} \omega_j \theta_j + \frac{K}{2N^2} \sum_{k,j=1}^{N} \left(1 - \cos(\theta_j - \theta_k)\right),
\]

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

\[
\langle v, w \rangle_N = \frac{v \cdot w}{N}.
\]

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

\[
\begin{aligned}
\dot{\Theta}(t) &= -\nabla_N V(\Theta(t)), \\
\Theta(0) &= \Theta_0,
\end{aligned}
\]

where \( \nabla_N \) 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

\[
r(\Theta) e^{i\phi(\Theta)} = \frac{1}{N} \sum_{k=1}^{N} e^{i\theta_k},
\]

then, we have that the potential reads

\[
V(\Theta) = -\frac{1}{N} \sum_{j=1}^{N} \omega_j \theta_j + \frac{K}{2} \left(1 - r^2(\Theta)\right),
\]

and the gradient slope take the form

\[
|\nabla_N V(\Theta)|_N^2 = \frac{1}{N} \sum_{j=1}^{N} \left|\omega_j - Kr \sin(\theta_j - \phi)\right|^2.
\]
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 \( \theta_i \) within \( \Theta \) tend to be synchronized around the same phase value. Moreover, using them we can rewrite system (1.1) as follows

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

for every \( i = 1, \cdots, N \). Without lost of generality we may assume that, the natural frequencies are centered, i.e.,

\[
\frac{1}{N} \sum_{i=1}^{N} \omega_i = 0.
\]

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, such condition is necessary to show the existence of stationary states and we shall assume it throughout the paper. For any such a stationary state \( \Theta_\infty \) so that \( r_\infty > 0 \), we must have that \( \nabla V(\Theta_\infty) = 0 \). Using (1.8), we readily obtain that at equilibria the following condition holds

\[
\max_{1 \leq j \leq N} |\omega_j| \leq Kr_\infty,
\]

and phases \( \theta_j \) must take some of the following two forms

\[
\begin{align*}
\theta_{j,\infty} &= \phi_\infty + \arcsin\left( \frac{\omega_j}{Kr_\infty} \right), \\
\theta_{j,\infty} &= \phi_\infty + \pi - \arcsin\left( \frac{\omega_j}{Kr_\infty} \right),
\end{align*}
\]

(1.10)

for every \( j = 1, \ldots, N \).

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

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

(1.11)

\( D_N^2 V \) denotes the Hessian operator with respect to the scaled inner product (1.5) and \( v = (v_1, \ldots v_N) \) is contained in \( \mathbb{R}^N \). From this, after accounting for the rotational invariance of the model, we deduce that the stable equilibrium must satisfy that

\[
\theta_{j,\infty} = \phi + \arcsin\left( \frac{\omega_j}{Kr_\infty} \right),
\]

for every \( j = 1, \ldots, N \).

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 rotation 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 Appendix A from [22]). In the nonidentical case, this structure is not strictly available. Nonetheless, in our analysis, we use several techniques and objects inspired by theory of gradient flows in the
space of probability measures. Similarly, if we consider the continuous version of the order parameters

\[ \Re e^{i\phi} = \int_{T \times \mathbb{R}} e^{i\theta} f(t, \theta, \omega) \, d\theta \, d\omega, \]

equation (1.12) can be restated as follows

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

Again, without loss of generality, we can assume that \( g \) is centered as well, i.e.,

\[ \int_{\mathbb{R}} \omega g(\omega) \, d\omega = 0. \]

Again, this is a necessary condition for equilibria to exist and we shall assume it throughout the paper. For any such equilibria \( f_\infty \) with corresponding \( R_\infty > 0 \), we obtain that \( \text{supp} \, g \subseteq [-KR_\infty, KR_\infty] \), and \( f_\infty \) takes the form

\[ f_\infty(\theta, \omega) = g^+(\omega) \otimes \delta_{\vartheta^+(\omega)}(\theta) + g^-(\omega) \otimes \delta_{\vartheta^-(\omega)}(\theta), \]

where,

\[ g = g^- + g^+, \]

for some non-negative \( g^- \), and \( g^+ \) and

\[ \vartheta^+(\omega) = \phi_\infty + \arcsin \left( \frac{\omega}{KR_\infty} \right), \]

\[ \vartheta^-(\omega) = \phi_\infty + \pi - \arcsin \left( \frac{\omega}{KR_\infty} \right), \]

for each \( \omega \in \text{supp} \, g \). As it will become apparent later along the paper, 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 (1.11), 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 one, the system (1.1) should converge to such equilibria if the coupling strength is sufficiently large. Such phenomenon has been widely observed in numerical simulations. However, to the date, this result is absent from the literature. It has only been verified for restricted initial configurations where all of the oscillators are constrained in an arc of the circle \( [12] \).

There have been many approaches in the literature to show the convergence of the system to the critical points of (1.11) in the large coupling strength regime. Since stable equilibria have oscillators contained within an interval of size less than \( \pi \), 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 [12, 20] a system of differential inequalities was found for the phase and frequency diameter, that 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 [24, 29] the authors derived the convergence to phase-locked states using Lojasiewicz gradient’s inequality for analytic potentials [30] and it was used to obtain convergence rates (after some unquantified initial time) in some particular cases where the Lojasiewicz exponent can be explicitly computed. For general initial data along the whole circle, the literature is rare and the main contribution is [21], but rates are not available. One of the main difficulties when trying to use standard theory from dynamical systems to show this is the fact that critical points of (1.11) are not isolated (see Remark 1.1).

In the continuum case, accumulation of oscillators in the hemisphere opposite of the order parameter was excluded in [22]. However, convergence towards a stationary solution was not established yet for generic initial data. See [10] for a particular proof when the phase diameter is smaller than $\pi$. 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.

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. 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, we can 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 [14, 15, 16, 17], 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 L. Desvillettes and C. Villani:

*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 later has to be estimated by techniques which would be well-adapted to the nonlinear equation.*

Here is where our contribution takes place, and this is why we shall not rely on linearization techniques. Instead, we shall 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

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

Then, there exists a universal constant $C$ such that if (1.15)
then we can find a time $T_0$ with the property that
\begin{equation}
T_0 \lesssim \frac{1}{KR_0} \log \left( 1 + W^{1/2} ||f_0||_2 + \frac{1}{R_0} \right),
\end{equation}
and
\[ W_2(f, f_\infty) \lesssim e^{-\frac{1}{4}K(t-T_0)}, \]
for every $t$ in $[T_0, \infty)$. Here, 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 $h_1 \leq Ch_2$. Since our argument is constructive, every time we use such a 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.

**Corollary 1.1.** Let $\mu_t^N$ 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 $\phi_\infty$ of the global equilibrium $f_\infty$. Then, there exists a positive time $T_0$ satisfying (1.16) and an integer $N^*$ with the property that
\[ \log N^* \lesssim \frac{1}{R_0} \log \left( 1 + W^{1/2} ||f_0||_2 + \frac{1}{R_0} \right), \]
and for any $N \geq N^*$ and any $s$ contained in the interval
\[ [T_0, T_0 + \frac{1}{25} \log \left( \frac{N}{N^*} \right)], \]
we can quantify the probability of mass concentration and diameter contraction of the particle system with $N$ oscillators. Indeed, we have that
\[ \mathbb{P}\left( \forall t \geq s, \exists L^N_s(t) \subseteq \mathbb{T} : L^N_s(t) = L \text{ and (M) - (D) holds} \right) \geq 1 - C_1e^{-C_2N^{\frac{1}{2}}}. \]

Here, conditions (M) and (D) yield mass concentration and diameter contraction. More precisely, such properties are given by
\begin{align*}
(M) & \quad \mu^N_t(L^N_s(t) \times \mathbb{R}) \geq 1 - \frac{1}{5}e^{-\frac{K}{4}(s-T_0)}, & \text{for every } t \in [s, \infty), \\
(D) & \quad \text{diam}(L^N_s(t)) \leq \max \left\{ \frac{4}{5}e^{-\frac{K}{4}(t-s)}, 12 \frac{W}{K} \right\}, & \text{for every } t \in [s, \infty). 
\end{align*}

Additionally, $C_1$ and $C_2$ are universal positive constants which could be explicitly computed.
1.4. **Ingredients.** The proof of Theorem [1.14](#) is the first quantitative proof for the relaxation problem for Kuramoto oscillators with generic initial data. 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 the probability measures under the Wasserstein distance, first introduced by F. Otto in [35](#), which we address in Sections 2.3 and 5. See also [22](#), Appendix A] for an overview in the context of Kuramoto-Sakaguchi with identical oscillators.
- A fibered Wasserstein distance $W_{2,g}$ presented independently in [32](#) and [38](#). Such a distance is well adapted to the nonlinear problem. By using this distance, in Section 3 we will derive new logarithmic Sobolev and Talagrand type inequalities associated with it (see [36](#)).
- A quantitative instability estimate excluding the equilibria with mass in the opposite pole of the order parameter, that we derive in Section 4. A form of such an estimate was originally presented in [22](#), but we use a more refined version in this work.
- 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 4.1.

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

## 2. Strategy

In this section, we shall 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 most attractive 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 shall overcome three crucial difficulties. First, the order parameter $R$ defined in [1.12](#) 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 plays an essential role in the particle system as well). The second difficulty is the fact that 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 that

$$
\frac{d}{dt} V(\Theta(t)) = -|\nabla_N V(\Theta(t))|^2_N.
$$
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 (1.11) of the energy functional for the particle system. Such an expression implies that $D^2_N V(\Theta(t))$ is bounded from above (as a quadratic form) by $Kr(\Theta(t))$, that is, $$\langle D^2_N V(\Theta(t))v, v \rangle_N \leq Kr(\Theta(t)) |v|^2_N,$$ 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 (1.7) $V(\Theta(t))$ and $1 - r^2(t)$, are related up to lower-order terms that can be neglected thanks to condition (1.15). Similarly, considering the time derivative of the above quantities, we have that the following two expressions $$\frac{|\nabla_N V(\Theta(t))|^2}{K} \quad \text{and} \quad \frac{dr^2}{dt}(t),$$ should also differ by a lower-order term that, again, can be controlled using (1.15). This justifies that, in the large coupling strength regime, we indistinctly call $\frac{dr^2}{dt}$ and $|\nabla_N V(\Theta(t))|^2_N$ 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 introduced by F. Otto for the space of probability measures (see [22, 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 (1.8) given, by, $$\mathcal{I}[f] := \int_{T \times \mathbb{R}} (\omega - KR \sin(\theta - \phi))^2 f \, d\theta \, d\omega.$$ We shall again call this quantity the dissipation. Indeed, notice that taking derivatives in (1.12), one clearly obtains the following dynamics of the order parameters

$$\dot{R} = -\int_{T \times \mathbb{R}} \sin(\theta - \phi)(\omega - KR \sin(\theta - \phi)) f \, d\theta \, d\omega,$$

$$\dot{\phi} = \frac{1}{R} \int_{T \times \mathbb{R}} \cos(\theta - \phi)(\omega - KR \sin(\theta - \phi)) f \, d\theta \, d\omega.$$ Using it, we will show, in Lemma 3.2 that dissipation and time derivative of the order parameter are again related up to lower-order terms that can be controlled by condition (1.15), 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,

$$-2KRI[f] \leq \frac{d}{dt}\mathcal{I}[f] \leq 2KI[f].$$
In Section 2.3, we will describe how this relationship along with the principle of entropy production, can be used to provide a universal lower bound of $R(t)$ of the form $\lambda R_0$, for some $\lambda$ in $(0, 1)$. In fact, we will show that by making $K$ sufficiently large we can make $\lambda$ as close to one as needed.

2.1. Displacement concavity and entropy production. Before entering into the details of the entropy production principle, we set some necessary notation. We define a dynamic neighborhood of the order parameter $\phi$ and its antipode as follows.

**Definition 2.1.** Given an angle $\alpha$ in $(0, \frac{\pi}{2})$, we denote by $L^+_{\alpha}(t)$ the interval (arc) in $T$ 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 $T$ 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 $B \subseteq T$ we define

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

and more generally, we will let

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

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

Now we describe the entropy production principle in our context. Roughly speaking, it will quantify the following 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 a lot in the next few instants of time.*

To make it rigorous, let us come back to the dissipation functional (2.1). As for the particle system (1.8), notice that $I[f]$ vanishes if, and only if, $f$ is an equilibrium. Hence, $I[f]$ can be thought of a natural measure of how close a given $f$ is to the family of equilibria (1.14). Notice that such expression of equilibria (1.14) guarantees that, by our assumption (1.15) on $W_K$, all the possible equilibria in our analysis have phase support confined to small arcs centered around $\phi$ and its antipode $\phi + \pi$. Since the diameter of the neighborhood can be made arbitrarily small due to hypothesis (1.15), then we can fix any small enough value of $\alpha$ for the size of the neighborhood $L^+_{\alpha}(t) \cup L^-_{\alpha}(t)$. For simplicity, we will set $\alpha = \pi/6$ all along the paper.

The entropy production principle then shows that, in the large coupling strength regime, if entropy production is small (i.e., the time derivative of the order parameter is small),
then most of the mass of the system lies in the neighborhood \( L^+_\alpha(t) \cup L^-_\alpha(t) \) of \( \phi(t) \) and its antipode. Specifically, in the proof of Proposition 3.4 we will quantify such assertion as follows

\[
(2.3) \quad \rho(T \setminus L^+_\alpha(t) \cup L^-_\alpha(t)) \leq \frac{1}{KR^2 \cos^2 \alpha} \frac{d}{dt} R^2 + \frac{W^2}{K^2 R^2 \cos^2 \alpha}.
\]

In other words, (2.3) suggests that when \( f \) is sufficiently far from the family of equilibria (1.14) (i.e. it has enough mass outside the time-dependent neighborhood \( L^+_\alpha(t) \cup L^-_\alpha(t) \)), then the dissipation \( I[f] \) is large. Consequently, the time derivative of the order parameter is large in this case as well, and this produces an entropy production of the system.

In the Lemma below, we quantify the corresponding gain in the order parameter.

**Lemma 2.1. (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 \( \lambda \) be contained in \((0, 1)\). Additionally, suppose that \( \sqrt{2} R_0 \geq R(t_0) > \lambda R_0 \) and \( \dot{R}(t_0) \geq \frac{K}{4} \cos^2 \alpha \lambda^3 R_0^3 \).

Then, there exists a universal constant \( C \) such that if

\[
(2.4) \quad \frac{W}{K} \leq C \lambda^2 R_0^2,
\]

then,

\[
(2.5) \quad R^2(t_0 + d) - R^2(t_0) \geq \frac{1}{40} \lambda^4 R_0^3.
\]

Moreover, we can select \( d \) in such a way that

\[
(2.6) \quad d \leq \frac{1}{3KR_0} \log 10,
\]

and

\[
R \leq \frac{3}{2} R_0 \quad \text{in} \quad [t_0, t_0 + d].
\]

**2.2. Small dissipation regime and lower bounds in 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, Proposition 3.4 in Section 3 will show that when \( \dot{R} \) is below a critical threshold, we achieve the following differential inequality

\[
(2.7) \quad \frac{d}{dt} R^2 > \frac{K}{2} \left( -R^3 + |\lambda R_0 + \frac{3}{5}(1 - \lambda) R_0| R^2 - \frac{3}{5}(1 - \lambda) \lambda^2 R_0^3 \right),
\]

which hold in any time interval \([t_1, t_2]\) such that

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

The estimate (2.3) will be crucial to derive such a proposition. Additionally, note that the right-hand side of (2.6) vanishes when \( R = \lambda R_0 \). In Corollary 3.6 we will combine this inequality with the above entropy production in Lemma 2.1 to quantify a universal lower bound \( R(t) \geq \lambda R_0 \) of 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 difficulty we need to quantify the instability of the antipodal equilibrium, that 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 follows

$$\chi_{\alpha,\delta_0}(\theta) = \xi_{\alpha,\delta_0}(\theta - \phi - \pi),$$

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

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

As for $\alpha$, we can take $\delta_0$ as small as desired. For notational simplicity we will set $\xi := \xi_{\alpha,1/2}$ and $\chi := \chi_{\alpha,1/2}$.

Additionally, we will use the notation

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

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

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

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

$$\frac{d}{dt} f^2(\chi^-_\alpha(t)) \leq -KR \sin \alpha f^2(\chi^-_\alpha(t)) + 4K f^2_1(\mathbb{T}) \left[ \frac{W}{K} + \sqrt{\frac{2R}{KR} + \frac{1}{R^2} \frac{W^2}{K^2}} - R \cos \alpha \right].$$

Although this inequality is a variant of an estimate previously introduced in [22], we prove it in Proposition 4.1 because it fits better 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 inequality to quantify the convergence rates, but the dissipation is not sufficiently small, one sees that the term $f^2_1(\mathbb{T})$ represents an obstacle. Specifically, it stands to reason that one can produce examples in which $f^2_1(\mathbb{T})$ 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 subsection.

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 \( X_{t_0,t}(\theta, \omega) = (\Theta_{t_0,t}(\theta, \omega), \omega) \) denote the forward flow map, that is,

\[
\begin{cases}
\frac{d}{dt}X_{t_0,t}(\theta, \omega) = (v[f], 0), \\
X_{t_0,t_0}(\theta, \omega) = (\theta, \omega),
\end{cases}
\]

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

For any measurable set \( A \subseteq T \times \mathbb{R} \), we will denote the image \( X_{t_0,t}(A) \) by \( A_{t_0,t} \). For simplicity, when considering a time-dependent set \( A(t) \), we will use the notation \( A(t_0) \) to denote \( A(t_0) \).

Additionally, given a measurable set \( B \subseteq T \), we will use \( B_{t_0,t} \) to denote the projection of \( (B \times [-W, W])_{t_0,t} \) into \( T \). Again, if \( B(t) \) is a time-dependent set in \( T \), we will use \( B(t_0) \) to denote the projection of \( (B(t_0) \times [-W, W])_{t_0,t} \) into \( T \).

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

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

and holds for any measurable set \( A \subseteq T \times \mathbb{R} \). We prove such 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 regime. We perform this analysis in Section 4.1.

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

\[
\begin{align*}
(2.8) \quad \frac{d}{dt} f^2(A_{t_0,t}) & \leq KR \left( \sup_{(\theta, \omega) \in A_{t_0,t}} \cos(\theta - \phi(t)) \right) f^2(A_{t_0,t}), \\
(2.9) \quad -2KRI[f] & \leq \frac{d}{dt}I[f] \leq 2KI[f], \\
(2.10) \quad I[f] - W^2 & \leq K \frac{d}{dt}R^2 \leq 3I[f] + W^2, \\
(2.11) \quad \frac{d}{dt}f^2(\chi^-(t)) & \leq -KR \sin \alpha f^2(\chi^-(t)) + 4Kf^2(\mathbb{T}) \left[ \frac{W}{KR} + \sqrt{\frac{2R}{KR} + \frac{1}{R^2} \frac{W^2}{K^2} - R \cos \alpha} \right]^{+}, \\
(2.12) \quad \frac{d}{dt}R^2 & > K \left( -R^3 + [\lambda R_0 + \frac{3}{5}(1 - \lambda)R_0]R^2 - \frac{3}{5}(1 - \lambda)\lambda^2 R_0^3 \right),
\end{align*}
\]

where the first inequality holds for any measurable set \( A \subseteq T \times \mathbb{R} \), the last inequality holds in any interval \([t_1, t_2]\) satisfying the hypotheses of Proposition 3.4 and all of the other inequalities above holds for every \( t \) in \([0, \infty)\).
The goal of such a system is to derive an explicit bound on the time \( T_0 \) in 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 4.1. On the other hand, we recover the approach developed by L. Desvillettes and C. Villani in [13] in our setting. Such an argument is described in detail in Section 5 and it consists of 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 (2.9) in the form of our entropy production estimate to quantify the increase of the order parameter. Conversely, if the dissipation is small, we use (2.11) 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 (2.8) and our analysis on the dynamics of sets from Section 4.1.

As a result of the above analysis, we obtain the following corollary:

**Corollary 2.1.** 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) and let \( \beta = \pi/3 \). Then, there exists a universal constant \( C \) such that if \( W \leq CR_0^3 \),

then we can find a time \( T_0 \) with the property that

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

and

\[
R(t) \geq \frac{3}{5} \quad \text{and} \quad \rho(\mathbb{T}\backslash L_{\beta}^+(t)) \leq e^{-\frac{1}{20}K(t-T_0)},
\]

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

Such a Corollary is the starting point of the last part of our strategy.

### 2.6. Local displacement convexity and Talagrand type inequalities

At the particle level, we see that the Hessian operator (1.11) is positive definite in the subspace orthogonal to rotation whenever the oscillators are strictly contained on a suitable interval. As mentioned in Section 1, the classical theory of gradient flows allows deriving convergence rates towards equilibrium when the energy is strictly convex. Thus, once the mass enters exponentially fast to the region of convexity after \( T_0 \), one may hope to recover such a convergence result for our system. Indeed, inspired by the arguments in [36] on their proof of 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 \( W_{2,g} \), independently introduced in [32] and [38], which is well adapted to the nonlinear problem. The proof of such inequalities is 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 Kuramoto-Sakaguchi
equation (1.2). Firstly, recall that from many preceding works (see e.g., [5, 10, 22]) 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 Corollary 2.1 where such a convexity area is described by a dynamic neighborhood of the order parameter $\phi$.

The proof of such result is postponed to forthcoming sections and becomes the cornerstone of this paper. We devote this part to study the other main feature of the dynamics. Specifically, we show that although the system is not a Wasserstein gradient flow, the generalized dissipation functional that has been introduced in (2.1) satisfies an appropriate Hessian-type inequality after the solution has entered into the concentration regime quantified in Corollary 2.1. The final step is inspired in [36] about the derivation of the logarithmic Sobolev and Talagrand inequalities for gradient flows in the Wasserstein space. Indeed, we shall show that despite the fact that our system is not a Wasserstein gradient flow due to the presence of heterogeneities introduced by $\omega$, some dissipation-transportation inequality still can be achieved for an adequate distance on the space of probability measures. Such inequality along with the exponential decay of the dissipation guarantee 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,

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

**Proof.** Taking derivatives yields the Wasserstein two terms

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

where each of them takes the form

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

$$I_2 := \int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi))^2 \phi f d\theta d\omega.$$
Let us use (2.2) and substitute the formulas for $\dot{R}$ and $\dot{\phi}$ in each term. By doing this, we get that

\begin{equation}
I_1 = 2K \int_{\mathbb{T}^2 \times \mathbb{R}^2} \left( \omega - KR \sin(\theta - \phi) \right) \left( \omega' - KR \sin(\theta' - \phi) \right) \\
\times \left( \sin(\theta - \phi) \sin(\theta' - \phi) - \cos(\theta - \phi) \cos(\theta' - \phi) \right) f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega',
\end{equation}

and

\begin{equation}
I_2 = \int_{\mathbb{T} \times \mathbb{R}} \partial_\theta \left[ (\omega - KR \sin(\theta - \phi))^2 \right] (\omega - KR \sin(\theta - \phi)) f \, d\theta \, d\omega,
\end{equation}

where we have used the Kuramoto-Sakaguchi equation (1.2) and integration by parts. Notice that by definition of the order parameter (1.12), we obtain

\begin{equation}
R \cos(\theta - \phi) = \int_{\mathbb{T} \times \mathbb{R}} \cos(\theta - \theta') f(t, \theta, \omega') \, d\theta' \, d\omega'.
\end{equation}

Using such identity in the above formula for $I_2$ implies

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

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

\begin{equation}
I_2 = -K \int_{\mathbb{T}^2 \times \mathbb{R}^2} \left( (\omega - KR \sin(\theta - \phi))^2 + (\omega' - KR \sin(\theta' - \phi))^2 \right) \\
\times \cos(\theta - \theta') f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega'.
\end{equation}

Finally, putting (3.1) and (3.4) together and completing the square yield 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,

\begin{equation}
-2KR \mathcal{I}[f] \leq \frac{d}{dt} \mathcal{I}[f] \leq 2K \mathcal{I}[f],
\end{equation}

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

\begin{equation}
\mathcal{I}[f](t_0) e^{-2K \int_{t_0}^t R(s) \, ds} \leq \mathcal{I}[f](t) \leq \mathcal{I}[f](t_0) e^{2K(t-t_0)},
\end{equation}

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

Proof. Note that the second chain of inequalities follows from by integration on \((3.5)\) with respect to time. Then, we focus on the proof of \((3.5)\), that we divide in two steps associated with the upper 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 \(f\):

\[
\frac{d}{dt} I[f] \leq \int_{\mathbb{T} \times \mathbb{R}^2} \left((\omega - KR \sin(\theta - \phi)) - (\omega' - KR \sin(\theta' - \phi))\right)^2 f(t, \theta, \omega') d\theta' d\omega' d\omega
\]

\[
= 2K \int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi))^2 f d\theta d\omega
\]

\[
- 2K \left(\int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi)) f d\theta d\omega\right)^2.
\]

Using the definition \((1.12)\) of \(R\) and \(\phi\) along with the assumption \((1.13)\), we clearly obtain that the second term vanishes and we conclude the upper bound.

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

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

\[
+ 2K \int_{\mathbb{T} \times \mathbb{R}^2} (\omega - KR \sin(\theta - \phi))(\omega' - KR \sin(\theta' - \phi)) \cos(\theta - \theta')
\]

\[
\times f(t, \theta, \omega) f(t, \theta', \omega') d\theta d\theta' d\omega d\omega'
\]

\[
= -2KR \int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi))^2 \cos(\theta - \phi) f d\theta d\omega
\]

\[
+ 2K \left| \int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi)) e^{i(\theta - \phi)} f d\theta d\omega\right|^2
\]

\[
\geq -2KR \int_{\mathbb{T} \times \mathbb{R}} (\omega - KR \sin(\theta - \phi))^2 f d\theta d\omega,
\]

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. \(\square\)

### 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 fiber-wise way, we call it the fibered quadratic Wasserstein distance. See also \([32]\) and \([38]\), where it was introduced independently by both authors. For the reader convenience, we recall it here and introduce some of the main properties that will be used throughout the paper.

**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 $\omega$-marginal agrees with $g$, i.e.,
$$P_g(T \times \mathbb{R}) := \{ \mu \in P(T \times \mathbb{R}) : (\pi_\theta)_{\#} \mu = g \}.$$  
We define the fibered quadratic Wasserstein distance on $P_g(T \times \mathbb{R})$ as follows
\begin{equation}
W_{2,g}(\mu, \nu) := \left( \int_\mathbb{R} W_2(\mu(\cdot | \omega), \nu(\cdot | \omega))^2 \, d\omega g \right)^{1/2},
\end{equation}
for any $\mu, \nu \in P_g(T \times \mathbb{R})$. Here, we denote family of conditional probabilities (or disintegrations) of $\mu$ with respect to the fiber $\omega \in \mathbb{R}$ as follows
$$\omega \in \mathbb{R} \mapsto h_{\tau} := \mu(\cdot | \omega) \in P(T),$$
that is a Borel measurable function defined by the following formula
$$\int_{T \times \mathbb{R}} \varphi(\theta, \omega) \, d(\theta, \omega) = \int_{\mathbb{R}} \left( \int_{T} \varphi(\theta, \omega) \, d\theta \mu(\cdot | \omega) \right) \, d\omega g,$$
for any test function $\varphi \in C_b(T \times \mathbb{R})$.

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

**Proposition 3.1.** Consider $g \in P(\mathbb{R})$ and let $f^1, f^2 \in P_g(T \times \mathbb{R})$. For $g$-a.e. value of $\omega \in \mathbb{R}$, let us consider some Wasserstein geodesic $\tau \in [0,1] \longrightarrow h_\tau(\cdot | \omega) \in P(T)$ that joins the conditional probabilities with respect to $\omega$, that is
$$h_{\tau=0}(\theta) = f^1(\cdot | \omega) \quad \text{and} \quad h_{\tau=1}(\theta) = f^2(\cdot | \omega).$$
This is an absolutely continuous family with respect to the Wasserstein distance on $T$ and it has an associated family of potentials $\tau \in [0, T] \longrightarrow \psi_\tau(\cdot, \omega)$ so that
\begin{equation}
\begin{aligned}
\frac{\partial}{\partial \tau} h_\tau(\cdot | \omega) + \text{div}_\theta (\nabla_\theta \psi_\tau(\cdot, \omega) h_\tau(\cdot | \omega)) &= 0, \\
\frac{\partial}{\partial \tau} \psi_\tau(\cdot, \omega) + \frac{1}{2} |\nabla_\theta \psi_\tau(\cdot, \omega)|^2 &= 0, \quad \psi_{\tau=0}(\cdot, \omega) = \psi_0(\cdot, \omega),
\end{aligned}
\end{equation}
for some $\frac{d^2}{2}$-concave function $-\psi_0$ with respect to $\theta$, in the distributional/viscosity sense. Then, the following identity holds true
\begin{equation}
W_{2,g}(f^1, f^2)^2 = \int_0^1 \int_{T \times \mathbb{R}} |\nabla_\theta \psi_\tau|^2 \, dh_\tau \, d\tau,
\end{equation}
where we denote $h_\tau$ to the measure that can be recovered from the conditional probabilities $h_\tau(\cdot | \omega)$ with marginal $g$, that is, for any test function $\varphi \in C_b(T \times \mathbb{R})$ the disintegration formula holds
$$\int_{T \times \mathbb{R}} \varphi(\theta, \omega) \, dh_\tau = \int_{\mathbb{R}} \left( \int_{T} \varphi(\theta, \omega) \, d\theta h_\tau(\cdot | \omega) \right) \, d\omega g.$$  
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 in the textbooks [2, 4] and [33, 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 follows

\[(3.9) \quad W^2_{2,g}(f^1, f^2)^2 = \int_{T \times \mathbb{R}} |\nabla_{\theta} \psi_{\tau}|^2 h_{\tau} \, d\theta \, d\omega,\]

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

An interesting fact is that this new fibered quadratic Wasserstein distance and the classical quadratic Wasserstein distances in \(P_2(T \times \mathbb{R})\) are appropriately ordered. Before we state the relation, let us remark the following fact.

Remark 3.2. The classical quadratic Wasserstein distance \(W^2_2\) in \(P_2(T \times \mathbb{R})\) is defined as the transportation cost associated with the standard Riemannian distance in the product space \(T \times \mathbb{R}\). That is, \(W^2_2\) is defined by

\[W^2_2(\mu^N_0, f_0) = \left( \inf_{\gamma \in \Pi(\mu^N_0, f_0)} \int_{T^2 \times \mathbb{R}^2} \left( d(\theta, \theta')^2 + (\omega - \omega')^2 \right) \, d\gamma \right)^{1/2},\]

for any \(\mu, \nu \in P_2(T \times \mathbb{R})\). Here, \(d(\theta, \theta')\) denotes the canonical Riemannian distance in between any two point \(\theta\) and \(\theta'\) in \(T\).

For our purposes, such distance is not appropriate as it is not dimensionally correct. Indeed, \(\theta\) and \(\omega\) have different physical units and considering \(W^2_2\) causes problems to derive asymptotic behavior of solutions.

The above remark suggests considering the following correction of the classical quadratic Wasserstein distance in \(P_2(T \times \mathbb{R})\).

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

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

We define the scaled quadratic Wasserstein distance on \(P_2(T \times \mathbb{R})\) by the transportation costs associated with the above scaled Riemannian distance, that is,

\[SW^2_2(\mu^N_0, f_0) = \left( \inf_{\gamma \in \Pi(\mu, \nu)} \int_{T^2 \times \mathbb{R}^2} \left( d(\theta, \theta')^2 + \frac{(\omega - \omega')^2}{K^2} \right) \, d\gamma \right)^{1/2},\]

for any \(\mu, \nu \in P_2(T \times \mathbb{R})\).

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

Proposition 3.2. Consider \(g \in P_2(T)\). Then we obtain

\[SW_2(\mu, \nu) \leq W^2_{2,g}(\mu, \nu),\]

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

\[W_2(\mu, \nu) \leq W^2_{2,g}(\mu, \nu).\]

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 \(\gamma \in P(T^2 \times \mathbb{R}^2)\) given by

\[(3.10) \quad \gamma := \gamma_{0, \omega}(\theta, \theta') \otimes \delta_\omega(\omega') \otimes g(\omega).\]
Let us see first that it defines a transference plan, that is, $\gamma \in \Pi(\mu, \nu)$. To such end, consider any test function $\varphi \in C_b(\mathbb{T} \times \mathbb{R})$ and note that
\[
\int_{\mathbb{T} \times \mathbb{R}} \varphi d(\theta, \omega) (\pi(\theta, \omega) \# \gamma) = \int_{\mathbb{T}^2 \times \mathbb{R}^2} \varphi(\theta, \omega) d(\theta, \omega) \gamma_{0, \omega} d\omega (\delta_\omega) d\omega
\]
\[
= \int_{\mathbb{T}^2 \times \mathbb{R}^2} \varphi(\theta, \omega) d(\theta, \omega) \gamma_{0, \omega} d\omega
\]
\[
= \int_{\mathbb{T}^2 \times \mathbb{R}^2} \varphi(\theta, \omega) d(\pi \# \gamma_{0, \omega}) d\omega
\]
\[
= \int_{\mathbb{T} \times \mathbb{R}} \varphi(\theta, \omega) d\theta \mu (|\omega|) d\omega
\]
Then, $\pi(\theta, \omega) \# \gamma = \mu$. Similarly, note that
\[
\int_{\mathbb{T} \times \mathbb{R}} \varphi d(\theta', \omega') (\pi(\theta', \omega') \# \gamma) = \int_{\mathbb{T}^2 \times \mathbb{R}^2} \varphi(\theta', \omega') d(\theta', \omega') \gamma_{0, \omega'} d\omega' (\delta_\omega) d\omega
\]
\[
= \int_{\mathbb{T}^2 \times \mathbb{R}^2} \varphi(\theta', \omega') d(\theta', \omega') \gamma_{0, \omega'} d\omega'
\]
\[
= \int_{\mathbb{T} \times \mathbb{R}} \varphi(\theta', \omega') d\theta' \nu (|\omega|) d\omega'
\]
Then we also recover $\pi(\theta', \omega') \# \gamma = \nu$. Also note that by definition
\[
W_{2,g}(\mu, \nu)^2 = \int_{\mathbb{R} \times \mathbb{T}^2} d(\theta, \omega)^2 d(\theta, \omega) \gamma_{0, \omega} d\omega g = \int_{\mathbb{T}^2 \times \mathbb{R}^2} d(\theta, \omega)^2 d(\theta, \omega) (\pi(\theta, \omega)) \gamma
\]
\[
= \int_{\mathbb{T}^2 \times \mathbb{R}^2} d_K((\theta, \omega), (\theta', \omega')) d(\theta, \omega) (\pi(\theta, \omega)) \gamma \geq SW_{2}(\mu, \nu)^2,
\]
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
\[
W_{2,g}(\mu, \nu)^2 = \varepsilon_\theta^2 \quad \text{and} \quad SW_{2}(\mu, \nu)^2 = \frac{1}{K^2} \min\{\varepsilon_\theta^2, \varepsilon_{\omega}^2\}.
\]
Consequently, we obtain that
\[
SW_{2}(\mu, \nu) < W_{2,g}(\mu, \nu), \quad \text{if} \quad \frac{\varepsilon_{\omega}}{K} < \varepsilon_\theta,
\]
\[
SW_{2}(\mu, \nu) = W_{2,g}(\mu, \nu), \quad \text{if} \quad \frac{\varepsilon_{\omega}}{K} \geq \varepsilon_\theta.
\]
We are now ready to state the main relation between this fibered transportation distance (3.4) 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 \eqref{eq:KS}. 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 \cite{38} Theorem 4.4. There, the author used the definition of $W_{2,g}$ in \eqref{eq:W2} for general measures that may enjoy atoms eventually. In this result, we sketch a simpler proof that used the representation formula of the derivative of Wasserstein conditional probability with respect to $\omega$ of \eqref{eq:W2} so that it represents a Wasserstein geodesic joining the conditional probabilities $\mathbb{P}_f$ and $\mathbb{P}_{f_s}$.

Proof of Lemma 3.1. Since $f$ satisfies the Kuramoto-Sakaguchi equation \eqref{eq:KS}, then each conditional probability with respect to $\omega \in \mathbb{T}$ verifies the following continuity equation

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

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

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

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 $\omega \in \text{supp} g$, see \cite{2} Theorem 8.4.6 or \cite{43} Theorem 23.9]. In particular, we can take derivatives almost everywhere and obtain the formula

$$\frac{d}{ds} \frac{1}{2} W_{2,g}(f_t, f_s)^2 = - \int_{\mathbb{T}} \left\langle v^\omega_s(\theta), \nabla \psi^{\ast, r}_{T=0}(\theta, \omega) \right\rangle f_s(\theta|\omega) d\theta,$$

for almost every $t \geq 0$, where the family $\tau \in [0,1] \mapsto (h^{s,t}_\tau, \psi^{s,t}_\tau)$ has been chose according to \eqref{eq:geodesic} 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 \eqref{eq:geodesic} we obtain that

$$\frac{d}{ds} \frac{1}{2} W_{2,g}(f_t, f_s)^2 = - \int_{\mathbb{T} \times \mathbb{R}} \left\langle v^\omega_s(\theta), \nabla \psi^{s,t}_{\tau=0}(\theta, \omega) \right\rangle f_s(\theta|\omega) g(\omega) d\theta d\omega \leq \mathcal{I}[f] \frac{1}{2} W_{2,g}(f_t, f_s),$$

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

$$W_{2,g}(f_t, f_s) \leq \int_t^s \mathcal{I}[f_r]^{1/2} \, dr, \quad \text{for all } s \geq t.$$  

### 3.2. Convergence and uniqueness of the global equilibria.

In this section, we shall show the claimed result about convergence to the global equilibria. Before we proceed with the proof, let us first show that such equilibrium is unique up to phase rotations. That result is not new and was first proved in \([10]\) via a strict contractivity estimate in such region of convexity for an appropriate Wasserstein distance $\widetilde{W}_p$ in $\mathcal{P}_2([0,2\pi) \times \mathbb{R})$. Notice that the geometry of $T$ has been disregarded in $\widetilde{W}_p$. Indeed, the distance $\widetilde{W}_2$ is strictly larger $W_{2,g}$ because the geometry of the $T$ reduces the transportation cost of mass between phases separated by distances larger that $\pi$ (when viewed in the real line). We show that the uniqueness result is also true using this new fibered distance and we leave the full study of similar strict contractivity of $W_{2,g}$ to future works.

**Proposition 3.3.** Let $f_\infty$ and $f'_\infty$ be stationary measure-valued solutions to \((1.2)\) and assume that they have the same distribution $g$ of natural frequencies and, $\text{diam}(\text{supp}_g f_\infty)$ and $\text{diam}(\text{supp}_\theta f'_\infty)$ are less 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 $\theta$

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

and define the following optimization problem

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

Such 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 $f'_\infty$ with $\mathcal{T}_{c_0}[f'_\infty]$ and it does not change thesis of this result. On the one hand, let us consider the following continuity equation

\[(3.14)\quad \begin{cases} \frac{\partial}{\partial t} f'_s + \text{div}_\theta(e^{\theta} f'_s) = 0, \\ f'_s|_{s=0} = f'_\infty, \end{cases}\]

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 the problem \((3.13)\), then we obtain a critical value at $c = 0$, i.e.,

\[(3.15)\quad \frac{d}{ds} \bigg|_{s=0} W_{2,g}(f_\infty, f'_s)^2 = 0.\]

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 represent it through a family

\[(3.16)\quad \tau \in [0, T] \rightarrow (h^s(\tau, \psi^*_s) \text{ with } h^s_{s=0}(\cdot | \omega) = f'_\infty(\cdot | \omega), h^s_{s=1}(\cdot | \omega) = f'_s(\cdot | \omega),\]

and define the following optimization problem

\[\mathcal{V} \cdot | \omega) = f'_\infty(\cdot | \omega), h^s_{s=1}(\cdot | \omega) = f'_s(\cdot | \omega),\]
as in (3.7) in Proposition 3.1. Here, although (3.7) holds only on the viscosity/distributional
such fact can be handled by nowadays standard regularization arguments, we refer the
reader to [43, Chapter 13]. (In particular our dissipation functional \( I[f] \) 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 \( \theta \in T \rightarrow e^{i\theta} \). Then, the same ideas in the
proof of Lemma 3.1 (see [2, Theorem 8.4.6] or [43, Theorem 23.9]), we obtain

\[
\frac{d}{ds} \bigg|_{s=0} \frac{1}{2} W_2(f_\infty(\cdot|\omega), f_s'(\cdot|\omega))^2 = \int_T \left\langle e^{i\theta}, \nabla_{\theta} \psi_s^{s=0} \right\rangle d_\theta f_\infty'(\cdot|\omega).
\]

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

\[
\int_{T \times \mathbb{R}} \left\langle e^{i\theta}, \nabla_{\theta} \psi_s^{s=0} \right\rangle d_{(\theta,\omega)} f_\infty'=0. 
\]

Indeed, using the equations for \( h_s^{s=0} \) and \( \phi_s^{s=0} \) in (3.17), it is clear that the above implies

(3.17) \[
\int_{T \times \mathbb{R}} \left\langle e^{i\theta}, \nabla_{\theta} \psi_s^{s=0} \right\rangle d_{(\theta,\omega)} h_s^{s=0}=0,
\]

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

\[
\frac{\partial f_\infty}{\partial t} + \text{div}_\theta ((\omega - K R_\infty \sin(\theta - \phi_\infty)) e^{i\theta} f_\infty) = 0,
\]

\[
\frac{\partial f_\infty'}{\partial t} + \text{div}_\theta ((\omega - K R_\infty' \sin(\theta - \phi_\infty')) e^{i\theta} f_\infty') = 0.
\]

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

\[
0 = \frac{d}{dt} \frac{1}{2} W_2(f_\infty(\cdot|\omega), f_\infty'(\cdot|\omega))^2 = \int_T \left\langle (\omega - K R_\infty \sin(\theta - \phi_\infty)) e^{i\theta}, \nabla_{\theta} \psi_s^{s=0} \right\rangle d_\theta f_\infty'(\cdot|\omega)
\]

\[
- \int_T \left\langle (\omega - K R_\infty \sin(\theta - \phi_\infty')) e^{i\theta}, \nabla_{\theta} \psi_s^{s=0} \right\rangle d_\theta f_\infty(\cdot|\omega),
\]

Here on we shall omit the superscripts \( s = 0 \) of \( h_s^{s=0} \) and \( \psi_s^{s=0} \) for simplicity, as it is
clear from the context. Then, integrating against \( g \) and using the fundamental theorem of
calculus in \( \tau \) yields

(3.18) \[
\int_0^1 \frac{d}{d\tau} \int_{T \times \mathbb{R}} \left\langle (\omega - K R_\tau \sin(\theta - \phi_\tau)) e^{i\theta}, \nabla_{\theta} \phi_\tau \right\rangle d_{(\theta,\omega)} h_\tau d\tau = 0,
\]

where \( R_\tau \) and \( \phi_\tau \) are order parameters associated with the displacement interpolation \( h_\tau \).

Let us now expand the derivative in (3.18) and use the Hamilton–Jacobi equation for \( \psi_\tau \)
and the continuity equation for \( h_\tau \) in (3.7) (see [43, Chapter 13]). Then we obtain that

\[
A + B + C = 0,
\]
where each term reads
\begin{align*}
A &:= \int_0^1 \int_{\mathbb{R}} \left( \nabla_{\theta} \left( -\frac{1}{2} |\nabla_{\theta}\psi_{\tau}|^2 \right), (\omega - KR_{\tau} \sin(\theta - \phi_{\tau})e^{i\theta}) \right) d(\theta, \omega) h_{\tau} d\tau, \\
B &:= \int_0^1 \int_{\mathbb{R}} \frac{d}{d\tau} (\omega - KR_{\tau} \sin(\theta - \phi_{\tau})) e^{i\theta}, \nabla_{\theta}\psi_{\tau} \right) d(\theta, \omega) h_{\tau} d\tau, \\
C &:= \int_0^1 \int_{\mathbb{R}} \left( \nabla_{\theta} \left( \nabla_{\theta}\psi_{\tau}, (\omega - KR_{\tau} \sin(\theta - \phi_{\tau}))e^{i\theta} \right), \nabla_{\theta}\psi_{\tau} \right) d(\theta, \omega) h_{\tau} d\tau.
\end{align*}

On the one hand, taking the sum of $A$ and $C$ we can simplify into
\begin{equation}
A + C = -K \int_0^1 \int_{\mathbb{R}} R_{\tau} \cos(\theta - \phi_{\tau}) |\nabla_{\theta}\psi_{\tau}|^2 d(\theta, \omega) h_{\tau} d\tau
\end{equation}
\begin{equation}
= -K \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \cos(\theta - \theta') |\nabla_{\theta}\psi_{\tau}|^2 d(\theta, \omega) h_{\tau} d(\theta', \omega') h_{\tau} d\tau
d(\theta', \omega') h_{\tau} d\tau
\end{equation}

where in the second line we have used the properties of the order parameters $R_{\tau}$ and $\phi_{\tau}$ of the interpolation $h_{\tau}$, namely
\begin{align*}
R_{\tau} &= \int_{\mathbb{R}} \cos(\theta' - \phi_{\tau}) d(\theta', \omega') h_{\tau}, \\
0 &= \int_{\mathbb{R}} \sin(\theta' - \phi_{\tau}) d(\theta', \omega') h_{\tau}.
\end{align*}

and in the third line we have used a clear symmetrization argument. Let us now differentiate with respect to $\tau$ and use the continuity equation for $h_{\tau}$ to obtain the formulas
\begin{align*}
\frac{dR_{\tau}}{d\tau} &= -\int_{\mathbb{R}} \sin(\theta' - \phi_{\tau}) \left( e^{i\theta'}, \nabla_{\theta}\psi_{\tau}(\theta', \omega') \right) d(\theta', \omega') h_{\tau}, \\
R_{\tau} \frac{d\phi_{\tau}}{d\tau} &= \int_{\mathbb{R}} \cos(\theta' - \phi_{\tau}) \left( e^{i\theta'}, \nabla_{\theta}\psi_{\tau}(\theta', \omega') \right) d(\theta', \omega') h_{\tau}.
\end{align*}

Then, the term $B$ can be written as follows
\begin{align}
B &= \int_0^1 \int_{\mathbb{R}} \left( e^{i\theta}, \nabla_{\theta}\psi_{\tau} \right) \frac{d}{d\tau} \left( -K \frac{dR_{\tau}}{d\tau} \sin(\theta - \phi_{\tau}) + KR_{\tau} \frac{d\phi_{\tau}}{d\tau} \cos(\theta - \phi_{\tau}) \right) d(\theta, \omega) h_{\tau} d\tau
\end{align}
\begin{align}
= K \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \cos(\theta - \theta') \left( e^{i\theta}, \nabla_{\theta}\psi_{\tau}(\theta, \omega) \right) \left( e^{i\theta'}, \nabla_{\theta}\psi_{\tau}(\theta', \omega') \right) d(\theta, \omega) h_{\tau} d(\theta', \omega') h_{\tau} d\tau
d(\theta', \omega') h_{\tau} d\tau
\end{align}

Putting the formulas (3.19) and (3.20) into (3.18) entails
\begin{align}
0 &= -\frac{K}{2} \int_0^1 \int_{\mathbb{R}} \int_{\mathbb{R}} \cos(\theta - \theta') \left( \left( e^{i\theta}, \nabla_{\theta}\psi_{\tau}(\theta, \omega) \right) - \left( e^{i\theta'}, \nabla_{\theta}\psi_{\tau}(\theta', \omega') \right) \right)^2 \\
&\quad \times d(\theta, \omega) h_{\tau} d(\theta', \omega') h_{\tau} d\tau.
\end{align}

Since there exists $0 < \delta < \pi/2$ such that
$$\text{diam}(\text{supp}_{\theta} f_{\infty}) < \delta \quad \text{and} \quad \text{diam}(\text{supp}_{\theta} f'_{\infty}) < \delta.$$
The same is true for the interpolations \( h_\tau \) and, consequently. 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 that

\[
0 \leq -\frac{K}{2} \cos \delta \int_0^1 \int_{\mathbb{T} \times \mathbb{R}} \left( e^{i\theta} \nabla_{\theta} \psi_\tau(\theta, \omega) - e^{i\theta'} \nabla_{\theta} \psi_\tau(\theta', \omega') \right)^2 d(\theta, \omega) h_\tau \ d(\theta, \omega) h_\tau \ d\tau
\]

\[
= -K \cos \delta \int_0^1 \int_{\mathbb{T} \times \mathbb{R}} |\nabla_{\theta} \psi_\tau|^2 d(\theta, \omega) h_\tau \ d\tau + K \cos \delta \int_0^1 \left( \int_{\mathbb{T} \times \mathbb{R}} \left( e^{i\theta} \nabla_{\theta} \psi_\tau \right) d(\theta, \omega) h_\tau \right)^2 \ d\tau.
\]

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

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

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

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

thus ending the proof.

We now come back to the proof of Theorem 1.1. First, we show that once the concentration regime in Corollary 2.1 takes place, Theorem 3.1 guaranteed 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., (1.13)). Consider the unique global-in-time classical solution \( f = f(t, \theta, \omega) \) to (1.2). Then, the following holds true

\[
\frac{dI[f]}{dt} \leq -2K \cos(\beta) I[f] + 24K(W + K)^2 \rho_1(\mathbb{T} \setminus L_{\beta}^0(t)),
\]

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 follows

\[
\frac{dI[f]}{dt} = I_1 + I_2,
\]

where each factor reads

\[
I_1 = -K \int_{L_{\beta}^0(t) \times L_{\beta}^0(t) \times \mathbb{R} \times \mathbb{R}} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times \cos(\theta - \theta') f(t, \theta, \omega) f(t, \theta', \omega') \ d\theta \ d\theta' \ d\omega \ d\omega',
\]

\[
I_2 = -K \int_{(\mathbb{T} \times \mathbb{T}) \setminus (L_{\beta}^0(t) \times L_{\beta}^0(t))} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times \cos(\theta - \theta') f(t, \theta, \omega) f(t, \theta', \omega') \ d\theta \ d\theta' \ d\omega \ d\omega'.
\]
On the one hand, it is clear that
\begin{equation}
(3.22)
I_1 \leq -K \cos(\beta) \int_{L^+_\beta(t) \times L^+_\beta(t) \times \mathbb{R} \times \mathbb{R}} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega' \\
= -K \cos(\beta) \int_{\mathbb{T}^2} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega' \\
\quad + K \cos(\beta) \int \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega' \\
=: I_{11} + I_{12},
\end{equation}
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 (1.13) we get
\begin{align*}
I_{11} &= -K \cos(\beta) \int_{\mathbb{T}^2} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \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) I[f].
\end{align*}
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
\begin{align*}
I_3 \leq 2K \int \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) \right)^2 \times f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega'.
\end{align*}
In other words, we achieved the estimate
\begin{equation}
(3.23)
\frac{dI[f]}{dt} \leq -2K \cos(\beta) I[f] + I_3.
\end{equation}
Our last goal is to estimate the remainder $I_3$. Define the following time-dependent sets
\begin{align*}
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{align*}
Since we have that $((\mathbb{T} \times \mathbb{T}) \setminus (L^+_\beta(t) \times L^+_\beta(t))) \times \mathbb{R} \times \mathbb{R} = A_1 \cup A_2 \cup A_3$, then we can split $I_3$ as follows
\begin{align*}
I_3 \leq I_{31} + I_{32} + I_{33},
\end{align*}
where each integral takes the following form
\begin{align*}
I_{3i} &= 2K \int_{A_i} \left( (\omega - KR \sin(\theta - \phi) - (\omega' - KR \sin(\theta' - \phi))) f(t, \theta, \omega) f(t, \theta', \omega') \, d\theta \, d\theta' \, d\omega \, d\omega',
\end{align*}
for every $i = 1, 2$. 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 follows

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

Then, we obtain

$$I_{31}(t) \leq 8K(W + K)^2 \rho_t(T \setminus L_t^+(t)),$$

for every $t \geq 0$. Exactly the same argument allows estimating $I_{33}$ and obtaining an identical bound. Putting everything together into (3.23) finishes the proof.

Now, we can apply Gronwall’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., (1.13)). 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

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

then there exists a time $T_0$ with the property that

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

and

$$I[f_t] \lesssim K^2 e^{-\frac{1}{30} K(t-T_0)}.$$

**Proof.** Let us adjust $C$ small enough so that we meet the hypotheses of Corollary 2.1. Then, there exists such a time $T_0$ so that

$$\rho_t(T \setminus L_t^+(t)) \leq Me^{-\frac{1}{36} K(t-T_0)},$$

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

$$\frac{d}{dt} I[f] \leq -2K \cos(\beta) I[f] + 24K(W + K)^2 Me^{-\frac{1}{36} K(t-T_0)},$$

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

$$I[f_t] \leq I[f_{T_0}] e^{-2K \cos(\beta)(t-T_0)} + \frac{24K(W + K)^2 M}{2K \cos(\beta) - \frac{1}{36} K} \left(e^{-\frac{1}{36} K(t-T_0)} - e^{-2K \cos(\beta)(t-T_0)}\right),$$

$$\lesssim (W + K)^2 e^{-\frac{K}{36} (t-T_0)} \lesssim K^2 e^{-\frac{K}{36}(t-T_0)},$$

where in the second inequality we have used that

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

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

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,

$$W_{2, g}(f_t, f_s) \lesssim e^{-\frac{K}{36} (t-T_0)} - e^{-\frac{K}{36} (s-T_0)},$$

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 \( (P_g(\mathbb{T} \times \mathbb{R}), W^2, g) \). Notice that it is a complete metric space. Consequently, there exists some probability measure \( f_\infty \in P_g(\mathbb{T} \times \mathbb{R}) \) 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

\[
W^2, g(f_t, f_\infty) \lesssim e^{-\frac{K}{40}(t-T_0)},
\]

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, \( f_\infty \) is an equilibrium of the Kuramoto-Sakaguchi equation (1.2) and the asymptotic concentration estimate in Corollary 2.1 guarantees that

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

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

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

The main objective of this part is the proof of the entropy production estimate Lemma 2.1. As a byproduct in Corollary 3.6 we will obtain a universal lower bound on the order parameter. Before we begin the proof of the entropy production estimate, we will need a relationship between the time derivative of the order parameter and the dissipation (2.10). That is the content of the following lemma.

**Lemma 3.2.** Assume that \( f_0 \) is contained in \( C^1(\mathbb{T} \times \mathbb{R}) \) and that \( g \) is compactly supported in \([-W,W]\). Then, the inequality

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

holds.

**Proof.** By (2.2) we have that

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

Consequently, by young’s inequality, we obtain that

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

and

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

Hence, the desired result follows. □
Now we are ready to prove our entropy production estimate.

**Proof of Lemma 2.1**

Without loss of generality, we can assume that

\[ R < \frac{3}{2} R_0 \text{ in } [t_0, t_0 + \frac{1}{3K R_0} \log 10]. \]  

Otherwise, if this condition fails for some \( s \) in the above interval, then we set \( d = s - t_0 \) and (2.5) would follow. Thanks to the inequalities (3.25) and (3.26), we arrive at the following estimate

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

Let us integrate the above inequality on the interval \([t_0, t_0 + d]\) for some \( d \) in \([0, \frac{1}{3K R_0} \log 10]\), which we will choose appropriately after the calculations below. By doing this and using (3.26), we deduce that

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

Thus, by choosing \( d = \frac{1}{3K R_0} \log 10 \), we obtain that

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

Consequently, by selecting \( C \) appropriately in (2.5) we conclude that

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

Hence, since \( \alpha = \pi/6 \) the desired result follows. \( \square \).

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 3.3.** Assume that \( f_0 \) is contained in \( C^1(T \times \mathbb{R}) \) 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 that

\[
|\dot{\phi}| \leq \frac{1}{R} \sqrt{K \frac{d}{dt} R^2 + W^2}. 
\]
Proof. By (2.2), and Jensen inequality, we have that
\[
R|\dot{\phi}| \leq \int |\cos(\theta - \phi)(\omega - KR \sin(\theta - \phi))| f \, d\theta d\omega \\
\leq \int |(\omega - KR \sin(\theta - \phi))| f \, d\theta d\omega \\
\leq \left( \int |(\omega - KR \sin(\theta - \phi))|^2 f \, d\theta d\omega \right)^{\frac{1}{2}} \\
= I^{\frac{1}{2}} \\
\leq \sqrt{K \frac{d}{dt} R^2 + W^2},
\]
where in the last inequality, we have used (3.25). Thus, the desired result follows. □

We will derive a global lower bound on the order parameter as an application of the entropy production estimate (2.1). To achieve this, we consider the following lemma, which controls the rate at which the order parameter can decrease.

Lemma 3.4. (Rate of decrease and mass monotonicity) Let $\lambda$ be contained in $(2/3, 1)$, assume that $f_0$ is contained $C^1(T \times \mathbb{R})$ 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 $\gamma$ be a positive number in $(\pi/6, \pi/2)$, and let $\alpha$ be as specified in Section 2. Then, we have that
\[
\frac{d}{dt} R^2 \geq \frac{KR^2 \cos^2 \gamma}{2} \left( 1 - \frac{2W^2}{K^2 R^2 \cos^2 \gamma} - \frac{R}{\sin \gamma} - \frac{1 + \sin \gamma}{\sin \gamma} f(\chi^-) \right),
\]
and
\[
\frac{d}{dt} f(\chi^-) \leq 4K \left[ \frac{W}{K} + \sqrt{\frac{2R}{KR} + \frac{1}{R^2} \frac{W^2}{K^2} - R \cos \alpha} \right]^+,
\]
for all $t \geq 0$.

Moreover, suppose that $\dot{R}(t_0) \leq 0$, $R(t_0) \geq R_0$, 
\[
\dot{R} \leq \frac{K \cos^2 \alpha \lambda^3 R_0^3}{4} \quad \text{in} \quad [t_0, t_0 + d] \quad \text{and} \quad \cos^2 \gamma = \frac{1 - \lambda}{5} R_0,
\]
for some non-negative numbers $d$ and $t_0$. Then, there exist a universal constant $C$ such that if
\[
W \leq C(1 - \lambda) \lambda^2 R_0^2,
\]
then,
\[
\frac{d}{dt} R^2 \geq \frac{K \cos^2 \gamma}{2 \sin \gamma} \left( -R^3 + \left[ \lambda R_0 + \frac{3}{5}(1 - \lambda) R_0 \right] R^2 - \frac{3}{5}(1 - \lambda) \lambda^2 R_0^3 \right),
\]
in $[t_0, t_0 + d]$. Consequently,
\[
R \geq \lambda R_0 \quad \text{in} \quad [t_0, t_0 + d].
\]
Proof. We divide the proof into the following steps:

**Step 1:** Derivation of estimate (3.28).

Recall that \( \chi_{\alpha}(\theta) = \xi_{\alpha}(\theta - \phi - \pi) \), with \( \xi_{\alpha} \) as defined in (2.7). Then, by direct computation, we have that

\[
\frac{d}{dt}f(\chi_{\alpha}) = \frac{d}{dt}\int_{\mathbb{T}_R} \xi_{\alpha}(\theta - \phi - \pi) f \, d\theta d\omega
\]

\[
= \int_{\mathbb{T}_R} \xi_{\alpha}'(\theta - \phi - \pi)(\omega - KR\sin(\theta - \phi) - \dot{\phi}) f \, d\theta d\omega
\]

\[
\leq f(|\xi_{\alpha}'|)[W + |\dot{\phi}|] + KR\int_{\mathbb{T}_R} \xi_{\alpha}'(\theta - \phi - \pi) \sin(\theta - \phi - \pi)f \, d\theta d\omega
\]

\[
\leq f(|\xi_{\alpha}'|)[W + |\dot{\phi}| - KR\cos\alpha]
\]

\[
\leq f(|\xi_{\alpha}'|) \left[ W + \frac{1}{R} \sqrt{2KR\frac{d}{dt}R + W^2 - KR\cos\alpha} \right].
\]

Notice that in the last inequality we have used Lemma 3.3 in order to estimate \(|\dot{\phi}|\) and the only thing that remains to show is the bound of the second term in the third line. Firstly, the support of \(\xi_{\alpha}'(\theta - \phi - \pi)\) consists of \(S^+ \cup S^-\) where each set stands for

\[
S^+ := \left[ \phi + \frac{3\pi}{2} - \alpha, \phi + \frac{3\pi}{2} - \alpha + \frac{1}{2} \right] \quad \text{and} \quad S^- := \left[ \phi + \frac{\pi}{2} + \alpha - \frac{1}{2}, \phi + \frac{\pi}{2} + \alpha \right].
\]

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

\[
\theta \in S^+ \implies \xi_{\alpha}'(\theta - \phi - \pi) \leq 0 \quad \text{and} \quad \sin(\theta - \phi - \pi) \geq \cos\alpha,
\]

\[
\theta \in S^- \implies \xi_{\alpha}'(\theta - \phi - \pi) \geq 0 \quad \text{and} \quad \sin(\theta - \phi - \pi) \leq -\cos\alpha.
\]

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, (3.28) follows.

**Step 2:** Derivation of estimate (3.27).

By the first equation in (2.2), we obtain the following lower bound on \(\dot{R}\)

\[
\frac{K}{2} \frac{d}{dt} R^2 = -\int_{\mathbb{T}_R} KR\sin(\theta - \phi)(\omega - KR\sin(\theta - \phi)) f \, d\theta d\omega
\]

\[
\geq \int_{\mathbb{T}_R} (KR\sin(\theta - \phi))^2 f \, d\theta d\omega - \int \omega(KR\sin(\theta - \phi)) f \, d\theta d\omega
\]

\[
\geq \frac{1}{2} \int_{\mathbb{T}_R} (KR\sin(\theta - \phi))^2 f \, d\theta d\omega - \frac{W^2}{2}
\]

\[
\geq \frac{1}{2} K^2 R^2 \cos^2 \gamma f(\mathbb{T} \setminus (L^+_\gamma(t) \cup L^-_\gamma(t))) - \frac{W^2}{2}.
\]

Then, we obtain

\[
(3.31) \quad f(\mathbb{T} \setminus (L^+_\gamma(t) \cup L^-_\gamma(t))) \leq \frac{1}{KR^2 \cos^2 \gamma} \frac{d}{dt} R^2 + \frac{W^2}{K^2 R^2 \cos^2 \gamma}.
\]
Additionally, using a similar argument on (1.12), where we split the integral into the sectors \( L^+_\gamma, L^-_\gamma \) and \( \mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma) \), allows getting the lower bound
\[
R \geq \sin \gamma f(L^+_\gamma) - \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - f(L^-_\gamma)
\]
\[
= \sin \gamma \left(1 - f(L^-_\gamma) - f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma))\right) - \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - f(L^-_\gamma)
\]
\[
= \sin \gamma - 2 \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - (1 + \sin \gamma)f(L^-_\gamma)
\]
\[
\geq \sin \gamma - 2 \sin \gamma \left(\frac{1}{KR^2 \cos^2 \gamma} \frac{d}{dt} R^2 + \frac{W^2}{K^2 R^2 \cos^2 \gamma}\right) - (1 + \sin \gamma)f(L^-_\gamma).
\]
Here, we have used the estimate (3.31) in the last inequality. Then, (3.27) follows.

- **Step 3: Upper bound on** \( f(L^-_\gamma) \).

Let us first achieve a lower bound of \( f(L^+_\gamma) \). To such end, we use a similar procedure and reverse the inequalities that we have considered in the preceding step. Specifically, notice that a similar split in (1.12) allows obtaining
\[
R \leq f(L^+_\gamma) + \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - \sin \gamma f(L^-_\gamma)
\]
\[
= f(L^+_\gamma) + \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - \sin \gamma (1 - f(L^+_\gamma) - f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)))
\]
\[
= (1 + \sin \gamma)f(L^+_\gamma) + 2 \sin \gamma f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma)) - \sin \gamma.
\]

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
\[
f(L^-_\gamma) = 1 - f(L^+_\gamma) - f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma))
\]
\[
\leq 1 - \frac{\sin \gamma}{1 + \sin \gamma} - \frac{R}{1 + \sin \gamma} - \frac{1 - \sin \gamma}{1 + \sin \gamma} f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma))
\]
\[
\leq \frac{1}{1 + \sin \gamma} - \frac{R}{1 + \sin \gamma}.
\]

Notice that since \( \dot{R}(t_0) \leq 0 \) we can select \( C \) appropriately in (3.29) to guarantee that
\[
\frac{W}{K} + \sqrt{\frac{2 \dot{R}(t_0)}{K R(t_0)} + \frac{1}{R(t_0)^2} \frac{W^2}{K^2} - R(t_0) \cos \alpha} \leq \frac{W}{K} + \sqrt{\frac{1}{R_0^2} \frac{W^2}{K^2} - \lambda R_0 \cos \alpha} < 0.
\]

Then, estimate (3.28) implies
\[
\left. \frac{d}{dt} \right|_{t=t_0} f(\chi^-_\alpha)(t) \leq 0.
\]

By continuity, and, inequalities (3.28) and (3.33), \( f(\chi^-_\alpha)(t) \) remains non increasing along \([t_0, t_0 + \delta]\) for small enough \( \delta > 0 \). Hence, we obtain that
\[
f(L^-_\gamma)(t) \leq f(\chi^-_\alpha)(t)
\]
\[
\leq f(\chi^-_\alpha)(t_0)
\]
\[
\leq f(L^-_\gamma)(t_0) + f(\mathbb{T} \setminus (L^+_\gamma \cup L^-_\gamma))(t_0)
\]
\[
\leq \frac{1}{1 + \sin \gamma} - \frac{R(t_0)}{1 + \sin \gamma} + \frac{W^2}{K^2 \cos^2 \gamma R^2(t_0)}.
\]
for all $t$ in $[t_0, t_0 + \delta]$. Here, we have used the estimates (3.31) and (3.32) along with the hypothesis $\dot{R}(t_0) \leq 0$.

- **Step 4**: Derivation of (3.30) and lower bound of $R$ in $[t_0, t_0 + \delta]$.

Putting the last estimate (3.34) and (3.27) together, we obtain the differential inequality

$$
\frac{dR^2}{dt} \geq \frac{KR^2 \cos^2 \gamma}{2 \sin \gamma} \left[ R(t_0) - R - (1 - \sin \gamma) - \frac{2 \sin \gamma W^2}{K^2 \cos^2 \gamma R^2} - \frac{1 + \sin \gamma W^2}{K^2 \cos^2 \gamma R^2(t_0)} \right]
$$

(3.35)

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

$$
b(t_0) := R(t_0) - \cos^2 \gamma - \frac{2W^2}{K^2 \cos^2 \gamma R^2(t_0)},
$$

$$
c(t_0) := \frac{2W^2}{K^2 \cos^2 \gamma}.
$$

Notice that in the last inequality in (3.35) 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 (3.29) we can guarantee that

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

Arguing in a similar way and making $C$ smaller if necessary in (3.29), we can guarantee that

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

Consequently, we have that

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

in $[t_0, t_0 + \delta]$. Since $\lambda R_0$, is the biggest root of the polynomial

$$
p(r) = -r^3 + \left[ \lambda R_0 + \frac{3}{5}(1 - \lambda) R_0 \right] r^2 - \frac{3}{5}(1 - \lambda) \lambda^2 R_0^3,
$$
we obtain desire lower bound $R \geq \lambda R_0$ in $[t_0, t_0 + \delta]$ by an elementary continuity method argument (we can see that $\lambda R_0$ is the biggest root of $p$ from the inequality $p(0) < 0$ and the fact that $\lambda$ being contained in $(2/3, 1)$ implies that $p'(\lambda R_0) < 0$).

Step 5: Propagation of (3.30) and the lower bound of $R$ in $[t_0, t_0 + d]$.

The main idea is supported by a continuity method. We proceed by contradiction. Specifically, define the time

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

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 in Step 4, we have that

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

By (3.28) and the fact that $\dot{R} \leq \frac{K \cos^2 \alpha \lambda^3 R_0}{4}$ in $[t_0, t_0 + d]$, making $C$ smaller in (3.29) if necessary, we can guarantee that,

$$\begin{align*}
\frac{W}{K} + \sqrt{\frac{2R(t)}{K R(t)}} + \frac{1}{R(t)^2} \frac{W^2}{K^2} - R(t) \cos \alpha &\leq \frac{W}{K} \\
&+ \sqrt{\frac{\lambda^2 R_0^2 \cos^2 \alpha}{2} + \frac{1}{\lambda^2 R_0^2} \frac{W^2}{K^2}} - \lambda R_0 \cos \alpha < 0,
\end{align*}$$

for all $t \in [t_0, t_*]$. In particular, by (3.28) and continuity we have that $f(\chi^-)$ is non increasing in $[t_0, t_* + \delta_*]$ and some small enough $\delta_* > 0$. Hence, we can repeat the train of thoughts in Step 4 to extend the upper bound of $f(\chi^-)(t)$ in (3.34) 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 $t_*$. \qed

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

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

then, we have that

$$R \geq \lambda R_0,$$

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 Lemma 2.1 and 3.4. We claim that either one of the following two conditions holds:

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

(ii) There exist a time $t^*$ and an increasing and strictly positive universal function $h$, satisfying that $R \geq \lambda R_0$ 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 times 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, increasing, and universal 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 3.4.

• Step 2: We show the claim.

For this purpose suppose that (i) does not hold, that is the set

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

is not empty. To show that (ii) holds in this case, let us consider the smallest time $t_1$ such that $\dot{R}(t_1) \geq \frac{K}{4}\lambda^3 R_0^3 \cos^2 \alpha$. Now, let $t_2$ denote the biggest time $t_2$, bigger or equal to $t_1$, such that $\dot{R} \geq \frac{K}{4}\lambda^3 R_0^3 \cos^2 \alpha$ in $[t_1, t_2]$. Notice that the existence of $t_2$ follows by the boundedness of $R$.

Now, observe that, by definition of $t_1$ Lemma 3.3 implies that $R \geq \lambda R_0$ in $[0, t_1]$. Moreover, by construction

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

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 2.1 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 \( \lambda \), we have that
\[
R(t_2 + d)^2 = R(t_2)^2 + \frac{\lambda^4}{40}R_0^3
\geq \lambda^2 R_0^2 + \frac{\lambda^4}{40}R_0^3
\geq R_0^2 + \frac{\lambda^4}{40}R_0^3 - (1 - \lambda^2)R_0^2
> R_0^2 + \left(\frac{5}{240}R_0 - 2(1 - \lambda)\right)R_0^2
> R_0^2 + \frac{1}{240}R_0^3.
\]
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 that \( R(t_2)^2 - R_0^2 > R_0^2/240 \). Hence, the desired result holds for \( t^* = t_2 \).

\[4. \quad \text{Instability of antipodal equilibria and Sliding norms}\]

We now start implementing the program outlined in sections 2.3 and 2.4. To do this, we first derive inequalities (2.8) and (2.11).

**Proposition 4.1. (Instability of antipodal equilibria)** Assume that \( f_0 \) is contained in \( C^1(T \times \mathbb{R}) \) 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 \( \alpha \) be as specified in Section 2. Then, we have that
\[
\frac{d}{dt}f^2(\chi^-_\alpha(t)) \leq -K R \sin \alpha f^2(\chi^-_\alpha(t)) + 4K f^2(\mathbb{T}) \left[\frac{W}{K} + \sqrt{\frac{2R}{KR} + \frac{W^2}{R^2 K^2}} - R \cos \alpha \right]^+,
\]
and
\[
(4.1) \quad \frac{d}{dt}f^2(\mathbb{T}) \leq KRf^2(\mathbb{T}).
\]
Moreover, with the hypothesis (3.29) and notation from Proposition 3.4, if \([t_1, t_2]\) is a time interval such that
\[
(4.2) \quad \hat{R} \leq K \frac{\lambda^3 R_0^3 \cos^2 \alpha}{4} \quad \text{in} \quad [t_1, t_2],
\]
then, we have that
\[
(4.3) \quad \frac{d}{dt}f^2(L^-_\alpha(t)) \leq -K \lambda R_0 \sin \alpha f^2(L^-_\alpha(t)) \quad \text{in} \quad [t_1, t_2].
\]
Proof. We begin with the first inequality in the Proposition. Arguing as in Step 1 of the proof of Lemma 3.4 we obtain that

\[
\frac{d}{dt} \int \chi_\alpha^-(\theta - \phi + \pi) f^2 \, d\theta d\omega = \int \dot{\phi} \chi_\alpha^-(\theta - \phi + \pi) f^2 \, d\theta d\omega + 2 \int \chi_\alpha^-(\theta - \phi + \pi) f \partial_t f \, d\theta d\omega
\]

(4.4)

The first inequality in the proposition follows from Lemma 3.3 and the same arguments as in Step 1 from Proposition 3.4. Inequality (4.1) follows from similar arguments to those of (4.4) by replacing \(\chi_\alpha\) with the constant function that is equal to one in \(T\). Finally, to derive inequality (4.3), recalling the notation introduced in Section 2.3, replacing \(\chi_\alpha^-\) with \(\chi_{\alpha,\varepsilon}^-\) in (4.4) and arguing as in Step 1 from Proposition 3.4 we get that

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

Now, we observe that as in (3.36), we can see that the second term of the above inequality vanishes on the interval \([t_1, t_2]\). Consequently, such a term is independent of \(\varepsilon\) and thus (4.3) follows by letting \(\varepsilon \to 0\).

\[\square\]

A form of the above Lemma was one of the main tools used to derive the main result in [22]. 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 analog 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 that

\[
\frac{d}{dt} f^2(A_{t_0}, t) \leq KR \left( \sup_{(\theta, \omega) \in A_{t_0}, t} \cos(\theta - \phi(t)) \right) f^2(A_{t_0}, t).
\]
Proof. By the change of variable theorem, we have that
\[
\frac{d}{dt} \frac{1}{2} \int_{A_{t_0,t}} f^2 \, d\theta d\omega = \frac{d}{dt} \left|_{t=t_0} \right. \frac{1}{2} \int_A f_t^2(\Theta_{t_0,t}(\theta,\omega),\omega) \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
= \int_A f_t(\Theta_{t_0,t}(\theta,\omega)) \partial_\theta f(\Theta_{t_0,t}(\theta,\omega),\omega) \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
+ \partial_\theta f(\Theta_{t_0,t}(\theta,\omega)) \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
- \frac{1}{2} K R \int_A \cos(\Theta_t(\theta,\omega) - \phi) \partial_\theta \Theta_{t_0,t} f^2 \, d\theta d\omega
\]
\[
= \int_A f_t(\Theta_{t_0,t}(\theta,\omega)) \left[ - \partial_\theta (\omega f - K R \sin(\Theta_{t_0,t}(\theta,\omega) - \phi)) \right] \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
+ (\omega - K R \sin(\Theta_{t_0,t}(\theta,\omega) - \phi)) \partial_\theta f(\Theta_{t_0,t}(\theta,\omega),\omega) \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
+ \frac{1}{2} K R \int_A \cos(\Theta_{t_0,t}(\theta,\omega) - \phi) f^2 \partial_\theta \Theta_{t_0,t} \, d\theta d\omega
\]
\[
= \frac{1}{2} K R \int_A \cos(\Theta_{t_0,t}(\theta,\omega) - \phi) f_t^2(\theta,\omega) \partial_\theta \Theta_{t_0,t} \, d\theta d\omega.
\]
where for \( t \) and each \( \omega \), \( \partial_\theta \Theta_{t_0,t}(\cdot,\omega) \) denotes the Jacobian of the map \( \theta \to \Theta_{t_0,t}(\theta,\omega) \). Hence, the desired result follows. \( \square \)

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. **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 2.3.

**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 to
\[
\begin{align*}
&f = f(t, \theta, \omega). \\
&T_{t_0} \geq 0 \text{ be an initial time in } [0, \infty) \text{ and } L \subset \mathbb{T} \text{ be an interval.} \\
&\text{Now, assume that, initially we have that} \\
&\rho_{t_0}(L) \geq m, \quad \text{and} \quad p = \inf_{\theta, \theta' \in L} \cos(\theta - \theta'),
\end{align*}
\]
for some positive numbers \( m \) and \( p \) in \((0,1)\). Additionally, suppose that
\[
mp - (1 - m) \geq \sigma \quad \text{and} \quad \frac{W^2}{K^2} \leq \frac{(1 - p)\sigma^2}{4},
\]
for some \( \sigma > 0 \).
Then, if we set
\[
\begin{align*}
\mathcal{P}(t) &= \inf_{\theta, \theta' \in L_{t_0,t}} \cos(\theta - \theta'), \\
\rho(L_{t_0,t}) &= m,
\end{align*}
\]
the following bounds hold true
\[
\text{for some } \sigma > 0.
\]
We will first prove inequality (4.8) and then prove the remaining inequalities afterward. Indeed, let us define

\[ L \]

(4.8) holds, for every \( t \) such that (4.5) implies the uniform lower bound

\[ 1 - P(t) \leq \max \left( (1 - p)e^{-\frac{\kappa s}{4}(t-t_0)}, \frac{4}{\sigma^2} \frac{W^2}{K^2} \right), \]

for every \( t \) in \([t_0, \infty[\).

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

\[ \frac{dP}{dt} \geq 2K \sqrt{1-P^2} \left[ R \left( \inf_{\theta \in L_{t_0,t}} \cos(\theta - \phi) \right) \sqrt{\frac{1-P}{2}} - \frac{W}{K} \right], \]

which hold when

\[ P = \cos(\Theta_{s,t}(\omega,\theta) - \Theta_{s,t}(\omega',\theta')), \]

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

We will first prove inequality (4.8) and then prove the remaining inequalities afterward. 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

\[ 1 - P(t') = \max \left( (1 - p)e^{-\frac{\kappa s}{4}(t'-t_0)}, \frac{4}{\sigma^2} \frac{W^2}{K^2} \right). \]

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 that

\[ \frac{dP}{dt} \bigg|_{t=t'} \geq 2K \sqrt{1-P^2} \left[ R \left( \inf_{\theta \in L_{t_0,t}} \cos(\theta - \phi) \right) \sqrt{\frac{1-P}{2}} - \frac{W}{K} \right] \]

\[ \geq 2K \sqrt{1-P^2} \left[ mP - (1-m) \right] \sqrt{\frac{1-P}{2}} - \frac{W}{K} \]

\[ \geq 2K \sqrt{1-P} \left[ \frac{\sqrt{2}}{2} \sigma(1-P) - \frac{W}{K} \sqrt{1-P} \right]. \]

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 \( p \leq P \). Now, let \((\theta,\omega)\) and \((\theta',\omega')\) be any couple of points contained in \( L_{t_0,t'} \times [-W,W] \) such that the corresponding \( P \) satisfies that

\[ 1 - P(t') = 1 - P(t') = \max \left( (1 - p)e^{-\frac{\kappa s}{4}(t'-t_0)}, \frac{4}{\sigma^2} \frac{W^2}{K^2} \right). \]

Note that since \( L \) is compact, then \( L_{t_0,t'} \times [-W,W] \) is compact as well. Thus, the set of such pairs \((\theta,\omega)\) and \((\theta',\omega')\) in \( L_{t_0,t'} \) whose corresponding \( P \) (obtained via (4.10)) satisfies (4.12)
is a compact set as well. We shall denote such a set by \( \mathcal{P} \subset L_{l_0,t} \times [-W,W] \times L_{l_0,t'} \times [-W,W] \).

To continue our proof observe that by using the assumption (4.12), we get that
\[
\frac{\sigma \sqrt{1 - P(t')}}{2} \geq \frac{W}{K},
\]
for any couple of characteristics in \( \mathcal{P} \) and, consequently, by (4.11) we obtain that
\[
\left. \frac{d}{dt} \right|_{t=t'} (1 - P(t')) \leq -2K\sqrt{1 - P(t')} \left[ \frac{\sqrt{2}}{2} \sigma(1 - P(t')) - \frac{\sigma(1 - P(t'))}{2} \right]
\]
\[
\leq -\frac{2}{5} K \sigma((1 - P(t'))).
\]

Since the right-hand side of the above inequality is uniform in the set of pairs in \( \mathcal{P} \) and the set \( \mathcal{P} \) is compact, we can find \( \varepsilon > 0 \) such that if \( \mathcal{P}_\varepsilon \) is an \( \varepsilon \)-neighborhood of \( \mathcal{P} \), then we have that
\[
\left. \frac{d}{dt} \right|_{t=t'} (1 - \cos(\Theta_{t',\iota}(\theta,\omega') - \Theta_{t',\iota}(\theta',\omega'))) < -\frac{1}{3} K \sigma((1 - P(t')))
\]
\[
\leq \left\{ \begin{array}{ll}
-\frac{K}{4} \sigma(1 - p) e^{-\frac{K}{4} (t' - t_0)} & \text{if } \frac{4}{\sigma^2} W^2 < 1 - P(t'), \\
0 & \text{if } \frac{4}{\sigma^2} W^2 = 1 - P(t'),
\end{array} \right.
\]
(4.13)

for any \((\theta,\omega),(\theta',\omega')\) in \( \mathcal{P}_\varepsilon \). This implies the existence of \( \delta \) and thus concludes the continuity method argument. Indeed, for characteristics with initial data in \( \mathcal{P}_\varepsilon \) the existence of the time interval \([t', t' + \delta)\), follows by the fact that the inequality in (4.13) is strict and uniform in \( \mathcal{P}_\varepsilon \). Similarly, for characteristics in \( (L_{l_0,t} \times [-W,W] \times L_{l_0,t'} \times [-W,W])/\mathcal{P}_\varepsilon \), the existence of the uniform time \( \delta \) follows by the fact that the characteristics have uniformly bounded speed and \( \varepsilon \) provides a uniform separation distance.

(Indeed, by continuity and compactness, we can find a uniform time neighborhood of \( t' \), in which the infimum for \( P \) is attained in \( \mathcal{P}_{\varepsilon/2} \), and we have already shown the existence of \( \delta \) 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**: Proof 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
\[
\inf_{\theta \in L_{t_0,t}} R \cos(\theta - \phi) = \inf_{\theta \in L_{t_0,t}} \langle e^{i\theta}, \int e^{i\theta} f d\omega' d\theta' \rangle \\
\geq \inf_{\theta \in L_{t_0,t}} \int \cos(\theta - \theta') f' d\theta' d\omega' \\
\geq \inf_{\theta \in L_{t_0,t}} \left[ \int_{(L \times [-W,W])_{t_0,t}} \cos(\theta - \theta') f' d\theta' d\omega' \\
+ \int_{T \times \mathbb{R} \setminus (L \times [-W,W])_{t_0,t}} \cos(\theta - \theta') f' d\theta' d\omega' \right] \\
\geq m P - (1 - m).
\]
(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 \((\theta, \omega)\) and \((\theta', \omega')\) be contained in \( L \times [-W,W] \). Additionally, let us set \( \Theta(s) := \Theta_{t_0,s}(\theta, \omega) \), and \( \Theta'(s) := \Theta_{t_0,s}(\theta', \omega') \).

Then,
\[
\frac{d}{ds} \bigg|_{s=t} \cos(\Theta - \Theta') = -\sin(\Theta - \Theta')(\dot{\Theta} - \dot{\Theta}') \\
= -\sin(\Theta - \Theta') \left( (\omega - \omega') - KR(\sin(\Theta - \phi) - \sin(\Theta' - \phi)) \right) \\
= -\sin(\Theta - \Theta') \left[ (\omega - \omega') - 2KR \cos \left( \frac{\Theta + \Theta'}{2} - \phi \right) \sin \left( \frac{\Theta - \Theta'}{2} \right) \right] \\
= -2 \cos \left( \frac{\Theta - \Theta'}{2} \right) \left[ (\omega - \omega') \sin \left( \frac{\Theta - \Theta'}{2} \right) \\
- 2KR \cos \left( \frac{\Theta + \Theta'}{2} - \phi \right) \sin^2 \left( \frac{\Theta - \Theta'}{2} \right) \right] \\
\geq 4KR \cos \left( \frac{\Theta - \Theta'}{2} \right) \left[ \cos \left( \frac{\Theta + \Theta'}{2} - \phi \right) \frac{1 - \cos(\Theta - \Theta')}{2} \right. \\
\left. - \frac{W}{KR} \sqrt{1 - \cos(\Theta - \Theta')} \right] ,
\]
(4.15)

where we have used several standard trigonometric formulas. Now, notice that
\[
\frac{\Theta_{t_0,t}(\theta, \omega) + \Theta_{t_0,t}(\theta', \omega')}{2} \text{ is contained in } L_{t_0,t},
\]
since it is a convex combination of two points in \( L_{t_0,t} \). Thus, when \( s = t_0 \) 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 in which the system forms an attractor.

**Lemma 4.3 (First invariant set).** Assume that \( f_0 \) is contained in \( C^1(T \times \mathbb{R}) \) 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
\[ \cos^2 \gamma = \frac{1}{30} R_0. \]

Then, we can find a universal constant \( C \) such that if
\[ \frac{W}{K} \leq C R_0^2, \]
then there exists a positive time \( T_{-1} \) satisfying that
\[ T_{-1} \lesssim \frac{1}{K R_0^3}, \]
and the bounds
\[ \rho(L_+^+(T_{-1})) \geq \frac{1 + \frac{4}{5} R_0}{2}, \]
\[ \inf_{\theta \in L_+^+(T_{-1})} R \cos(\theta - \phi) \geq \frac{3}{5} R_0, \]
and
\[ \inf_{\theta, \theta' \in L_+^+(T_{-1})} \cos(\theta - \theta') \geq 1 - \frac{1}{15} R_0, \]
hold true for every \( t \) in \([T_{-1}, \infty)\).

**Proof.** Define the time
\[ T_{-1} = \inf \left\{ t \geq 0 : \dot{R} \leq \frac{K R_0^3}{4 \cdot 30^2} \right\}, \]
and note that by construction, (4.18) follows by 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 \( L = L_+^+(T_{-1}) \) and \( t_0 = T_{-1} \). To verify the corresponding hypotheses, first, we begin by controlling the mass in \( L_+^+(T_{-1}) \). Indeed, by the decomposition of the integral (1.12) that defines \( R \) into three parts \( L_+^+(T_{-1}), L_-^-(T_{-1}), \) and \( \mathcal{T} \setminus (L_+^+ \cup L_-^-) \), we obtain the inequality:
\[ R \leq (1 + \sin \gamma) \rho(L_+^+) - \sin \gamma + 2 \sin \gamma \rho(\mathcal{T} \setminus (L_+^+ \cup L_-^-)). \]
Consequently, using (3.31) to control \( \rho(\mathcal{T} \setminus (L_+^+ \cup L_-^-)) \), we deduce that
\[ \rho(L_+^+) \geq \frac{R + \sin \gamma}{1 + \sin \gamma} - \frac{2 \sin \gamma}{1 + \sin \gamma} \rho(\mathcal{T} \setminus (L_+^+ \cup L_-^-)) \]
\[ \geq \frac{1}{1 + \sin \gamma} \left[ R + \sin \gamma - 2 \left( \frac{1}{K R^2 \cos^2 \gamma} \frac{d}{dt} R^2 + \frac{W^2}{K^2 R^2 \cos^2 \gamma} \right) \right] \]
\[ = \frac{1}{1 + \sin \gamma} \left[ R + 1 + (\sin \gamma - 1) - 2 \left( \frac{2 \hat{R}}{K R \cos^2 \gamma} + \frac{W^2}{K^2 R^2 \cos^2 \gamma} \right) \right]. \]
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

$$\rho(L_+^+(T_{-1})) \geq \frac{1}{2} \left[ R_0 + 1 + (\sin \gamma - 1) - 2 \left( \frac{2 \hat{R}(T_{-1})}{KR_0 \cos^2 \gamma} + \frac{W^2}{K^2 R_0^2 \cos^2 \gamma} \right) \right]$$

$$\geq \frac{1}{2} \left[ R_0 + 1 - \frac{R_0}{30} - 2 \left( 30 \frac{2 \hat{R}(T_{-1})}{K R_0^2} + \frac{30 W^2}{R_0 K^2 R_0^2} \right) \right]$$

$$\geq \frac{1}{2} \left( 1 + \frac{4}{5} R_0 \right),$$

where we have used the fact that $1 - \sin \gamma \leq \frac{1}{2} - \sin \gamma = \cos^2 \gamma = R_0/30$.

Second, we estimate the infimum of the cosine of the difference of angles in $L_+^+(T_{-1})$, that is

$$\inf_{\theta, \theta' \in L_+^+(T_{-1})} \cos(\theta - \theta') = \cos(\pi - 2\gamma)$$

$$= \cos \left( 2 \left( \frac{\pi}{2} - \gamma \right) \right)$$

$$= 2 \cos^2 \left( \frac{\pi}{2} - \gamma \right) - 1$$

$$= 2 \sin^2 \gamma - 1$$

$$= 1 - \frac{1}{15} R_0.$$

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

$$mp - (1 - m) \geq \frac{1 + \frac{4R_0}{2}}{2} \left( 1 - \frac{1}{15} R_0 \right) + \left( \frac{4R_0 - 1}{2} \right)$$

$$\geq \frac{1}{2} \left( \frac{8}{5} R_0 - \frac{1}{15} R_0 - \frac{4}{75} R_0^2 \right)$$

$$> \frac{3}{5} R_0.$$

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) taking $C$ small enough. \qed

In the next corollary, we shall explain in which sense the sets whose formation we showed above have an attractive property. Before stating it 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 $L_+^+(t_0,t_1)$, as follows. First, we define

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

for any $\epsilon$ in $[R_0/15, 1)$. Second, using the notation in subsection 2.3, we will denote the $\theta$-projection of the image of $(L_+^+(t_0,t_1))_\epsilon$ under the characteristic flow, that is $\Theta_{t_1,t}(L_+^+(t_0,t_1))_\epsilon \times [-W, W)$, by

$$(L_+^+(t_0,t_1))_{\epsilon, t}. $$
for any \( t > t_1 \). 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

\[
W/K < CR_0^2,
\]

then

\[
\rho(L_{-1}^+(T_{-1}, t, t, t)) \geq \frac{1 + \frac{4}{5}R_0}{2},
\]

\[
\inf_{\theta \in (L_{-1}^+(T_{-1}), t, t)} R \cos(\theta - \phi) \geq \frac{1}{2}R_0,
\]

and

\[
\inf_{\theta, \theta' \in (L_{-1}^+(T_{-1}), t, t)} \cos(\theta - \theta') \geq 1 - \frac{1}{3}R_0,
\]

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 \( L := L_{-1}^+(T_{-1}, t, t, t) \). To verify the corresponding hypotheses; first, we begin by controlling the mass in \( L \). Indeed, by Lemma 4.3 we have that

\[
\rho(L_{-1}^+(T_{-1}, t, t, t)) \geq \rho\left(L_{-1}^+(T_{-1}, t_1, \epsilon)\right) \geq \frac{1 + \frac{4}{5}R_0}{2}.
\]

Second, we estimate the infimum over the cosine of the difference of angles in \( L_{-1}^+(T_{-1}, t, t, t) \) for this purpose let \( \bar{\theta} \) be contained in \( L_{-1}^+(T_{-1}, t_1, \epsilon) \). Then, for any \( \theta \) and \( \theta' \) in \( (L_{-1}^+(T_{-1}, t_1, \epsilon)) \), we have that

\[
\cos(\theta - \theta') = \cos(\theta - \bar{\theta} + \bar{\theta} - \theta')
\]
\[
= \cos(\theta - \bar{\theta}) \cos(\bar{\theta} - \theta) - \sin(\theta - \bar{\theta}) \sin(\bar{\theta} - \theta)
\]
\[
\geq \left[1 - \frac{1}{15}R_0\right]^2 - \left[1 - \left[1 - \frac{1}{15}R_0\right]^2\right]
\]
\[
\geq 2\left[1 - \frac{1}{15}R_0\right]^2 - 1
\]
\[
\geq 1 - \frac{1}{3}R_0.
\]

Thus, since \( \theta \) and \( \theta' \) were arbitrary, we deduce that

\[
\inf_{\theta, \theta' \in L_{-1}^+(T_{-1}, t, t, t)} \cos(\theta - \theta') \geq 1 - \frac{1}{3}R_0.
\]
Finally, considering \( m = \frac{1+4R_0}{3} \) and \( p = 1 - \frac{1}{3}R_0 \) and using the bounds in (4.30) and (4.31), we obtain that
\[
mp - (1 - m) \geq \frac{1 + 4R_0}{2} \left( 1 - \frac{1}{3}R_0 \right) + \left( \frac{4R_0}{3} - \frac{1}{2} \right)
\]
\[
\geq \frac{1}{2} \left( \frac{8}{5}R_0 - \frac{1}{3}R_0 - \frac{4}{15}R_0^2 \right)
\]
\[
> \frac{1}{2} \left( \frac{24 - 5}{15} \right) R_0
\]
\[
= \frac{1}{2} R_0.
\]
Therefore, the desired result follows by choosing \( C \) appropriately in (4.26) so that (4.5) holds and applying Lemma 4.2 with \( \sigma = \frac{R_0}{2} \).

\[\square\]

In the next Lemma, we derive an estimate that we will use in Section 5. Such an estimate shows that if the entropy production vanishes over sufficiently long intervals of time, then \( L^2 \) norm of the solution in \( T \setminus (L^+_{T-1}(T_1))_t \), will begin to decrease exponentially.

**Lemma 4.4.** Let \([t_1,t_2] \) be a time interval in \([T-1,\infty)\), such that
\[
\dot{R} \leq K e^{\lambda^2 R_0^2 \log \frac{1}{R_0}} \cos^2 \alpha
\]
and \( R < 2R_0 \) in \([t_1,t_2]\).

with \( \alpha \) as specified in Section 2. Assume that \( \epsilon = R_0/15 \) and \( \lambda \) is contained in \((2/3,1)\). Then there exists a universal constant \( C \) and some \( \delta > 0 \) such that if
\[
W_K < C \lambda^2 R_0^2 (1 - \lambda), \quad \text{and} \quad t_2 - t_1 \geq \delta
\]
then, we have that
\[
f^2(T \setminus (L^+_T(T_1))_t) \leq f^2(L^-_{T}(t_1)) e^{K(2\delta R_0 - \frac{(t_2 - t_1) R_0 \sin \alpha}{2})} \quad \text{in} \quad [t_1 + \delta, t_2].
\]

Moreover, we can choose \( \delta \) so that
\[
\delta \lesssim \frac{1}{K \lambda R_0 \cos \alpha} + \frac{\sin \alpha}{K \lambda R_0 \log \frac{1}{R_0}}.
\]

**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 3.4 and Lemma 4.3. The proof is based on Lemma 3.4, Proposition 4.1, Lemma 4.3, and the following differential inequalities:
\[
dt P \geq K \lambda R_0 \sqrt{1 - P^2} \left( \sqrt{1 - P^2} - \frac{4 \cos \alpha}{5} \right) \quad \text{in} \quad [t_1,t_2] \cap \{ P \leq \sin \alpha \},
\]
\[
dt (1 - P) \leq -\frac{1}{4} \sin \alpha K \lambda R_0 (1 - P) \quad \text{in} \quad [s,t_2] \cap \{ P \leq 1 - R_0/15 \}.
\]

Such inequalities hold when \( P = \cos(\Theta_{r,\gamma}(\theta,\omega) - \phi) \) for any \( r \) and for any \( \theta \) satisfying that \( \cos(\theta - \phi(r)) = -\sin \alpha \) in \([t_1,t_2]\), and when \( P = \cos(\Theta_{r,\gamma}(\theta,\omega) - \Theta_{T-1,\gamma}(\theta',\omega')) \) for any \( r' \) in \([t_1,t_2]\) and any \( \theta \) and \( \theta' \) such that \( \cos(\theta - \phi(r')) \geq \sin \alpha \) and \( \theta' \) is contained in \( L^+_T(T_1) \).

Here, \( \omega \) and \( \omega' \) are contained in \([-W,W]\).
We claim such inequalities imply that there exists $\delta > 0$ satisfying (4.34) such that
\begin{equation*}
T \setminus (L^+_\gamma(T-1)s) \subset L^-_\alpha(s - \delta)s,
\end{equation*}
for any $s$ in $[t_1 + \delta, t_2]$. Here, we are using the notation introduced in Section 2.3 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 3.4 and Proposition 4.1 on the interval $[t_1, s - \delta]$ we obtain that
  \begin{equation*}
  f^2(L^-_\alpha(s - \delta)) \leq f^2(L^-_\alpha(t_1))e^{-K\{(s - \delta - t_1)KR_0\sin \alpha\}}.
  \end{equation*}
Consequently, once the claim is proved, 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 $P(r) = -\sin \alpha$, the first inequality in (4.35) implies that there exists $\delta > 0$ such that
  \begin{equation*}
  \frac{d}{dt}P \geq \frac{KR_0\cos^2 \alpha}{5} \quad \text{in } [r, r + \hat{\delta}].
  \end{equation*}
Consequently, in particular we can find $\hat{\delta}$ such that the above property holds, $P(r + \hat{\delta}) = \sin \alpha$ and
  \begin{equation*}
  \hat{\delta} \leq \frac{10\alpha}{K\lambda R_0 \cos^2 \alpha}.
  \end{equation*}
By the definition of $P$ this implies that
\begin{equation*}
T \setminus L^+_\alpha(s) \subset L^-_\alpha(s - \hat{\delta})s,
\end{equation*}
for any $s$ in $[t_1 + \delta, t_2]$. To derive such implication, we have set $s = r + \hat{\delta}$.
Consequently, if we let $\theta$ be any element $T \setminus L^-_\alpha(s - \hat{\delta})s$ and we set $r' = r + \hat{\delta}$ in the definition of $P$ then, by Lemma 4.3 and construction, we have that $P(r') > -1$. Moreover, by integrating the second inequality in (4.35) we have that we can find $\overline{\delta} > 0$ such that $P(s + \hat{\delta} + \overline{\delta}) \geq 1 - \frac{R_0}{15}$ and
\begin{equation*}
\overline{\delta} \leq \frac{\sin \alpha}{K\lambda R_0 \log \frac{1}{R_0}}.
\end{equation*}
Thus, by the construction of $P$ we obtain that
\begin{equation*}
T \setminus (L^+_\gamma(T-1)s + \hat{\delta} + \overline{\delta}) \subset T \setminus L^+_\alpha(r + \hat{\delta} + \overline{\delta})r + \hat{\delta} + \overline{\delta}.
\end{equation*}
Consequently, the claim follows by selecting $s = r + \hat{\delta} + \overline{\delta}$ and $\delta = \hat{\delta} + \overline{\delta}$.

• **Step 3**: We derive (4.35).
  Let us denote:
  \begin{equation*}
  \Theta = \Theta_{r,t}(\theta, \omega), \quad \Theta = \Theta_{r',t}(\theta, \omega), \quad \text{and} \quad \Theta' = \Theta_{T-1,t}(\theta', \omega).
  \end{equation*}
To derive the first inequality, observe that thanks to Lemma 3.3 and our assumption on $\dot{R}$, we can select the constant in (4.32) appropriately so that we can guarantee that
\[
\frac{d}{dt} \cos(\Theta - \phi) = -\sin(\Theta - \phi(t))(\dot{\Theta} - \dot{\phi})
\]
\[
= -\sin(\Theta - \phi)(\omega - KR \sin(\Theta - \phi)) - \dot{\phi}
\]
\[
\geq -|\sin(\Theta - \phi)| \left[ \frac{1}{R} \sqrt{K \frac{d}{dt} R^2 + W^2 + W - KR \sin(\Theta - \phi)} \right]
\]
\[
\geq |\sin(\Theta - \phi)|((KR \sin(\Theta - \phi)) - \frac{4K\lambda R_0 \cos \alpha}{5}).
\]
Here, in the third inequality, we have used Lemma 3.3. Consequently, $P$ satisfies the inequality
\[
\frac{d}{dt} P \geq K\lambda R_0 \sqrt{1 - P^2} \left( \sqrt{1 - P^2} - \frac{4 \cos \alpha}{5} \right).
\]
Thus, the first inequality in (4.35) follows.

Finally, to derive the second inequality we use the same argument in the derivation of (4.9) to obtain that
\[
\frac{dP}{dt} \geq 2K \sqrt{1 - P^2} \left[ R \cos \left( \frac{\Theta + \Theta'}{2} - \Phi \right) \sqrt{\frac{1 - P}{2} - \frac{W}{R}} \right] \text{ in } [t_1, t_2],
\]
Now, using the same arguments as in the proof of inequality in (4.35) and equation (4.20), we obtain that
\[
\cos \left( \frac{\Theta + \Theta'}{2} - \Phi \right) = \cos \left( \frac{(\Theta - \phi) + (\Theta' - \phi)}{2} \right)
\]
\[
\geq \frac{\cos(\Theta - \phi) + \cos(\Theta' - \phi)}{2}
\]
\[
\geq \frac{\sin(\alpha) + \frac{3}{5} R_0}{2}
\]
\[
\geq \frac{\sin \alpha}{2}.
\]
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 that
\[
\frac{dP}{dt} \geq 2K \sqrt{1 - P^2} \left[ \frac{R_0 \lambda \sin \alpha}{2} \sqrt{\frac{1 - P}{2} - \frac{W}{R}} \right]
\]
\[
\geq 2K \sqrt{1 - P} \left[ \frac{\sqrt{2}}{4} R_0 \lambda \sin \alpha (1 - P) - \frac{W}{R} \sqrt{1 - P} \right].
\]
Consequently, by choosing $C$ appropriately in (4.32) so that
\[
\frac{W}{R} \sqrt{1 - P} < CR_0^2 < \frac{R_0 \lambda \sin \alpha}{20} (1 - P) \text{ whenever } 1 - P \geq R_0/15,
\]
we can guarantee that
\[
\frac{d}{dt} P \geq \frac{K\lambda R_0}{4} (1 - P), \text{ whenever } P \leq 1 - \frac{R_0}{15}.
\]
Hence, the desired result follows. \qed
We close this section with a lemma that will allow us to control the $L^2$ norm of the solution in $\mathbb{T}\backslash (L^+ (T-1))_t$ 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 $R < 2R_0$, in $[t_1, t_2]$.

Then, we have that

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

**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 to 4. We shall demonstrate that this system implies the control on the time $T_0$ presented in Theorem 1.1. We begin by describing a subdivision of the interval $[0, T_0]$ inspired by the treatment of L. Desvillettes and C. Villani in [13].

We first subordinate the subdivision to different scales of values of the order parameter. Then, we classify the intervals (of such subdivision) into intervals where dissipation is above and below a certain threshold. Such threshold depends on the scale of the order parameter.

#### 5.1. The subdivision.

Now, we give the precise construction of our subdivision. Before we enter into details, we shall introduce further notation that we will use along 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 $\{r_k\}_{k \in \mathbb{N}}$. Such a partition will be used in this part and is set according to a dyadic behavior of the square of the order parameter $R^2$. Namely, such sequence provides the first times at which $R^2$ doubles its value. To such an 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

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

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

$$\frac{W}{K} \leq CR_0^3 \text{ and } 1 - \lambda \leq \frac{\cos^2 \alpha}{180} R_0,$$

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

Now, let us set

$$\mu_k := \frac{\cos^2 \alpha}{4} \lambda^3 R_k^3, \quad d_k := \frac{1}{3KR_k} \log 10, \quad \text{and} \quad \delta_k := \frac{1}{KR_k} \log \left(\frac{1}{R_k}\right).$$
Observe that (5.2) implies that \( \frac{W}{K} \leq C\lambda^2(1 - \lambda)R_k^2 \), for any \( k = 0, \ldots, k_* \) with the same universal constant \( C \). In particular, we can use Lemma 3.4 and obtain that
\[
R(t) \geq \lambda R_k, \quad \text{for all } t \in [r_k, r_{k+1}).
\]

\[\text{• Initial time of the subdivision:}\] Let us use Lemma 4.3 to define the corresponding times of formation of attractors that is, we set
\[
T_{-1}^k := \inf \left\{ t \geq r_k : \frac{dR}{dt} \leq KQ R_k^3 \right\},
\]
where \( k = 0, \ldots, k_* \) 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 let
\[
t_0 := \min \{ T_{-1}^k : k = 0, \ldots, k_* \},
\]
and
\[
k_0 := \max \{ k \in \mathbb{Z}_+^* : r_k \leq t_0 \}.
\]
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 \( (L_+^+ (t_0))_t \). Here, and throughout the rest of this section we will choose \( \gamma \) by the condition
\[
\cos^2 \gamma = \frac{1}{30} R_{k_0}.
\]
We have done so according to condition (4.16).

\[\text{• The subdivision:}\] 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 \( \{ 0, \ldots, k_* \} \) such that \( t_l \) is contained in \( [r_{k(l)}, r_{k(l)}+1) \). Then, we will distinguish between two different situations:

1. If \( \tilde{R}(t_l) < K\mu_{k(l)} \), then we set
\[
t_{l+1} := \sup \{ t \in [t_l, r_{k(l)}+1) : \tilde{R}(s) < K\mu_{k(l)}, \forall s \in [t_l, s) \},
\]

2. If \( \tilde{R}(t_l) \geq K\mu_{k(l)} \), then we first compute
\[
\tilde{t}_{l+1} := \sup \{ t \in [t_l, r_{k(l)}+1) : \tilde{R}(s) \geq K\mu_{k(l)}, \forall s \in [t_l, s) \},
\]
and set \( t_{l+1} \) via the following correction:
\[
t_{l+1} = \begin{cases} 
\tilde{t}_{l+1} + d_{k(l)} & \text{if } \tilde{t}_{l+1} + d_{k(l)} \leq r_{k(l)}+1, \\
r_{k(l)}+1 & \text{otherwise.}
\end{cases}
\]

\[\text{• The good and the bad sets:}\] 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 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 $R_{k(l)}^2$ in the above dyadic hierarchy as depicted in (5.3). For this reason, we shall collect all the indices $l$ of good and bad sets associated to the index $k$ of the dyadic hierarchy as follows.

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

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

\begin{align}
g_k & := |G_k|, \\
b_k & := |B_k|,
\end{align}

for every $k = 0, \ldots, k_*$. Notice that as a consequence of the 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 $[t_l, t_{l+1})$ according to (5.11) 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 $[t_{l+1}, t_{l+1})$ and we then lose the control of what is next: either $G_k$ or $B_k$ set. Nevertheless, this is enough to show that

\begin{equation}
b_k \leq g_k + 1, \quad \text{for all } k = 0, \ldots, k_*.\end{equation}

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 shall sometimes enumerate the indices in $G_k$ in an increasing manner, namely,

\[G_k = \{ t_k^m : m = 1, \ldots, g_k \},\]

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

5.1.1. *Bound of the size of $t_0$. By Lemma 4.3 we have that that each $T_{k-1}^k$ 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 that

\[t_0 \lesssim \frac{1}{K R_0^2}.\]

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

\[R(r_{k+1}) - R(r_k) = \int_{r_k}^{r_{k+1}} \dot{R}(t) \, dt \geq K Q R_k^2 (r_{k+1} - r_k),\]

and

\[R(t_0) - R(r_{k_0}) = \int_{r_{k_0}}^{t_0} \dot{R}(t) \, dt \geq K Q R_{k_0}^2 (t_0 - r_{k_0}),\]

for any $k = 0, \ldots, k_0 - 1$. Here, we have used the fact that $r_k \leq t_0 \leq T_{k-1}^k$ for every $k = 0, \ldots, k_0$. By estimate (5.0) and the definition of $T_{k-1}^k$ in (5.0) 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

\[
 r_{k+1} - r_k \leq Q \frac{(R(r_{k+1}) - R(r_k))}{KR_k^3} \leq \frac{1}{2} \frac{Q}{KR_k^2},
\]

and

\[
 t_0 - r_{k_0} \leq Q \frac{(R(t_0) - R(r_{k_0}))}{KR_{k_0}^3} \leq \frac{1}{2} \frac{Q}{KR_{k_0}^2},
\]

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

\[
 t_0 = t_0 - r_{k_0} + \sum_{k=0}^{k_0-1} (r_{k+1} - r_k) \leq \frac{1}{2} \frac{Q}{KR_{k_0}^2} \sum_{k=0}^{k_0} \left( \frac{1}{2} \right)^k \leq \frac{Q}{KR_0^2}.
\]

\[ \square \]

5.1.2. **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 \( \{k_0, \ldots, k_\alpha \} \), 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 Collorary 4.3. In the following Lemma, we will see that assumption (5.2) implies that the loss in \( R^2 \) in smaller than \( 4/5 \) of the gain (except on possibly the last interval of \( [t_{l(k)}, t_{l(k)+1}) \).

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

\[
 R^2(t_l) - R^2(t_{l+1}) \leq \frac{4}{5} \left( R^2(t_{l+m+1}^k) - R^2(t_m^k) \right),
\]

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

**Proof.** Thanks to Corollary 3.6 and Lemma 2.1 we have that

\[
 R^2(t_l) - R^2(t_{l+1}) \leq (1 - \lambda^2)R^2(t_l) \leq 4(1 - \lambda)R_k^2 \quad \text{and} \quad R^2(t_{l+m+1}^k) - R^2(t_m^k) \geq \frac{1}{40} \lambda^4 R_k^3.
\]

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

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

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

\[ \square \]

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

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

\[
 \max(b_k, g_k) \lesssim \frac{1}{R_k}.
\]
Proof. To prove this, recall that by Lemma \ref{lem:5.4}, we have that

\begin{equation}
\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}.
\end{equation}

Thus, Lemma \ref{lem:5.2} implies

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

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 on whether $l(k)$ is in $B_k$ or $G_k$. By doing this, we obtain that

\begin{equation*}
R^2_{k+1} - R^2_k \geq \frac{g_k}{200} \lambda^4 R_k^3 - \frac{1}{40} \chi(l(k) \in G_k) \lambda^4 R_k^3.
\end{equation*}

Hence, we deduce the bound

\begin{equation}
g_k \leq \frac{200(2 - \lambda^2) R_k^2}{R_k^3} + 5.
\end{equation}

Here, we have used the fact that assumption \ref{lem:5.2} implies that $\lambda > 179/180$. Hence, the desired result follows. \qed

5.1.4. 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 $G_k$ and $B_k$. The set $G_k$ 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 $G_k = G_k$. On the other hand, the set $B_k$ is obtained 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 $B_k = B_k$. Now, we are ready to state our control.

Lemma 5.4. The sum of the lengths of the interval $[t^k_{m}, t^k_{m+1}]$ satisfies

\begin{equation*}
\sum_{m=1}^{g_k} \left( t^k_{m+1} - t^k_{m} \right) \lesssim \frac{1}{K R_k^2}.
\end{equation*}

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

\begin{equation}
t^k_{m+1} - t^k_{m} = (\tilde{t}^k_{m+1} - t^k_{m}) + d_k.
\end{equation}

Our next goal is to estimate the first term. To such end, we shall use the idea in Lemma \ref{lem:5.3} and the fundamental theorem of calculus to write

\begin{equation*}
R(\tilde{t}^k_{m+1}) - R(t^k_{m}) = \int_{t^k_{m}}^{\tilde{t}^k_{m+1}} \dot{R}(t) \, dt \geq \frac{\cos^2 \alpha}{4} K \lambda^3 R_k^3 \left( \tilde{t}^k_{m+1} - t^k_{m} \right),
\end{equation*}

for all $m = 1, \ldots, g_k$. Here, we have used \eqref{eq:5.9} to bound the time derivative of $R$. Hence, we obtain

\begin{equation}
\tilde{t}^k_{m+1} - t^k_{m} \leq \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} \left( R(\tilde{t}^k_{m+1}) - R(t^k_{m}) \right),
\end{equation}

\begin{equation}
t^k_{m+1} - t^k_{m} \leq \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} \left( R(\tilde{t}^k_{m+1}) - R(t^k_{m}) \right),
\end{equation}

\begin{equation}
\sum_{m=1}^{g_k} \left( t^k_{m+1} - t^k_{m} \right) \lesssim \frac{1}{K R_k^2}.
\end{equation}
for all \( m = 1, \ldots, g_k \). By summing over all the intervals of type \( G_k \) we obtain that
\[
\sum_{l \in G_k} (\tilde{t}_{l+1} - t_l) \leq \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} \sum_{l \in G_k} (R(\tilde{t}_{l+1}) - R(t_l))
\]
\[
= \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} \sum_{l \in G_k} \left[ (R(t_{l+1}) - R(t_{l_m})) - (R(t_{l+1}) - R(\tilde{t}_{l+1})) \right],
\]

Let us add and subtract to the first term in (5.21) the oscillations of \( R \) over all the sets of type \( B_k \). 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 turns, 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 of the order parameter given by (5.4). Then, we obtain
\[
\sum_{l \in G_k} (\tilde{t}_{l+1} - t_l) \leq \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} (R_{k+1} - \lambda R_k)
\]
\[
- \frac{4}{\cos^2 \alpha K \lambda^3 R_k^3} \left[ \sum_{l \in B_k} (R(t_l) - R(t_{l+1})) + \sum_{l \in G_k} (R(t_{l+1}) - R(\tilde{t}_{l+1})) \right].
\]

Our goal is to show that the term in the second line is non-positive. Indeed, let us use lemmas 2.1 and 5.2 in the second term of (5.22) to obtain that
\[
\sum_{l \in G_k} (\tilde{t}_{l+1} - t_l) \leq \frac{4(2 - \lambda)}{K \cos^2 \alpha \lambda^3 R_k^2} - \frac{4}{5 \cos^2 \alpha K \lambda^3 R_k^3} \sum_{l \in G_k} (R(t_{l+1}) - R(\tilde{t}_{l+1}))
\]
\[
\leq \frac{4(2 - \lambda)}{\cos^2 \alpha K \lambda^3 R_k^2}.
\]

Hence, by lemmas 2.1 and 5.3 and (5.19) we deduce that
\[
\sum_{m=1}^{g_k} (t_{l_{m+1}} - t_{l_m}) \leq d_k g_k + \tilde{t}_{l_{g_k+1}} - t_{l_{g_k}} + \sum_{l \in G_k} (\tilde{t}_{l+1} - t_l)
\]
\[
\leq \frac{1}{K R_k},
\]
where we have used (5.20) and our usual bound on the oscillation to control the difference
\[
\tilde{t}_{l_{g_k+1}} - t_{l_{g_k}}.
\]

Thus, the desired result follows. \( \square \)

5.1.5. Growth of \( f^2(T \setminus (L^+(t_{0}) t_{L})) \). Our goal here is to control the growth of \( f^2(T \setminus (L^+(t_{0}) t_{L})) \) in each interval \([r_k, r_{k+1}]\), where the parameter \( \epsilon \) of the neighborhood is set once for all as follows
\[
\epsilon := \frac{R_0}{15}.
\]
Notice that $\epsilon$ has been set so that the attractive property in Corollary 4.1 holds true. To initialize the iterative method, we need to control $f^2(\mathbb{T} \setminus (L^+_x(t_0)))$ at $t = t_0$. Hence, we begin by providing a control of the growth of $f^2_t(\mathbb{T})$ during the transient $[0,t_0]$.

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

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

**Proof.** Thanks to Proposition 4.1 we obtain that

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

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

$$\int_{t_0}^{t_0} R(s) \, ds = \sum_{k=0}^{k_0-1} \int_{r_k}^{r_{k+1}} R(s) \, ds + \int_{r_{k_0}}^{t_0} R(s) \, ds$$

$$\leq \sum_{k=0}^{k_0-1} R_{k+1}(r_{k+1} - r_k) + R_{k_0+1}(t_0 - r_{k_0})$$

$$\leq Q \sum_{k=0}^{k_0} \frac{R_k}{KR_k^2}$$

$$= Q \sum_{k=0}^{k_0} \frac{1}{KR_0} \left( \frac{\sqrt{2}}{2} \right)^k$$

$$\leq \frac{4Q}{KR_0}.$$

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 on 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

$$(5.23) \quad D_k := \max(b_k, g_k)(\delta_k + d_k) + \sum_{l=1}^{g_k} (\tilde{t}_{k_{m+1}} - \tilde{t}_{k_m}),$$

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):

$$(5.24) \quad D_k \lesssim \frac{1}{KR_k^2} + \frac{1}{R_k} \left[ \frac{1}{KR_k} \log \left( \frac{1}{R_k} \right) + \frac{1}{KR_k} \right]$$

$$\lesssim \frac{1}{KR_k^2} \log \left( 1 + \frac{1}{R_k} \right).$$

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

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

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

Lemma 5.6. Assume condition (5.2) holds, then we have that

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

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

Proof. By definition, we note that

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

and for every $k = k_0 + 1 \ldots, k_s$. Also, notice that by contruction, we have that

$$F_{k_0}(r_{k_0 + 1}) \leq ||f_0||_2^2 e^{\frac{4B}{2} D_{k_0}} e^{2KR_0 D_{k_0}}.$$ 

Then, a simple induction shows that

$$F_k(t) \leq ||f_0||_2^2 e^{\frac{B}{2} \log(1 + \frac{t}{R_0})} e^{2KR_0 D_k} = ||f_0||_2^2 \exp\left(\frac{4Q}{R_0} + \sum_{q=k_0}^k 2KR_q D_q\right).$$

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

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

Hence, the desired result follows.

The sequence $\{F_k\}_{k=k_0}^{k_s}$ has been constructed as a barrier in order to control the map $t \to f^2(T \setminus (L_\gamma^+(t_0)\epsilon))$ at each interval $[r_k, r_{k+1})$. We achieve this in the following theorem. Such a theorem is the main result in this section. As a byproduct, we derive corollaries 5.1 and 2.1 which provide the basis for our discussion in Section 3.1 and Section 3.2.

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

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

for each $k = k_0, \ldots, k_s$.

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 $f^2(T \setminus (L_\gamma^+(t_0)\epsilon))$ via lemmas 4.1, 4.3 and 4.5. Specifically, we shall distinguish between three different scenarios for each interval $[t_i, t_{i+1})$ with $t_i$ in $[r_{k_0}, r_{k_0+1})$: 

--


(1) If the interval is of type $G_{k_0}$, then $\hat{R}(t_0) \geq K\mu_{k_0}$ and Lemma 4.3 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 that

$$f^2(L_{a}^-(t)) \leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon,t})$$

$$\leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{2KR_{k_0}(t-t_0)}$$

$$\leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{2KR_{k_0}(t_{i+1}-t_t)}$$

for every $t$ in $[t_1, t_{l+1})$.

(2) If the interval is of type $B_{k_0}$, then two different possibilities can take place: either $[t_1, t_{l+1})$ is small or it is large.

(a) If $[t_1, t_{l+1})$ 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

$$f^2(L_{a}^-(t)) \leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{2KR_{k_0}(t-t_0)} \leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{2KR_{k_0}\delta_{k_0}}$$

for every $t$ in $[t_1, t_{l+1})$.

(b) Finally, if $[t_1, t_{l+1})$ 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 $[t_l, t_{l+1})$, we can only apply the same argument as before supported by Lemma 4.1 about sliding $L^2$ norm. Specifically, for any $t$ in $[t_l, t_{l+1})$ Lemma 4.1 implies

$$f^2(L_{a}^-(t)) \leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{2KR_{k_0}\delta_{k_0}}$$

Now, for any $t$ in $[t_l + \delta_{k_0}, t_{l+1})$ lemmas 4.4 and 4.5 yield

$$f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) \leq f^2(L_{a}^-(t_0)) e^{\frac{K\left(2R_{k_0}\delta_{k_0} - \frac{(t-l-t_{l+1})R_{k_0}\sin\alpha}{2}\right)}{2}}$$

$$\leq f^2(T \setminus (L_{a}^+(t_0)_t)_{\epsilon}) e^{\frac{K\left(2R_{k_0}\delta_{k_0} - \frac{(t-l-t_{l+1})R_{k_0}\sin\alpha}{2}\right)}{2}}.$$
By applying the above discussion in a recursive way, we obtain that
\[
I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0}
\leq I_{t_0}^2(T) \exp \left\{ 2R_{k_0} \sum_{l \in G_{k_0}} (t_{l+1} - t_l) + \sum_{l \in B_{k_0}^s} \delta_{k_0} \right\}
+ \sum_{l \in B_{k_0}^l} \left( 2R_{k_0} \delta_{k_0} - \frac{(t_{l+1} - t_l - \delta_{k_0})R_{k_0} \sin \alpha}{2} \right),
\]
(5.26)

Similarly, for any \( t \) in \( (t_p, t_p + \delta_{k_0}) \) we have that
\[
I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0}
\leq I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0} \exp \left\{ 2KR_{k_0} \left[ (t_{p+1} - t_p) \chi_{\{p \in G_{k_0}\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^s\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^l\}} \right] \right\}
\leq I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0} \exp \left\{ 2KR_{k_0} \left[ (t_{p+1} - t_p) \chi_{\{p \in G_{k_0}\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^s\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^l\}} \right] \right\}.
\]
(5.28)

Thus, for any \( t \) in \( [t_p + \delta_{k_0}, t_{p+1}] \) we obtain that
\[
I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0}
\leq I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0} \exp \left\{ 2KR_{k_0} \left[ (t_{p+1} - t_p) \chi_{\{p \in G_{k_0}\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^s\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^l\}} \right] + \left( \delta_{k_0} - \frac{(t_p - t - \delta_{k_0}) \sin \alpha}{4} \right) \chi_{\{p \in B_{k_0}^l\}} \right\}
\leq I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0} \exp \left\{ 2KR_{k_0} \left[ (t_{p+1} - t_p) \chi_{\{p \in G_{k_0}\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^s\}} + \delta_{k_0} \chi_{\{p \in B_{k_0}^l\}} \right] - \frac{(t - t_p - \delta_{k_0}) R_0 \sin \alpha}{4} \chi_{\{p \in B_{k_0}^l\}} \right\}.
\]
(5.27)

Putting (5.26), (5.27) and (5.28) together and recalling \( D_k \) in (5.23) implies
\[
I_{t_0}^2(T) \setminus (L_+^\gamma (t_0)_{t_0})_{t_0}
\leq I_{t_0}^2(T) \exp \left\{ 2KD_{k_0} R_{k_0} - \sum_{l \in B_{k_0}^s} K \frac{(t_{l+1} - t_l) R_{k_0} \sin \alpha}{2} \right\}
- \frac{K (t - t_p) R_{k_0} \sin \alpha}{2} \chi_{\{p \in B_{k_0}^l\}},
\]
(5.29)

where we have absorbed the \( \delta_{k_0} \) in the last term into \( D_{k_0} \).

On the other hand, notice that we can recover \( t \) from the following telescopic sum
\[
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}^s\}} + \sum_{l \in B_{k_0}^s} (t_{l+1} - t_l) + \sum_{l \in B_{k_0}^l} (t_{l+1} - t_l)
\leq t_0 + D_{k_0} + (t - t_p) \chi_{\{p \in B_{k_0}^s\}} + \sum_{l \in B_{k_0}^l} (t_{l+1} - t_l).
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,

\[
(5.30) \quad f^2_l(T \setminus (L^+_p(t_0)\chi)_e) \leq \begin{cases} 
 f^2_l(T)e^{2K_Dk_0}, & \text{for } t \in (t_p, t_p + \delta_{k_0}), \\
 f^2_l(T)e^{2K_Dk_0 - \frac{KR_{k_0}sin\alpha}{2}(t-t_0-D_{k_0})}, & \text{for } t \in [t_p + \delta_{k_0}, t_{p+1}]. 
\end{cases}
\]

Notice that the worst situation is the one where there is no intermediate fall-off, that is, \(B_{k_0,p} = \emptyset\). Since such scenario dominates all the other possibilities, we shall restrict to it without loss of generality. This amounts to the chain of inequalities

\[
t_p + \delta_{k_0} = t_0 + \delta_{k_0} + \sum_{l \in G_{k_0,p}} (t_{l+1} - t_l) + \sum_{l \in B^2_{k_0,p}} (t_{l+1} - t_l) + \sum_{l \in B^L_{k_0,p}} (t_{l+1} - t_l)
\]

\[
\leq t_0 + \sum_{l \in G_{k_0}} (t_{l+1} - t_l) + \max(g_k, b_k)\delta_k
\]

\[
\leq t_0 + D_{k_0},
\]

that is, \(t_p + \delta_{k_0} \leq t_0 + D_{k_0}\), that leads to restating (5.30) as follows

\[
f^2_l(T \setminus (L^+_p(t_0)\chi)_e) \leq \begin{cases} 
 f^2_l(T)e^{2K_Dk_0R_{k_0}}, & \text{for } t \in (t_0, t_0 + D_{k_0}), \\
 f^2_l(T)e^{2K_Dk_0R_{k_0} - \frac{KR_{k_0}sin\alpha}{2}(t-t_0-D_{k_0})}, & \text{for } t \in [t_0 + D_{k_0}, r_{k_0+1}]. 
\end{cases}
\]

Finally, use Lemma [5.4] to relate the \(L^2\) norm at \(t = t_0\) and at \(t = 0\). Thus, we have showed the claimed bound.

• **Step 2:** Inductive hypothesis.

Let us assume that for certain \(k_0 < k < k_*\) we have

\[
f^2(T \setminus (L^+_p(t_0)\chi)_e) \leq F_q(t), \quad t \in [r_q, r_{q+1}),
\]

for any \(q < k\).

• **Step 3:** Induction step.

The proof for the index \(k\) becomes a simple consequence of the inductive hypothesis where we need to apply again lemmas 4.1, 4.4 and 4.5 repeatedly in the spirit as in Step 1 for the base step.

As a consequence of Theorem 5.1 we obtain the following two Corollaries.

**Corollary 5.1.** Suppose assumption (5.2) holds. Then, we have that

\[
r_{k+1} - r_k \lesssim \frac{1}{KR_k R_0} \frac{1}{log \left(1 + \frac{1}{R_0} + W^{1/2}\|f_0\|_2\right)},
\]

for any \(k \leq k_*\).
Proof. Thanks to (5.24), we may assume, without lost of generality that \( r_{k+1} - r_k \geq D_k \).
Now, observe that, by Theorem 5.1 and (5.25) we have that
\[
f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)))_\varepsilon \leq F_k(t)
\]
\[
\leq \|f_0\|^2_2 \frac{2e^{\beta R_k}}{\sqrt{N_0}} \left( \prod_{q=k_0}^k e^{2KR_qD_q} \right) e^{-K \frac{R_k \sin \alpha}{2}(t-r_k-D_k)}.
\]
\[
\leq \|f_0\|^2_2 \frac{\sqrt{2}}{\sqrt{N_0}} e^{\alpha(1+e^{-\varepsilon})} e^{-K \frac{R_k \sin \alpha}{4}(t-r_k-D_k)},
\]
for every \( t \) in \([r_k + D_k, r_{k+1})\) and some universal constant \( Q'\).
On the other hand, by Jensen inequality, we have that
\[
\rho(\mathbb{T} \setminus (L_\gamma^+(t_0)))_\varepsilon \leq \sqrt{4\pi W f^2(\mathbb{T} \setminus (L_\gamma^+(t_0)))_\varepsilon}.
\]
Consequently, if we let \( m(s) = 1 - \rho(\mathbb{T} \setminus (L_\gamma^+(t_0)))_\varepsilon \), using Theorem 5.1 we deduce that
\[
1 - m(s) \leq 2\sqrt{\pi \|f_0\|^2_2} e^{\frac{\beta}{2N_0}} e^{\frac{\alpha(1+e^{-\varepsilon})}{N_0}} e^{-K \frac{R_k \sin \alpha}{4}(s-r_k-D_k)}.
\]
For any \( s \) in \([D_k + r_k, r_{k+1}].\) On the other hand, by lemmas 4.2 and 4.3 and Corollary 4.1 if we let
\[
P(t) = \inf_{\theta, \theta' \in (L_\gamma^+(t_0))_\varepsilon} \cos(\theta - \theta'),
\]
we have that
\[
1 - P(t) \leq \max \left[ \frac{1}{3} R_k e^{-K R_k (t-s)}, \frac{16}{R_k} \frac{W^2\pi^2}{K^2} \right]
\]
for every \( t \) in \([s, r_{k+1}].\)
Additionally, using lemmas 4.2 and 4.3 and Corollary 4.1 if we let \( L = (L_\gamma^+(t_0))_\varepsilon \) we have that
\[
R(t) \geq \inf_{\theta, \theta' \in L, \varepsilon} R \cos(\theta - \theta')
\]
\[
\geq m(s) P(t) - (1 - m(s))
\]
\[
= (1 - (1 - m(s))) P(t) - (1 - m(s))
\]
\[
\geq P(t) - 2(1 - m(s))
\]
\[
\geq 1 - (1 - P(t)) - 4\sqrt{\pi W^2 \|f_0\|^2_2} e^{\frac{\beta}{2N_0}} e^{\frac{\alpha(1+e^{-\varepsilon})}{N_0}} e^{-K \frac{R_k \sin \alpha}{4}(s-r_k-D_k)}.
\]
Now, observe that, by construction
\[
\frac{\sqrt{2}}{2} \geq R \in [r_k, r_{k+1}).
\]
Consequently, by (5.31) and (5.32), if we set \( t = r_{k+1} \) and \( s = r_{k+1} - \frac{8}{\pi R_k} \frac{\log \frac{1}{\varepsilon}}{\log \frac{1}{N_k}} \) in (5.33), and make \( C \) smaller within the constrains of (5.2) if necessary, we obtain that
\[
\frac{1}{3} R_k e^{-\frac{1}{2} \frac{\log \frac{1}{\varepsilon}}{\log \frac{1}{N_k}}} + 4\sqrt{\pi W^2} \|f_0\|^2_2 e^{\frac{\beta}{2N_0}} e^{\frac{\alpha(1+e^{-\varepsilon})}{N_0}} e^{-K \frac{R_k \sin \alpha}{4}(r_{k+1} - r_k - D_k - \frac{8}{\pi R_k} \frac{\log \frac{1}{\varepsilon}}{\log \frac{1}{N_k}})} \geq 1 - \frac{\sqrt{2}}{2}.
\]
Thus,
\[
4\sqrt{\pi}\|f_0\|_2 W^{1/2} e^{\frac{C_1}{R_0} \log(1 + \frac{1}{R_0})} e^{-K R_0 \sin \alpha \left( r_{k+1} - r_k - D_k \right)} \geq 1 - \frac{\sqrt{2}}{2} - \frac{1}{30} \geq \frac{1}{10},
\]
for some universal constant \( C_1 \).

Hence,
\[
\frac{4}{K R_k \sin \alpha} \log \left( 40 \sqrt{\pi} W^{1/2} ||f_0||_2 \right) + \frac{4C_1}{KR_0 R_k \sin \alpha} \log \left( 1 + \frac{1}{R_0} \right) + D_k \geq r_{k+1} - r_k.
\]
Consequently, using (5.24) the desired result follows. \( \square \)

5.2. **Proof of Corollary 2.1.** We will prove the Corollary, by proving that
\[
(5.35) \quad \rho(T \setminus (L^+_{t_0}(t_0), s, t) \leq e^{-\frac{1}{10} K \sin \alpha (t - T_0)},
\]
and
\[
(L^+_{t_0}(t_0), s, t) \subset L^+_{\beta}(t),
\]
for every \( t \in [T_0, \infty) \). Here,
\[
s = t - \frac{8}{KR_k} \log \frac{1}{40 R_k}.
\]
Additionally, recall that \( \gamma \) was chosen in (5.7).

We begin by showing the first equation in (2.13). To do this, we control \( r_k^* \) via the following telescopic sum and Corollary 5.1
\[
r_k^* = t_0 + \sum_{k=k_0}^{k^*} r_{k+1} - r_k \leq \frac{1}{KR_0^2} + \sum_{k=k_0}^{k^*} \frac{1}{KR_k R_0} \log \left( 1 + \frac{1}{R_0} + W^{1/2} ||f_0\|_2 \right) \]
\[
\leq \frac{1}{KR_0^2} + \sum_{k=k_0}^{k^*} \left( \frac{\sqrt{2}}{2} \right)^k \frac{1}{KR_0^2} \log \left( 1 + \frac{1}{R_0} + W^{1/2} ||f_0\|_2 \right).
\]
Consequently, by construction, to guarantee the first equation in (2.13) 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 that
\[
R(t) \geq \frac{\sqrt{2}}{2} \lambda \geq \frac{3}{5},
\]
for every \( t \) in \([r_{k^*}, \infty)\).
Now, we proceed to show that we can guarantee the second equation in (2.13) by selecting $T_0$ within the desired constraints. To achieve this, we argue as in equation (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 that,

(5.36)

$$\rho(T \setminus (L^+_{\epsilon, t} (t_0))) \leq 4\sqrt{\pi W^{1/2}} \|f_0\|_{2e^{2f_0}} \log \left(1 + \frac{1}{n_0}\right) e^{-K R_k \sin \alpha / 4} \left(t - r_k - D_k - \frac{8}{KR_k} \log \frac{1}{40R_k}\right),$$

and

(5.37)

$$\inf_{\theta \in (L^+_{\epsilon, t} (t_0)))} \cos(\theta - \phi) \geq 1 - \frac{1}{3} R_k e^{-\log \frac{1}{40R_k}},$$

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

(5.38)

$$\frac{1}{K R_0^2} \log \left(1 + \frac{1}{R_0} + W^{1/2} \|f_0\|_{L^2}\right) \geq T_0 \geq \frac{4}{KR_k \sin \alpha} \left[\log \frac{4\sqrt{\pi W^{1/2}} \|f_0\|_{L^2}}{R_k / 120} + \frac{Q^* + 16}{2KR_0} \log \left(1 + \frac{1}{40R_0}\right) + r_k + D_k\right],$$

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

(5.39)

$$\inf_{\theta \in (L^+_{\epsilon, t} (t_0)))} \cos(\theta - \phi) \geq \frac{59}{60},$$

and

$$\rho(T \setminus (L^+_{\epsilon, t} (t_0))) \leq \frac{1}{120} e^{-K R_k \sin \alpha / 4 (t - T_0)},$$

for every $t$ in $[T_0, \infty)$. Consequently, the desired result follows from the fact that (5.39) implies that $(L^+_{\epsilon, t} (t_0)) \subset L^+_{\beta}(t)$. □

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. Along 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 $W$ satisfies 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 [[42], Theorem 3.5]), let us consider a probability space \((E, \mathcal{F}, P)\) and set some sequence of random variables for \(k \in \mathbb{N}\)

\[
(\theta_k(0), \omega_k(0)) : E \rightarrow T \times \mathbb{R},
\]

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 agent-based system (1.1) issued at the above random initial data. Then, we define the associated random empirical measures as follows

\[
\mu_t^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{(\theta_i^N(t), \omega_i(0))}(\theta, \omega),
\]

for every \(t \geq 0\).

The proof of Corollary 1.1 gathers three different tools:

- First, we shall use our main Theorem 1.1 that quantifies the rate of convergence of the solution \(f = f(t, \theta, \omega)\) towards the global equilibrium \(f_\infty\) 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 \(P\) of \(\mu_0^N\) 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 \(\mu_t^N\) 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 \(f_\infty\). 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 if 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 \(\mu_0^N\) in Definition 6.1 approximate the initial datum \(f_0\) as \(N \to \infty\). Specifically, by the strong Law of Large Numbers (see [[11]]) we obtain that

\[
\mu_0^N \overset{\text{a.s.}}{\rightharpoonup} f_0, \quad P\text{-a.s.,}
\]
in the narrow topology of \(P(T \times \mathbb{R})\) 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 data \(f_0\) and the sequence of random empirical measures \(\mu_0^N\), see [[6, 7, 8]]. Specifically, some transportation-entropy inequality is required. To the best of our knowledge, the first result where those assumptions on can be removed was recently introduced in [[19]]. In our particular setting, it reads as follows.
Lemma 6.1. Let $f_0$ be 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
\begin{equation}
\mathcal{E}(g) := \int_{\mathbb{R}} e^{\omega^4} \, dg < \infty.
\end{equation}
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 $\mu_0^N$ according to Definition 6.7. Then,
\[ \mathbb{P} \left( W_2(\mu_0^N, f_0) \geq \varepsilon \right) \leq C_1 e^{-C_2 N \varepsilon^4}, \]
for every $\varepsilon > 0$ and $N$ in $\mathbb{N}$. Here, $C_1$ and $C_2$ are two positive constants that depend neither on $\varepsilon$ nor on $N$, but only depend on $\mathcal{E}(g)$.

Proof. Take $d = 2, p = 2, \gamma = 1$ and $\beta = 4$ in [19, 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 in Section 3, such 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 such situation by scaling $\omega$. Let us recall the scaled quadratic Wasserstein distance (see Definition 6.1),
\[ SW_2(\mu_0^N, f_0) = \left( \inf_{\gamma \in \Pi(\mu_0^N, f_0)} \int_{\mathbb{T}^2 \times \mathbb{R}^2} \left( d(\theta, \theta')^2 + \frac{(\omega - \omega')^2}{K^2} \right) \, d\gamma \right)^{1/2}. \]
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 $\omega$,
\[ D_K(\omega) := \frac{\omega}{K}, \quad \text{for } \omega \in \mathbb{R}. \]
Then, we can define the following scaled objects:
\[ f_{0,K} := (\text{Id} \otimes D_K)_{#} f_0 \quad \text{and} \quad \mu_N^{0,K} := (\text{Id} \otimes D_K)_{#} \mu_0^N. \]
Notice that $f_{0,K}$ is contained in $\mathbb{P}(\mathbb{T} \times \mathbb{R})$ and the empirical measures $\mu_0^{N,K}$ are i.i.d. variables with law $f_{0,K}$. Interestingly, we obtain the relation
\[ SW_2(\mu_0^N, f_0) = W_2(\mu_0^{N,K}, f_{0,K}). \]
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 $C^1(\mathbb{T} \times \mathbb{R})$, assume that the distribution of natural frequencies $g = (\pi_\omega)_{#} f_0$ has compact support in $[-W, W]$ and that condition (1.15) 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 $\mu_0^N$ according to Definition 6.7. Then,
\begin{equation}
\mathbb{P} \left( SW_2(\mu_0^N, f_0) \geq \varepsilon \right) \leq C_1 \exp \left( -C_2 N \varepsilon^4 \right),
\end{equation}
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 6.2, the above $C_1$ and $C_2$ only depend upon
$E(g_K)$ where $g_K := D_K # g$. Since $g$ has compact support in $[-W, W]$ we obtain that
$$1 \leq E(g_K) \leq e^{\frac{2K}{K}},$$
so that $C_1$ and $C_2$ will ultimately depend only on $\frac{W}{K}$. However, under the assumptions
$(1.15)$ in Theorem $7.2$ $\frac{W}{K}$ is smaller than a universal constant. Consequently, $E(g_K)$ can be
made smaller than a universal constant arbitrarily close to 1. This justifies that $C_1$ and $C_2$
can be considered universal constants.

6.2. Wasserstein stability estimate. The study of Wasserstein stability estimates or
Dobrushin-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 trans-
portation distance has to be considered. In particular, the starting works by R. Dobrushin
and H. Neunzert (see [18, 33]) show that the bounded-Lipschitz distance is appropriate for
Lipschitz-continuous interaction kernels. This type of inequalities has been generalized to
specific kernels with more limited regularity. In particular, the right transportation distance
for gradient flows associated with $W_2$ (see [11]). Indeed, we do not necessarily need an underlying gradient structure, but
only require that the interaction kernel is one-sided Lipschitz-continuous. This was proved
in [38, Theorem 4.7] for the Kuramoto model with weakly singular weights, that in our case
provides the following stability estimate for $W_2$

$$W_2(f_t, \bar{f}_t) \leq e^{\left(2K + \frac{4}{5}\right)t} W_2(f_0, \bar{f}_0),$$

which holds for any two measured valued solution 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 data $f_0$ and $\bar{f}_0 \in \mathbb{P}_2(\mathbb{T} \times \mathbb{R})$. Then, we have that

$$SW_2(f_t, \bar{f}_t) \leq e^{\frac{2}{5}Kt} SW_2(f_0, \bar{f}_0),$$

for every $t \geq 0$.

Proof. Consider an optimal transference plan $\gamma_0$ joining $f_0$ to $\bar{f}_0$, i.e.,

$$\gamma_0 \in \Pi(f_0, \bar{f}_0) := \{ \gamma \in \mathbb{P}((\mathbb{T} \times \mathbb{R}) \times (\mathbb{T} \times \mathbb{R})) : (\pi_1)_# \gamma = f_0 \text{ and } (\pi_2)_# \gamma = \bar{f}_0 \},$$

such that

$$SW_2(f_0, \bar{f}_0)^2 = \int_{\mathbb{T} \times \mathbb{R}} \int_{\mathbb{T} \times \mathbb{R}} d_{\mathbb{K}}((\theta_1, \omega_1), (\theta_2, \omega_2))^2 d_{((\theta_1, \omega_1), (\theta_2, \omega_2))} \gamma_0.$$ 

Here $\pi_1$ and $\pi_2$ represent the projections

$$\pi_1((\theta, \omega), (\theta', \omega')) = (\theta, \omega),$$

$$\pi_2((\theta, \omega), (\theta', \omega')) = (\theta', \omega').$$

Let us consider the following competitor at time $t$ via push-forward, namely,

$$\gamma_t := (X_{0,t} \otimes \bar{X}_{0,t})_# \gamma_0 \in \mathbb{P}((\mathbb{T} \times \mathbb{R}) \times (\mathbb{T} \times \mathbb{R})).$$
Our final goal is to derive some Gönwal-type inequality for $I$. Fix $(\theta_1, \omega_1), (\theta_2, \omega_2) \in T \times \mathbb{R}$ and define the following curves in $T$
\[ \Theta(t) := \Theta_{0,t}(\theta_1, \omega_1) \quad \text{and} \quad \Theta(t) := \Theta_{0,t}(\theta_2, \omega_2), \]
and the associated characteristic curves in $T \times \mathbb{R}$,
\[ X(t) := X_{0,t}(\theta_1, \omega_1) = (\Theta(t), \omega_1), \]
\[ \overline{X}(t) := \overline{X}_{0,t}(\theta_2, \omega_2) = (\Theta(t), \omega_2). \]
Set a minimizing geodesic $x_t : [0, 1] \to T \times \mathbb{R}$ joining $X(t)$ to $\overline{X}(t)$, for every fixed $t > 0$. Notice that the following function
\[ t \mapsto \frac{1}{2} d_K^2(X(t), \overline{X}(t)), \]
is Lipschitz continuous. Then, we can take derivatives and show that
\[
\frac{d}{dt} \frac{1}{2} d_K^2(X(t), \overline{X}(t)) \leq - \langle (v[f_t](X(t))), 0 \rangle - \langle (v[f_t](\overline{X}(t))), 0 \rangle, 
\]
for almost every $t > 0$. Let us now consider $\theta(t) := \overline{\Theta}(t) - \Theta(t)$, the representative of $\overline{\Theta}(t) - \Theta(t)$ modulo $2\pi$ that lies in $(-\pi, \pi]$. We find two different cases:
\begin{itemize}
  \item **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, the (6.5) reads
    \[
    \frac{d}{dt} \frac{1}{2} d_K^2(X(t), \overline{X}(t)) \leq (v[f_t](\overline{\Theta}(t)), \omega_2) - v[f_t](\Theta(t), \omega_1))\theta(t),
    \]
    for almost every $t \geq 0$.
  \item **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 follows
    \[
    \frac{d}{dt} \frac{1}{2} d_K^2(X(t), \overline{X}(t)) \leq (v[f_t](\overline{\Theta}(t)), \omega_2) - v[f_t](\Theta(t), \omega_1)))(\pm \pi),
    \]
    for almost every $t \geq 0$. To sum up, we achieve the following estimate
    \[
    \frac{d}{dt} \frac{1}{2} d_K^2(X_{0,t}(\theta_1, \omega_1), \overline{X}_{0,t}(\theta_2, \omega_2))
    \leq (v[f_t](\Theta_{0,t}(\theta_1, \omega_1)) - v[f_t](\overline{\Theta}_{0,t}(\theta_2, \omega_2)), \omega_1) - \overline{\Theta}_{0,t}(\theta_2, \omega_2),
    \]}
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{equation}
\frac{dI}{dt} \leq \int_{\mathbb{T} \times \mathbb{R}} \int_{\mathbb{T} \times \mathbb{R}} (v[f_1](\Theta_{0,t}(\theta_1, \omega_1), \omega_1) - v[\tilde{f}_1](\Theta_{0,t}(\theta_2, \omega_2), \omega_2)) \times \Theta_{0,t}(\theta_1, \omega_1) - \overline{\Theta}_{0,t}(\theta_1, \omega_1) \, d((\theta_1, \omega_1), (\theta_2, \omega_2)) \, \gamma_0,
\end{equation}

for almost every $t \geq 0$. Also, note that

\begin{align*}
v[f_1](\theta, \omega) &= \omega - K \int_{\mathbb{T} \times \mathbb{R}} \sin(\theta - \Theta_{0,t}(\theta_1, \omega_1)) \, d\bar{\theta}_1 \bar{\omega}_1 \tilde{f}_0, \\
v[\tilde{f}_1](\theta, \omega) &= \omega - K \int_{\mathbb{T} \times \mathbb{R}} \sin(\theta - \overline{\Theta}_{0,t}(\theta_2, \omega_2)) \, d\bar{\theta}_2 \bar{\omega}_2 \tilde{f}_0.
\end{align*}

Since $(\pi_1)\# \gamma_0 = f_0$ and $(\pi_2)\# \gamma_0 = \tilde{f}_0$, then

\begin{align*}
v[f_1](\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\bar{\theta}_1' \bar{\omega}_1' \gamma_0, \\
v[\tilde{f}_1](\theta, \omega) &= \omega - K \int_{\mathbb{T} \times \mathbb{R}} \int_{\mathbb{T} \times \mathbb{R}} \sin(\theta - \overline{\Theta}_{0,t}(\theta_2', \omega_2')) \, d\bar{\theta}_2' \bar{\omega}_2' \gamma_0.
\end{align*}

Putting (6.7) into (6.6) amounts to

\begin{equation}
\frac{dI}{dt} \leq \int_{(\mathbb{T} \times \mathbb{R})^4} (\omega_1 - \omega_2) \Theta_{0,t}(\theta_1, \omega_1) - \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 \\
- K \int_{(\mathbb{T} \times \mathbb{R})^4} (\sin(\Theta_{0,t}(\theta_1, \omega_1) - \Theta_{0,t}(\theta_1', \omega_1')) - \sin(\overline{\Theta}_{0,t}(\theta_2, \omega_2) - \overline{\Theta}_{0,t}(\theta_2', \omega_2'))) \times \Theta_{0,t}(\theta_1, \omega_1) - \overline{\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{equation}

for almost every $t \geq 0$. By Young’s inequality, it is clear that

\begin{equation}
(\omega_1 - \omega_2) \Theta_{0,t}(\theta_1, \omega_1) - \Theta_{0,t}(\theta_2, \omega_2) \leq K \Theta_{0,t}(\theta_1, \omega_1) - \Theta_{0,t}(\theta_2, \omega_2)^2 + \frac{(\omega_1 - \omega_2)^2}{2K} = \frac{K}{2} d_K(\mathcal{X}_{0,t}(\theta_1, \omega_1), \mathcal{X}_{0,t}(\theta_2, \omega_2))^2.
\end{equation}

This, along with a clear symmetrization argument in the second term implies

\begin{equation}
\frac{dI}{dt} \leq K I(t)
\end{equation}

\begin{align*}
- \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(\overline{\Theta}_{0,t}(\theta_2, \omega_2) - \overline{\Theta}_{0,t}(\theta_2', \omega_2'))) \times \left( \Theta_{0,t}(\theta_1, \omega_1) - \overline{\Theta}_{0,t}(\theta_2, \omega_2) - \Theta_{0,t}(\theta_1', \omega_1') - \overline{\Theta}_{0,t}(\theta_2', \omega_2') \right) \\
& \times d((\theta_1, \omega_1), (\theta_2, \omega_2)) \, \gamma_0 \, d((\theta_1', \omega_1'), (\theta_2', \omega_2')) \gamma_0
\end{align*}
for almost every $t \geq 0$. Now, using the Lipschitz property of the sine function we achieve the inequality
\[
\frac{dI}{dt} \leq (K + 4K)I, \quad \text{for a.e. } t \geq 0.
\]
Integrating the inequality and using that
\[
I(0) = \int_{\mathbb{T} \times \mathbb{R}} \int_{\mathbb{T} \times \mathbb{R}} \frac{1}{2} d_{\mathcal{K}}((\theta_1, \omega_1), (\theta_2, \omega_2))^2 d_{\mathcal{H}}((\theta_1, \omega_1), (\theta_2, \omega_2)) \gamma_0 = \frac{1}{2} SW_2(f_0, \bar{f}_0)^2,
\]
yields the desired result. \(\square\)

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 $\phi_\infty$ of $f_\infty$. Recall that by Corollary 2.13 we obtain
\[
R_\infty = \lim_{t \to \infty} R(t) \geq 3/5.
\]
Looking at the structure of the stable equilibria $f_\infty$ in (1.14) (that corresponds to $g^- = 0$, that is, no antipodal mass), we observe that for any $(\theta, \omega)$ in supp$f_\infty$ we have the relation
\[
\theta = \phi_\infty + \arcsin\left(\frac{\omega}{KR_\infty}\right).
\]
In particular,
\[
|\theta - \phi_\infty| \leq \arcsin\left(\frac{W}{KR_\infty}\right) \leq \arcsin\left(\frac{5W}{3K}\right).
\]
Then, we can select $C$ in (1.15), so that we have that
\[
(6.10) \quad \text{supp } f_\infty \subseteq L_{\frac{1}{2}} \times [-W, W].
\]
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 $\mu_t^N$ and $f_t$ in $\mathbb{T} \setminus L$, namely,
\[
(6.11) \quad \mu_t^N((\mathbb{T} \setminus L) \times \mathbb{R}) \leq 25 SW_2(\mu_t^N, f_\infty)^2,
\]
\[
(6.12) \quad \rho_t(\mathbb{T} \setminus L) \leq 25 SW_2(f_t, f_\infty)^2,
\]
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 $\mu_t^N$ and $f_\infty$ for the scaled Wasserstein distance $SW_2$. Then, we have that

$$SW_2(\mu_t^N, f_\infty)^2 = \int_{(\mathbb{T} \times \mathbb{R})^2} dK((\theta, \omega), (\theta', \omega'))^2 d\gamma_t$$

$$\geq \int_{((\mathbb{T} \setminus L) \times \mathbb{R}) \times (L_{1/2} \times \mathbb{R})} d(\theta, \theta')^2 d\gamma_t$$

$$\geq \frac{1}{25} \gamma_t((\mathbb{T} \setminus L) \times \mathbb{R}) \times (L_{1/2} \times \mathbb{R}))$$

$$= \frac{1}{25} \gamma_t(((\mathbb{T} \setminus L) \times \mathbb{R}) \times ((\mathbb{T} \setminus L_{1/2}) \times \mathbb{R}))$$

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 $\mu_t^N$ replaced with $f_t$, we deduce that (6.12).

**Step b:** We claim that we can select $T_0$ satisfying that

$$(6.13) \quad T_0 \lesssim \frac{1}{K R_0^2} \log \left(1 + W^1/2 ||f_0||_2 + \frac{1}{R_0}\right),$$

and with the additional property that

$$(6.14) \quad SW_2(f_t, f_\infty) \leq \frac{1}{\sqrt{500}} e^{-\frac{1}{2} t K (t - T_0)},$$

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

To show this, take $Q_1$ large enough and $T_0$ verifying

$$T_0 \leq \frac{Q_1}{K R_0^2} \log \left(1 + W^1/2 ||f_0||_2 + \frac{1}{R_0}\right),$$

so that we meet the constraints in Theorem 1.1. Then, using (3.24) and Proposition 3.2, we obtain that

$$(6.15) \quad SW_2(f_t, f_\infty) \leq Q_2 e^{-\frac{1}{2} t K (t - T_0)},$$

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

$$(6.16) \quad \mathbb{P} \left( SW_2(\mu_t^N, f_t) \leq \frac{1}{\sqrt{500}} e^{-\frac{1}{2} t K (t - T_0)} \right) \geq 1 - C_1 e^{-C_2 N^{1/2}},$$

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

$$(6.17) \quad \varepsilon_N := N^{-1/2}.$$
Now, we define $N^*$ as follows

\[ N^* := \min \left\{ N \in \mathbb{N} : \varepsilon_N e^{\frac{5K T_0}{2}} \leq \frac{1}{\sqrt{500}} \right\}, \tag{6.18} \]

so that, by definition, we get the bound

\[ N^* \geq 500^4 e^{20KT_0}. \]

Fix any $N \geq N^*$. Notice that $N^*$ has been defined in (6.18) so that there exists $d_N > 0$ with the property

\[ \varepsilon_N e^{\frac{5K}{2}(T_0 + d_N)} = \frac{1}{\sqrt{500}} e^{-\frac{1}{40}Kd_N}, \tag{6.19} \]

Indeed, by dividing (6.19) over (6.18), we can quantify $d_N$ in terms of $N^*$ as follows

\[ \frac{\varepsilon_N}{\varepsilon_{N^*}} e^{\frac{5K}{2} d_N} \geq e^{-\frac{1}{40}Kd_N}. \]

Consequently, we have that

\[ d_N \geq \frac{5}{101K} \log \frac{N}{N^*}. \]

By construction, letting $\varepsilon = \varepsilon_N$ in the concentration inequality (6.3) of Lemma 6.2, we obtain the following quantification

\[ P (SW_2(\mu_0^N, f_0) \geq \varepsilon_N) \leq C_1 e^{-C_2N^{\frac{1}{2}}}, \tag{6.20} \]

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 that

\[ C_1 e^{-C_2N^{\frac{1}{2}}} \geq P (SW_2(\mu_t^N, f_t) \geq \varepsilon_N) \geq P (SW_2(\mu_t^N, f_t) \geq \varepsilon_N e^{\frac{5K}{2}(T_0 + d_N)}) \]

\[ \geq P (SW_2(\mu_t^N, f_t) \geq \frac{1}{\sqrt{500}} e^{-\frac{1}{40}Kd_N}) \]

\[ = P (SW_2(\mu_t^N, f_t) \geq \frac{1}{\sqrt{500}} e^{-\frac{1}{40}K(t - T_0)}), \]

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 remaining ones 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).

\[ \bullet \text{ Step d: We quantify the probability of mass concentration of } \mu_t^N \text{ in the interval } L, \text{ namely,} \]

\[ P \left( \mu_t^N (L \times \mathbb{R}) \geq 1 - \frac{1}{5} e^{-\frac{1}{40}K(t - T_0)} \right) \geq 1 - C_1 e^{-C_2N^{\frac{1}{2}}}, \tag{6.21} \]

for every $t \in [T_0, T_0 + d_N]$ and any $N \geq N^*$. 

Now, by \((6.11), (6.14)\) and triangular inequality we have that
\[
\mu_t^N((T \setminus L) \times \mathbb{R}) \leq 25 SW_2(\mu_t^N, f_\infty)^2 \\
\leq 50 \left[ SW_2(\mu_t^N, f_t)^2 + SW_2(f_t, f_\infty)^2 \right] \\
\leq 50 \left[ SW_2(\mu_t^N, f_t)^2 + \frac{1}{500} e^{-\frac{1}{25}K(t-T_0)} \right],
\]
for every \(t \in [T_0, T_0 + d_N]\). Hence, we obtain
\[
\mu_t^N(L \times \mathbb{R}) \geq 1 - \frac{1}{10} e^{-\frac{1}{25}K(t-T_0)} - 50 SW_2(\mu_t^N, f_t)^2,
\]
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 \(\mu^N\) (recall Definition \(6.1)\) so that the condition within \((6.21)\) holds. Hence, by construction, we obtain that at such realization
\[
p := \inf_{\theta, \theta' \in L} \cos(\theta - \theta') \geq \frac{4}{5} \quad \text{and} \quad m := \mu_s^N(L \times \mathbb{R}) \geq 1 - \frac{1}{5} e^{-\frac{1}{25}K(s-T_0)} \geq \frac{4}{5}.
\]

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 the assumption \((1.15)\) in Theorem \([1.1]\) guarantee the condition \((4.5)\) within the hypotheses of Lemma \([4.2]\). Notice that such result also holds true for the particle system. Consequently, it asserts that for such realization of \(\mu^N\) we can consider a time-dependent interval \(L_s^N(t)\) with \(t \geq s\) so that \(L_s^N(s) = L\) and
\[
\mu_t^N(L_s^N(t) \times \mathbb{R}) \geq 1 - \frac{1}{5} e^{-\frac{1}{25}K(s-T_0)},
\]
\[
1 - \inf_{\theta, \theta' \in L_s^N(t)} \cos(\theta - \theta') \leq \max \left\{ \frac{1}{5} e^{-\frac{1}{25}k(t-s)}, \frac{25 W^2}{K^2} \right\},
\]
for any \(t \geq s\). Indeed, we have that \(L_s^N(t) = \pi_\theta(X_s^N(L \times [-W, W]))\), where \(X_s^N\) represents the flow of the particle system, that is, the flow of \(\mu^N\). Our final goal is to simplify the last condition in \((6.22)\). To such an end, let us consider \(D_s^N(t) := \text{diam}(L_s^N(t))\) and notice that such inequality implies that
\[
2 \left(\frac{D_s^N(t)}{5}\right)^2 \leq 1 - \cos(D_s^N(t)) \leq \max \left\{ \frac{1}{5} e^{-\frac{1}{25}k(t-s)}, \frac{25 W^2}{K^2} \right\},
\]
for any \(t \geq s\). In particular, we obtain \((12)\). Thus, Corollary \([1.1]\) follows.
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, Birkhäuser, Basel, 2008.

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.

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.

5. D. Benedetto, E. Caglioti, and U. Montemagno, On the complete phase concentration for the Kuramoto model in the mean field limit, Commun. Math. Sci. 13 (2015), no. 7, 1775–1786.

6. E. Boissard, Problèmes d’interaction discret-continu et distances de Wasserstein, Ph.D. thesis, Université de Toulouse III, December 2011.

7. , Simple bounds for convergence of empirical and occupation measures in $L^1$-Wasserstein distance, Electron. J. Probab. 16 (2011), no. 83, 2296–2333.

8. F. Bolley, A. Guillin, and C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, Probab. Theory Relat. Fields 137 (2007), no. 3–4, 541–593.

9. M. Breakspear, S. Heitmann, and A. Daffertshofer, Generative models of cortical oscillations: neurobiological implications of the Kuramoto model, Front. Hum. Neurosci. 4 (2010), 190.

10. J. A. Carrillo, Y.-P. Choi, S.-Y. Ha, M.-J. Kang, and Kim Y., Contractivity of transport distances for the kinetic Kuramoto equation, J. Stat. Phys. 156 (2014), no. 2, 395–415.

11. J. A. Carrillo, F. James, F. Lagoutièere, and V. Vauchelet, The Filippov characteristic flow for the aggregation equation with mildly singular potentials, J. Differential Equations 260 (2016), no. 1, 304–338.

12. Y.-P. Choi, S.-Y. Ha, S. Jung, and Y. Kim, Asymptotic formation and orbital stability of phase-locked states for the Kuramoto model, Physica D 241 (2012), no. 7, 735–754.

13. 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.

14. H. Dietert, Stability and bifurcation for the Kuramoto model, J. Math. Pures Appl. 105 (2016), no. 4, 451–489.

15. , Stability of partially locked states in the Kuramoto model through Landau damping with Sobolev regularity, 2017, arXiv:1707.03475.

16. H. Dietert and Fernandez B., The mathematics of asymptotic stability in the Kuramoto model, Proc. R. Soc. A 474 (2018), 20180467.

17. H. Dietert, B. Fernandez, and D. Gérard-Varet, Landau Damping to Partially Locked States in the Kuramoto Model, Commun. Pure Appl. Math. 71 (2018), no. 5, 953–993.

18. R. Dobrushin, Vlasov equations, Funct. Anal. Appl. 13 (1979), no. 2, 115–123.

19. N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probab. Theory Relat. Fields 162 (2015), no. 3–4, 707–738.

20. S.-Y. Ha, T. Y. Ha, and J.-H. Kim, On the complete synchronization of the Kuramoto phase model, Physica D 239 (2010), no. 17, 1692–1700.

21. 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.

22. S.-Y. Ha, Y.-H. Kim, J. Morales, and J. Park, Emergence of phase concentration for the Kuramoto–Sakaguchi equation, Physica D (2019), doi:10.1016/j.physd.2019.132154.

23. 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.

24. 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.

25. R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation, J. Stat. Phys. 29 (1989), no. 1–2, 1–17.

26. Y. Kuramoto, Self-entrainment of a population of coupled non-linear oscillators, International Symposium on Mathematical Problems in Theoretical Physics (H. Araki, ed.), Lecture Notes in Physics, vol. 30, Springer-Verlag, Kyoto, Japan, 1975, pp. 429–422.

27. , Chemical Oscillations, waves and turbulence, Springer-Verlag, Berlin, 1984.
28. C. Lancellotti, *On the Vlasov limit for systems of nonlinearly coupled oscillators without noise*, Transport Theor. Stat. Phys. **34** (2005), 523–535.

29. Z. Li, X. Xue, and D. Yu, *On the lojasiewicz exponent of Kuramoto model*, J. Math. Phys. **56** (2015), no. 2, 022704.

30. S. Lojasiewicz, *Une propriétée topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, vol. 117, Coll. du CNRS, Paris, 1962, pp. 87–89.

31. S. C. Manrubia, A. S. Mikhailov, and D. H. Zanette, *Emergence of Dynamical Order. Synchronization Phenomena in Complex Systems*, World Scientific Lecture Notes in Complex System, vol. 2, World Scientific, 2004.

32. J. Morales, *Least action principles with applications to gradient flows and kinetic equations*, Ph.D. thesis, The University of Texas at Austin, May 2017.

33. H. Neunzert, *An introduction to the nonlinear Boltzmann–Vlasov equation*, Kinetic Theories and the Boltzmann Equation (C. Cercignani, ed.), Lecture Notes in Mathematics, vol. 1048, Springer, Berlin, Heidelberg, 1984, pp. 60–110.

34. F. Otto, *Evolution of microstructure in unstable porous media: A relaxational approach*, Comm. Pure Appl. Math. **52** (1999), no. 7, 873–915.

35. F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Part. Differ. Equat. **26** (2001), no. 1–2, 101–174.

36. 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.

37. A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A universal concept in nonlinear sciences*, Cambridge University Press, Cambridge, 2001.

38. D. Poyato, *Filippov flows and mean-field limits in the kinetic singular Kuramoto model*, 2019, arXiv:1903.01305.

39. S. H. Strogatz, *From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators*, Physica D **143** (2000), no. 1–2, 1–20.

40. 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.

41. V. S. Varadarajan, *On the convergence of sample probability distributions*, Sankhya **19** (1958), no. 1–2, 23–26.

42. S. R. S. Varadhan, *Probability theory*, Courant Lecture Notes, vol. 7, American Mathematical Society, Basel, 2001.

43. C. Villani, *Optimal transport: old and new*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin, Heidelberg, 2009.

44. P. Villegas, P. Moretti, and M. A. Muñoz, *Frustrated hierarchical synchronization and emergent complexity in the human connectome network*, Sci. Rep. **4** (2014), 5990.

(Javier Morales) CSCAMM, 4141 CSIC 8169 Paint Branch Drive University of Maryland College Park, MD 20742-3289

E-mail address: javierm1@cscamm.umd.edu

(David Poyato) Departamento de Matemática Aplicada and Research Unit “Modeling Nature” (MNAT), Universidad de Granada, Granada, 18071, Spain

E-mail address: davidpoyato@ugr.es