ABSTRACT

Title of dissertation: Dispersion Properties of Transport Equations and Applications

Hsin-Yi Lin
Doctor of Philosophy, 2020

Dissertation directed by: Professor Pierre-Emmanuel Jabin
Department of Mathematics

The concept of transport is fundamental and has great influence in a wide range of fields across science. This dissertation provides three topics possessing the character of transport phenomena from the perspective of partial differential equations. The three parts include:

(1) Commutator method for averaging lemma: A new commutator method is introduced to prove a new type of averaging lemmas, the regularizing effect for the velocity average of solutions for kinetic equations. This novel approach shows a new range of assumptions that are sufficient for the velocity average to be in $L^2([0,T], H_x^{1/2})$ and improves the regularity result for the measure-valued solutions of scalar conservation laws in space one-dimensional case.

(2) Unmixing property of incompressible flows on 2d tori: The local Hamiltonian structure of a 2d torus is utilized to show that the unmixing property of incompressible flows can be preserved under a sup-norm perturbation on stream functions. With this perturbation result, a quantitative statement was provided

by considering vector fields in the form of a random Fourier series. This statement offers an interesting observation for the unmixing property from the perspective of Fourier analysis.

(3) Memory effect on animal migration: The goal of this work is to obtain a better understanding of the memory effect on the animals' migration patterns under periodic environments. A memory model and a corresponding memory-driven dynamic were constructed. Through simulations, it is discovered that in order to have periodic movement, the individual must be able to gather and carry sufficient information from both short-term memory and long-term memory, and possess the ability to discriminate which information is more important with appropriate time scales. Furthermore, our mathematical model is general and can be used to test the memory effect under different circumstances. Several interesting examples are demonstrated.

Dispersion Properties of Transport Equations and Applications

by

Hsin-Yi Lin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Pierre-Emmanuel Jabin, Chair/Advisor
Professor William Fagan, Dean's representation
Professor Matei Machedon
Professor Eitan Tadmor
Professor Konstantina Trivisa

Acknowledgments

It has been my pleasure to learn from and work with the most amazing people in CSCAMM and Department of Mathematics at University of Maryland.

My sincere and deep gratitude goes first and foremost to my advisor, Professor Pierre-Emmanuel Jabin for his invaluable and kind guidance on several intriguing and challenging projects. His creative thinking and intuition in mathematics have taught me a lot, which are extremely helpful in partial differential equations. I would also like to thank Professor Matei Machedon and Professor Dmitry Dolgopyat for sharing with me their deep insights for my works from different perspectives.

During this program, I have the honor to work with Professor Eitan Tadmor on the commutator method for averaging lemmas and Professor William Fagan on the memory effect on animal migration. I would like to express my gratitude for all the knowledge they are willing to share generously.

I am grateful for my family and good friends: Wenbo Li, Siddharth Sharma, Pranava Jayanti, Michael Rawson, Addison Bohannon, David Russell, Ruiwen Shu, Christopher Dock, Shujie Kang, Zhao Liu, Ke Xue, Yue Fan, Tengfei Su, Martin Fructuoso, Luyu Sun, Jing Zhou, Xin Dong. I apologize if there are indivduals I fail to recall and are inadvertently left out. Their companions and supports have been meaningful to me.

I would also like to acknowledge the help and support from some of the staff members: Cristina Garcia, Anil Zenginoglu, Agi Alipio, Sara Moran. They are the most accommodating and thoughtful persons and I sincerely appreciate their assistance.

This graduate experience in the University of Maryland shall be cherished forever with my deepest gratitude to all that made this thesis possible.

Table of Contents

Chapter 1: Background and overview

1.1 Outline of dissertation

This dissertation contains three parts, each of which addresses a seemingly different direction while they are all centered around a common topic of transport phenomena. One of the classical fields that analyzes transport phenomena is the kinetic theory. This thesis can be deemed as a modest approach to demonstrate the wide influence and essence that kinetic theory could have.

The first part studies the averaging lemmas for kinetic transport equations. The second part discusses the mixing property of incompressible flows on a 2-dimensional torus. Both averaging effect and flow mixing are regarded as a dispersive property of transport operators. The averaging effect is a dispersion in phase space. On the other hand, the flow mixing is a dispersive property in physical space.

In addition to the kinetic models and mixing of flows, the transport equations are found to be useful in biological modeling. The last part of this dissertation illuminates such topic, where the narrative turns slightly into modeling rather than classical style of analysis. This part of work is to render the application of transport equations to biology by investigating memory effects on animal migrations. One source of the ideas for our model origins from the optimal control theory.

The dissertation is organized as the following: (1) Introduce the backgrounds and characterize the author's contributions in Chapter 1. (2) In Chapter 2, averaging lemmas are described in details. (3) Discussion of flow mixing is in Chapter 3. (4) Modeling for memory effects on animal migrations is in Chapter 4.

1.2 Averaging effect in kinetic transport equations

1.2.1 Brief introduction to kinetic theory

The kinetic theory provides certain mathematical models to describe the dynamics of a large collection of particles, represented by density functions in phase space. The phase space includes both microscopic information (states) of the particle, as well as the macroscopic variables, such as position. In between microscopic and macroscopic models, the kinetic equations are sometimes called *mesoscopic* models. Due to the long history of development and extensive literature in kinetic theory, it is impossible to cover every aspect without losing focus. For this short introduction, three essential types of classical kinetic equations and their solutions shall be briefly discussed.

The simplest kinetic model is the free transport equation:

$$\partial_t f + v \cdot \nabla_x f = 0, \tag{1.1}$$

where x represents the position, v the velocity, and f is the density function of particles. This equation describes that each particle in the system travels at a constant velocity v. The solution of (1.1) can be expressed by the characteristic line $x - vt$:

$$f(t, x, v) = f_0(x - vt),$$

where f_0 is the density function at time zero.

When a force F is taken into consideration, the particles no longer travel in a straight line, and hence f follows a different equation:

$$\partial_t f + v \cdot \nabla_x f + div_v(Ff) = 0, \tag{1.2}$$

which is called the linear Vlasov equation. Two classical examples are the Vlasov-Poisson and Vlasov-Maxwell system. The Vlasov-Maxwell system characterizes charged particles in an electromagnetic field, and the Vlasov-Poisson system describes particles under the gravitational force. Both systems are modeled by (1.2), where F is coupled with the Poisson equations for the Vlasov-Poisson system, and with the Maxwell equations in the Vlasov-Maxwell system.

Both (1.1) and (1.2) do not consider the collisions between particles. The presence of collision can be modeled by adding the Boltzmann's quadratic collision operator $Q(f, f)$, and the density function f in this case satisfies the Boltzmann equation:

$$\partial_t f + v \cdot \nabla_x f + div_v(Ff) = Q(f, f). \tag{1.3}$$

Informally speaking, $Q(f, f) = \frac{\partial f}{\partial t}|_{collision}$ represents the "rate of particle number change" within a volume element. For detailed discussion for the Boltzmann collision operator, we refer to [103].

The proof of existence of weak solutions for the Vlasov-Poisson system is provided in [8], and the Vlasov-Maxwell system in [38]. A classical approach to obtain the weak solution of a nonlinear system is to first consider a modified problem, and then pass to the limit with a uniform a priori estimate. For instance, the elliptic regularity of the Poisson equation could be utilized to derive a priori estimate for the Vlasov-Poisson system. As for the Maxwell equation, one would need to use another method called the *velocity averaging*. Roughly speaking, an averaging lemma states that the macroscopic quantity, which is expressed as a velocity average of f, is smoother than f itself. This gain of regularity provides compactness for such averages and hence help obtain weak solutions.

The existence of global solutions for (1.3) has been proven in [39]. One of the obstacles one encounters in the Boltzmann equation is the lack of a priori estimate, and the condition on f is too weak to even apply averaging lemmas. To overcome this issue, a renormalization formulation was introduced in [39]. After renormalization, velocity averaging can be applied and an approximation argument is then used to reach the desired compactness.

The Boltzmann equation can be regarded as a stepping stone between the molecule dynamics to a continuum description. The idea that the Boltzmann equation can be considered as a consequence of Newton's laws was first proposed in [58], and the derivation of the incompressible Navier-Stokes equations from the Boltzmann equation has been proved in [54], where an averaging lemma in L^1 was shown and played a critical part in their main proof.

We have presented several important works that utilize averaging lemmas to prove the existence of solutions. For modern kinetic theory, averaging lemmas has become an indispensable method for studying transport equations. This shall be the focus for the rest of our discussion.

1.2.2 Averaging lemmas of kinetic transport equations

Consider the kinetic equations of the following form:

$$\varepsilon \partial_t f + a(v) \cdot \nabla_x f = (-\Delta_v)^{\alpha/2} g, \tag{1.4}$$

where $\varepsilon > 0$, $\alpha \geq 0$, $a : \mathbb{R}^n_v \to \mathbb{R}^n$ and $g : \mathbb{R}_t \times \mathbb{R}^n_v \times \mathbb{R}^n_x \to \mathbb{R}$ are given functions. While one often considers $a(v) = v$ for classical kinetic models, the needs for nonlinear $a(v)$ appear naturally from the kinetic formulation of scalar conservation law, as well as the kinetic models for the relativistic and quantum settings [42, 52].

Because the kinetic transport equations are of hyperbolic type, the solutions in general cannot be more regular than its initial conditions. Even so, it has been discovered that the velocity averages of the solution f of (1.4)

$$\rho_\phi(t, x) := \int f(t, x, v)\, \phi(v)\, dv, \quad \phi \in L^\infty_c \tag{1.5}$$

gain regularities in space variable, where L^∞_c is the space containing all the bounded and compactly supported functions. This type of results are called the *averaging lemmas.*

Classical averaging lemmas consider the case when f and g in (1.4) are both in one same L^p space. The L^2 setting was first considered and introduced independently in [2] and [51].

The derivation for the L^2 case usually involves the techniques in Fourier analysis. Consider the Fourier transform of the transport operator in x:

$$\mathcal{F}_{x\to\xi}\left(a(v)\cdot\nabla_x\right) = \left(a(v)\cdot\frac{\xi}{|\xi|}\right)\cdot|\xi|.$$

Informally speaking, the transport operator "possesses" a regularity of order 1 inside any good region $\left\{v : \left(a(v)\cdot\frac{\xi}{|\xi|}\right) > \alpha\right\}$ for any fixed ξ and $\alpha > 0$. The hope to rigorously prove a gain of regularity is to somehow control the singular part $\left\{v : \left(a(v)\cdot\frac{\xi}{|\xi|}\right) \le \alpha\right\}$, which leads to the need of conditions on $a(v)$. In fact, if there exists one direction $\sigma \in \mathbb{S}^{d-1}$ such that $a(v)\cdot\sigma = 0$ for all v, no additional regularity can be gained from the velocity averaging. For example when $g = 0$, consider $f = \phi(x\cdot\sigma)$ for some smooth ϕ.

To exclude this type of counterexamples, a condition on the measure of the singular part for all directions is needed. A classical assumption used by previous literature is defined as follows: There exists a constant c_0 such that

$$\mathcal{L}^n\left(\{v\in D : |a(v)\cdot\sigma| \le \alpha/2\}\right) \le c_0\,\alpha^\nu, \tag{1.6}$$

for all compact set $D \subset \mathbb{R}^n$ and $\sigma \in \mathbb{S}^{m-1}$, where $\mathcal{L}^n$ is the Lebesgue measure in $\mathbb{R}^n$. (1.6) is called the *non-degeneracy condition* for time-independent kinetic equations. The time-dependent case can be discussed similarly by considering the variable (t, x) and the coefficient $(1, a(v))$. With change of variables, (1.6) can be rewritten as

$$\mathcal{L}^n \left(\{ v \in D : |a(v) \cdot \sigma - \tau| \leq \alpha/2 \} \right) \leq c_0 \, \alpha^\nu, \tag{1.7}$$

where $\tau \in \mathbb{R}$. With (1.6) or (1.7), the averaging lemma in L^2 case can be concluded by decomposition in Fourier space and careful estimations on the singular parts.

The general L^p results, $1 < p < \infty$, are more delicate and one needs other techniques from harmonic analysis and interpolation arguments to reach a L^p result when $p \neq 2$ [18, 40]. The optimal Besov results has also been proved in [35] by using wavelet decomposition.

One difficulty one could encounter when proving L^p results with interpolations is that the limiting L^1 case in general is not true. A counterexample has been given in [51]. A classical way to bypass this issue is to replace the L^1 space by a Hardy space of product spaces [18, 40].

Although the gain of regularity in L^1 case does not hold, the compactness in L^1 can still be proved with an equi-integrability assumption just in v variable [56].

More recent developments study the averaging lemmas with f and g in different spaces. For instance, the case that f and g in the same Besov space in x but admits different integrability in v was investigated in [105]. Later general mixed norms assumptions were considered in [67,68]. Their work inspired [10] to consider the case when f and g have less integrability in x than v.

In a joint work with P.-E. Jabin and E. Tadmor, we introduce a *commutator method* to prove a new type of averaging lemmas when f and g are in general L^p spaces. The goal of the next subsection is to briefly introduce this work.

1.2.3 Commutator method for averaging lemmas

The goal of our work is to introduce a *commutator method* as a novel approach to averaging lemmas for (1.4). Commutator methods have been used for example in the studies for equations of Schrödinger type, where the commutator appear naturally from the Hamilton vector field; See for instance [30, 41, 72, 97].

One simple special case of our main result (Theorem 2) when $a(v) = v$ and $\alpha = 0$ can be stated as follows: For any $\varepsilon \le 1$, if $f \in L^\infty([0,T], L^p_{loc}(\mathbb{R}^n_x \times \mathbb{R}^n_v))$ solves

$$\partial_t f + v \cdot \nabla_x f = g, \tag{1.8}$$

for some $g \in L^1([0,T], L^q_{loc}(\mathbb{R}^n_x \times \mathbb{R}^n_v))$, where $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $\phi \in L^\infty_c$, $\rho_\phi \in L^2\left([0,T], H_{loc}^{\frac{1}{2(1+\alpha)}}\right)$.

Roughly speaking, our proof involves extracting the dispersion of transport operator by integrating along the characteristics in Fourier space. The dispersion is then transformed into a gain of regularity through the commutator of the transport operator and a deliberately selected multiplier operator. The basic structure of our argument can be formulated in the following:

Denote the free transport operator $v \cdot \nabla_x$ by B. Then for any time-independent operator Q, one has

$$\varepsilon \partial_t \int f \, \overline{Qf} \, dx \, dv = \int [B, Q] \, f \bar{f} \, dx \, dv + \int g \, \overline{Qf} \, dx \, dv + \int f \, \overline{Qg} \, dx \, dv, \tag{1.9}$$

where f solves (1.8). After integrating (1.9) over t, a bound for the commutator term reads

$$Re \int_0^T [B, Q]\, f \bar{f}\, dx\, dv\, dt \leq \sup_{t=0,T} \left| \int f\, \overline{Qf}\, dx\, dv \right| + \left| \int g\, \overline{Qf}\, dx\, dv\, dt \right| + \left| \int f\, \overline{Qg}\, dx\, dv\, dt \right| . \tag{1.10}$$

The idea is to find an operator Q which is bounded in some L^p spaces such that the commutator $[B, Q]$ is positive-definite and possesses extra regularity. If such an operator Q is found, (1.10) implies the gain of regularity of f.

In our work, the selected Q is a multipler operator, which is inspired by the multiplier method introduced in [76]. This multiplier method was used to prove moment lemmas for kinetic equations [91] and the local smoothing properties for dispersive equations through the Wigner transform [50].

The connection between moment lemmas and our result can be perceived through the Fourier transform in (x, v). Let (ξ, ζ) be the frequency dual of (x, v). Note that the Fourier transform of transport operator $v \cdot \nabla_x$ is again a transport operator $\xi \cdot \nabla_\zeta$ but in the frequency space. The moment lemma for $\xi \cdot \nabla_\zeta$ in frequency space then leads to a gain of regularity for $v \cdot \nabla_x$ in physical space.

When the special case above is extended to the general $a(v)$ and $\alpha > 0$, difficult technical issues emerge for our argument. Our proof overcomes these issues by using a change of variables and a regularization process, which has a connection with the renormalization introduced in [37].

Our method not only provides a new type of regularity results, but also has some nice features that could be beneficial for certain applications. One important feature that distinguishes our results from others in the literature is that the integrability of f and g can be of assistance to each other. Due to this feature, our method leads to exciting, novel results, especially when $p \geq 2$ and $\frac{1}{p} + \frac{1}{q} = 1$. Moreover, this type of duality condition fits nicely with the kinetic formulation of scalar conservation law. This will be an example of applications of our main result.

Another advantage of our result is the ε independence in (1.4). This could have applications to the compactness of solutions for rescaled kinetic equations, which frequently appear in the discussions of hydrodynamic limits. Moreover, unlike the classical approach, our argument does not perform the Fourier transform in the time variable, and hence possible extensions for time discretized kinetic equations or stochastic cases may be considered. For the existed works for averaging lemmas of time discretized kinetic equations, see for example [19] and [79, 80] for stochastic kinetic equations.

1.3 Mixing behaviors

1.3.1 Mixing from the analysis point of view

Mixing has attracted abundant research interests in various mathematical fields, such as dynamical systems, probability, and PDEs. Here our focus is on the mixing behavior of PDE solutions. The mixing in the ergodic sense is defined as following:

Let $(X, \mathcal{A}, \mu)$ be a normalized measure space, and $S : X \to X$ be a measure-preserving transformation. Then S^t is **mixing** if

$$\lim_{t\to\infty}\int_X \psi(S^t(x))\,\phi(x)\,d\mu = \int_X \psi(x)\,d\mu \int_X \phi(x)\,d\mu, \tag{1.11}$$

for any $\psi, \phi \in L^2(X, d\mu)$.

The transformation S^t of our interests solves a particular PDE. There is a huge amount of work in this field, and we only point out some of them here for this brief introduction. The equations considered are for instance the transport equation of a divergence free vector field [5, 75], the transport equation with a diffusion term [29], or the stochastic 2-dimensional Navier-Stokes equation [24]. Beside in fluid dynamics, mixing also plays an important role in kinetic models, especially for plasma physics. One famous example is the phase mixing for the linearized Vlasov equations, which is the cause of Landau damping. This is another vital research area with a lot of existing works. We refer to for example [15], [86], [104] and the references therein.

Our focus for this second part of dissertation is on the passive scalar mixing for the incompressible flows on the n-dimensional torus $\mathbb{T}^n = [-\pi, \pi]^n$. We refer to for instance [5], [75] and [81] for existing works for passive scalar mixing. The invariant measure μ in this case is the normalized Lebesgue measure on $\mathbb{T}^n$, and the measure-preserving transformation S^t is the solution of the transport equation with divergence free vector fields:

$$\begin{cases} \partial_t f + a(x)\cdot\nabla_x f = 0, \\ f(x, t=0) = f_0, \end{cases} \tag{1.12}$$

where $\nabla_x \cdot a(x) = 0$ and $f_0 \in L^2(\mathbb{T}^n)$. Let

$$\theta(x,t) := f(x,t) - \int_{\mathbb{T}^n} f_0(x,t)\, dx. \tag{1.13}$$

Then the mixing (1.11) in our case becomes

$$\lim_{t\to\infty} \int_{\mathbb{T}^n} \phi(x)\theta(x,t)\, dx = 0, \tag{1.14}$$

for all $\phi(x) \in L^2(\mathbb{T}^n)$ and $f_0 \in L^2(\mathbb{T}^n)$.

A lot of interests coming from analysis and applications prompt the search of a way to quantify the degree of mixing. The negative Sobolev norm $H^{-1/2}$ was proposed in [84]. This is later proved in [75] that

$$\|\theta\|_{H^{-s}} \to 0 \ \text{ as } t \to 0 \tag{1.15}$$

is equivalent to (1.14) for all $s > 0$. Beside negative Sobolev norms, another mixing scale was introduced in [23], related to rearrangements of sets. This mixing scale brings the perspective of geometric measure theory to the mixing problems and thus is frequently called the geometric mixing scale. The two mixing scales are not equivalent, but still closely related [108].

Recent progress has been made in the study of the decay rate of these two mix scales under energy constraints on the vector field. It has been shown that both mixing scales can decay at most exponentially [31,64,95]. The mixing phenomenon with an exponential decay rate is often referred to optimal mixing. An explicit example was constructed and analyzed in both mixing scales [4, 5]. For further discussions concerning the optimal mixing, see for instance [75, 81] for H^{-s} norm, and [31, 74] for the geometric mixing scale.

1.3.2 Hamiltonian structure for unmixing of flows

A divergence-free vector field $\mathbf{a} = (a_1, a_2) \in L^2(\mathbb{T}^2)$ on a two-dimensional torus is a local Hamiltonian system. Informally, this means that there exists a function $\psi \in H^1(\mathbb{T}^2)$ such that the following Hamilton's equations hold *locally*:

$$\mathbf{a} = \nabla^{\perp}\psi := (\partial_2\psi, -\partial_1\psi). \tag{1.16}$$

In our discussion, the (Hamiltonian) function ψ would be called a **stream function** for $\mathbf{a}$, which is commonly used in the analysis for incompressible flows.

Due to the Hamiltonian structure (1.16), ψ is constant on each characteristic of the corresponding vector field $\mathbf{a}$. As it is assumed here that ψ is time-independent, this statement can be reworded as follows: the characteristics are retained in the level sets of ψ. The physical interpretation of this statement in the fluid mixing scenario is that quantities can never leave the level set that they originally start with. This phenomenon suggests the fluid interaction near the local maximum and minimum of ψ is lacked and thus the system is unmixing.

In a joint work with P.-E. Jabin, it was shown that this phenomenon is preserved under a sup-norm perturbation on the stream functions. With this perturbation result, a probabilistic setting is considered in order to give a quantitative statement regarding how many vector fields are unmixing. We take a classical approach and consider the vector field $\mathbf{a}$ in the form of a random Fourier series,

$$\mathbf{a} = (a_1, a_2) = (\gamma_0^1, \gamma_0^2) + \sum_{k\in\mathbb{Z}^2\setminus\{0\}} k^{\perp}\frac{i\gamma_k}{(1+|k|)^{\theta}}e^{ik\cdot x}, \tag{1.17}$$

where for all $k \in \mathbb{Z} \setminus \{0\}$, $k^{\perp} \cdot k = 0$, and $\gamma_0^1, \gamma_0^2, \{\gamma_k\}_k$ are random variables.

Deriving from the perturbation in L^∞ norm, one can show that the probability of unmixing is positive under a proper condition on the distributions of the Fourier coefficients. This condition is important to guarantee there is a non-negligible contribution from different frequencies to our system. This effect makes the positive probability of unmixing possible.

1.4 Modeling of transport phenomena

1.4.1 Dynamic modeling for complex systems

A dynamic mathematical model for a complex system normally consists of a set of equations, describing the evolution or properties of each relevant component. Frequently, the system of equations coupled with each other because of the interaction between components. The coupled equations can be seen as an accurate and quantitative description for the dynamic behavior of a physical system.

Here we are in particular interested in the coupled systems with a transport character, such as the models consisting of continuity equations or advection-diffusion equations.

The classical *advection-diffusion equation* is written as follows:

$$\partial_t f + div_x \left(v(t,x)\, f - \sigma(x,v,f) \nabla_x f \right) = 0, \tag{1.18}$$

where f is a density function, x the physical position, and v a velocity field. An effect of diffusion is included in this equation with a strength $\sigma(x,v,f)$.

It is well known that (1.18) is inherently related to the stochastic differential equation:

$$dX = v\,dt + \sigma\,dW, \tag{1.19}$$

where X represents the position, v the velocity, and W is the Brownian motion. (1.19) can be regarded as the microscopic description of the advection-diffusion equation.

The velocity field v is often determined by another equation coupled with the transport equation describing the system. For instance, v solves the Poisson equation for the model of chemotaxis in [25], or the eikonal equation in the Hughes model for pedestrians [49, 62].

The use of advection-diffusion equation can be seen in the models for swarming [101] and homing behavior of animals [85]. It has been demonstrated that random search combined with a directed motion can provide a reasonable explanation for some observed animal homing and migration phenomena [87]. The advection-diffusion equation has been well-developed as one of the important mathematical techniques for the research on animal movement.

In a joint work with W. F. Fagan and P.-E. Jabin, a dynamical model for the movement of *one individual* is constructed using (1.19) coupled with the eikonal equation. The goal of this model is to investigate the memory effects on animal migrations. Before the overview of our work, a short discussion of the models for spatial memory and animal movements shall be presented in the following section.

1.4.2 Models for spatial memory in animal movements

Using statistical inference, several works have shown the important role of spatial memory in animal migrations [1, 22, 45]. However, since the dynamic interaction between memory and movement cannot be observed directly, it is often challenging to utilize statistical inference to investigate the underlying memory mechanism for animal migration. For a broader discussion concerning the spatial memory and movement from the biological prospective, we refer to an excellent review paper [43].

This short presentation is restricted to the models for animal movement with explicit description for the mechanism between the spatial memory and movement. One possible way to describe such mechanism consists of two steps. First, assign a desirability value to geographical region based on the animal's memory and perception. Second, provide an interface that outputs the animal's movement decision based on this desirability landscape. Our novel approach consists in introducing, for the first time, an interface that incorporates memory effect and is compatible with (1.18) or (1.19).

For the first step, it is not uncommon for the models of animal movement to consider multiple covariates when determining the desirability due to the system's high complexity. Frequently, each different source of information is simplified to a certain function. To comprehend all relevant information, one could consider a sum of these functions multiplying with weight functions. This type of combined model can be found in several literature and different forms of weight functions are available; See for instance the step selection functions introduced in [48] and the resources selection functions in [94].

In our interests of memory effect, the spatial memory and perception of the animal are included in the combined model. This combined model can then be considered as the desirability landscape of the environment.

Under the framework of advection-diffusion equation, the interface in the second step can be clarified by specifying how the velocity field v in (1.18) or (1.19) is determined based on the desirability landscape. Classically, the advection term in a biological system is often caused by some attraction in the environment, such as food or shelters. From this point of view, one simple model for the velocity v can be the vector that starts from the present position and points to the best location in the desirability landscape. However, this naive selection does not consider how far the best location is and the resulting movement may not be reasonable.

One way to improve the above simple model is to use the eikonal equation, which is written as follows:

$$|\nabla_x \psi| = \Phi, \tag{1.20}$$

where Φ is a smooth, positive function, often called the *potential* of the eikonal equation. The eikonal equation can be traced back to Fermat's principle in optics, solving for the shortest time path of light in medium.

In our work, the velocity v is modeled by $\nabla_x \psi$, which solves (1.20) with Φ depending on the animal's memory and perception. The solution of this eikonal equation can be regarded as a comprehensive movement decision based on memory, perception, as well as the cost of traveling.

The use of eikonal equation is inspired by the Hughes model for pedestrians [62]. In the Hughes model, the potential of eikonal equation depends on pedestrian density. This is one way to model the phenomenon that the pedestrians normally avoid crowded regions.

1.4.3 Application: Memory effect on migration under periodic environments

A brief overview for a joint work with W. F. Fagan and P.-E. Jabin shall be given in this subsection. The goal of this work is to obtain a better understanding of the effect that memory has on the animals' migration patterns.

There exist works investigating the animal movement affected by memory and changing landscapes [17, 44], but the underlying memory mechanism remain unclear. This motivates us to construct a memory-driven dynamical model, which describes an individual making travel decisions to optimize its fitness, based on its perception and memory toward the environment.

The components of our model includes the position $X(t)$, fitness $P(t)$, and environment $E(t,x)$ of the individual. The value of $E(t,x)$ indicates the condition of environment at time t and location x. The larger the value is, the more resources (or less predators) are available for the individual. Precisely, the dynamics is modeled by the following equation

$$dX = \sigma dW_t + (\overline{P} - P(t))e^{-E(t,X(t))}\nabla_x \psi(t, X(t))dt,$$

where $\sigma > 0$ is a small fixed parameter and W_t is a Brownian motion. $\bar{P} > 0$ is the optimal fitness that an animal can have, and ψ solves the eikonal equation: $|\nabla_x \psi(t,x)| = exp(-H(t,x))$, where H is a function depending on memory and perception. H shall be defined after memory models are specified. Once the model is complete, numerical experiments can be performed to visualize the memory mechanism.

In general, seasonal changes are considered an important factor in animal migration. It is therefore reasonable to include the periodicity in the environment of experiment. Our models are tested under a simple time-periodic environment, which is defined as follows:

Experiment setting: a simple periodic environment. It is assumed that there are two potential habitats, modeled by two disjoint circular regions A and B; See Figure 1.1. The location with positive value of E (*good area*) is alternating between A and B with duration $T = 1$. E is assumed uniformly negative (*poor area*) outside the single good region.

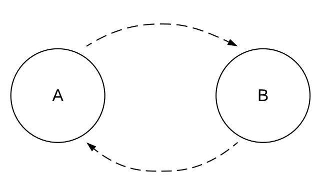

Figure 1.1: Time-periodic Setting: The good region is alternating between A and B with duration $T = 1$.

In our interests of memory effect, it is also assumed that the two habitats are far enough from each other so that the individual cannot see both A and B in the same time. Mathematically, it is assumed that

$$d(A, B) > \sup \{|x - y| : x, y \in supp(K)\}, \tag{1.21}$$

where K is the perception kernel.

As biological memory is complicated and not fully understood, a practical mathematical model for memory often needs to be simplified but still capture some essential features. By observing the simulation results for different memory models, the critical features of a memory system for periodic migration can be discovered. In the spirit of finding an appropriate approximation of memory models for periodic migration, our construction of memory model started simple, and the model complexity was then increased and adjusted until the expected periodic migration pattern was recovered in the simulations.

Two memory models I and II were constructed. The key assumption for our memory models is that the memory fades and is updated over time. The decay and update rates are assumed to be of the same order, which is called the time scale. The memory model I has only one time scale. From the experiment results, a memory system with only one time scale cannot produce a consistent periodic migration. For this reason, the memory model II was introduced with two time scales, which can be considered as the long-term and short-term memory. With proper choices of the two time scales, the desired periodic migration pattern can be recovered. The example of simulations for each case shall be presented after the memory models are specified:

Memory model I: Single memory channel and perception. The memory is modeled by $M(t,x)$, representing how the individual remember the environment condition at time t and location x. To introduce the perception to our model, a perceptual kernel $K(x,y) = k(|x-y|)$ is defined, where k is a positive function on $\mathbb{R}$, decreasing to zero within a finite distance, and with maximum 1. The evolution of $M(t,x)$ follows the equation:

$$\partial_t M(t,x) = -d \cdot sgn(M(t,x))\sqrt{|M(t,x)|} + u \cdot K(X(t),x)(E(t,x) - M(t,x)),$$

with $d > 0$ denoted as the decay rate, and $u > 0$ as the update rate.

With the memory model I, H is defined by

$$H(t,x) = K(X(t),x) \cdot E(t,x) + (1 - K(X(t),x)) \cdot M(t,x). \tag{1.22}$$

(1.22) indicates that the individual evaluates its environment by observation when a location is within its perception range, but when a location is too far to be seen, it evaluates it by its memory.

Memory model II: Long-term and short-term memory. Memory model II contains the long-term memory $M_\ell(t,x)$ and short-term memory $M_s(t,x)$. Both M_ℓ and M_s satisfy the same assumptions in memory model I, but $M_s(t,x)$ has smaller decay and update rates. Here H is defined by:

$$H(t,x) = M_s + M_\ell. \tag{1.23}$$

With (1.23), the individual decides a direction based more on its local environment when it is in an extreme condition. Otherwise, it tends to favor more on the long-term memory. The following experiment shows a successful result.

Experimental results. The dynamics generated with the memory model I is presented by the left picture in Figure 1.2. The resulting trajectory shows that the periodic migration breaks at a certain point. This is because the value of M was updated negative in A and B and the individual would rather exploring the other places that have not been visited before.

On the other hand, the memory model II successfully produced the desired migration patterns; See the right picture in Figure 1.2. Observe that the individual leaves an exhausted region after a bit of explorations because of M_s, and return to A or B based on M_ℓ.

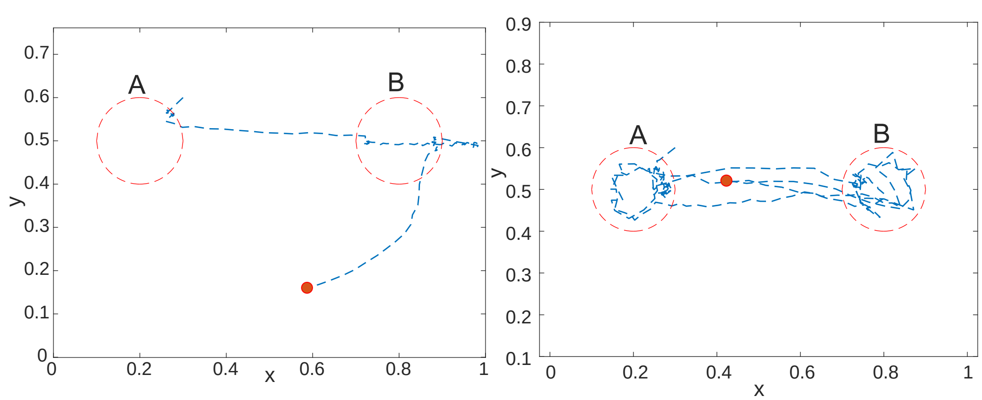

Figure 1.2: The left picture is the numerical experiment result for Memory Model I, and right picture for II. The blue dash line represents the trajectory and the red dot is the location of the animal at the end of the experiment.

Conclusion. Through simulations, it is discovered that in order to have periodic movement, the individual must be able to gather and carry sufficient information from both short-term memory and long-term memory, and possess the ability to discriminate which information is more important with appropriate time scales. A discussion regarding the time scales of long and short-term memory is also provided in Section 4.4.2.

Our model is general and can be utilized to test the memory effect for different circumstances. Some interesting examples are briefly summarized below, (for precise experiment settings see Section 4.5.):

1. A periodic migration pattern of three habitats can also be recovered if the time scales of memory are appropriate regarding the given time-periodic environment.

2. The phenomena of memory disruption can be observed by testing our model under an environment with changing habitats.

3. The phenomenon of a periodic migration route altered by newly discovered habitats can be observed in an environment set up with two major habitats and two nearby intermediate habitats.

4. The critical role of environmental persistence for periodic memory-based migration can be indicated by simulations with random seasonal changes.

Chapter 2: Commutator method for averaging lemmas

2.1 Introduction

2.1.1 Brief overview for averaging lemmas

Our goal of this paper is to introduce the **commutator method** for kinetic transport equations:

$$\varepsilon\partial_t f + a(v)\cdot\nabla_x f = (-\Delta_v)^{\alpha/2} g, \tag{2.1}$$

where $\varepsilon > 0$, $\alpha \geq 0$, $a : \mathbb{R}^n_v \to \mathbb{R}^n$ and $g : \mathbb{R}_t \times \mathbb{R}^n_v \times \mathbb{R}^n_x \to \mathbb{R}$ are given functions. ε is the macroscopic scale normally introduced when a hydrodynamics limit is considered. The nonlinear coefficients $a(v)$ in this setting appears in kinetic formulation of scalar conservation law, and also in kinetic models under relativistic and quantum setting [42], [52].

We shall utilize this method as a new approach to derive **averaging lemmas**, which state that by taking average in microscopic v variable, the velocity average of f

$$\rho_\phi(t,x) := \int f(t,x,v)\phi(v)\,dv, \quad \phi \in L^\infty_c,$$

has better regularity than f and g in x variable, where L^∞_c is the space containing all the bounded and compactly supported functions. There is a vast literature of averaging lemmas, and here we only mention few of them that are relatively closer to our discussion. This type of results is famous for getting compactness for the Vlasov-Maxwell system [38], renormalized solutions [39] and hydrodynamic limits for the Boltzmann equation [53], and the convergence of the renormalized solutions to the semiconductor Boltzmann-Poisson system [83]. It also contributes to the regularizing effect of solutions wherever the kinetic formulations exist, such as the isentropic gas dynamics [78], Ginzburg-Landau model [65], and scalar conservation laws [77].

Classical averaging lemmas were first introduced independently in [2] and [51] under L^2 setting. The derivation in [51] involves decomposition in Fourier space according to the order of $a(v) \cdot \xi$, and controlling the singular part $|a(v) \cdot \xi| < c$ with the non-degeneracy condition. Combining with interpolation arguments, it was later extended to general L^p, $1 < p < \infty$ by [18] and [40]. It was followed by the optimal Besov results proved in [35] by using wavelet decomposition. The regularity for the L^p case is further improved in the one-dimensional case, precisely from $\frac{1}{p}$ to $1 - \frac{1}{p}$ when $p > 2$ by [11], with dispersive property and dyadic decomposition.

Averaging lemmas under different conditions on f and g were further discussed. For instance, in [105] the author considered f and g in the same Besov space in x but can have different integrability in v. The results for general mixed norms assumptions were obtained in [67] [68]. Their work inspired [10] to consider the case when f and g have less integrability in x than v. Except for the explorations in the direction of general conditions, averaging results for a larger class of operators in the form of $a(v) \cdot \nabla_x - \nabla_x^\perp \cdot b(v) \nabla_x$ were acquired by [99]. They presented several applications for their results and especially, they improved the regularity of solutions for scalar conservation laws.

The limiting L^1 case for classical averaging lemmas in general is not true, and a counterexample was given in [51]. However, L^1 compactness can be proved with equi-integrability in only v variable [56], and was extended to more general transport equations in [9] and [59].

2.1.2 Commutator method with multiplier technique

In this work we use commutator method with multipliers to transform the dispersion of transport operator in Fourier space into gain of regularity in x variable. Let us introduce the commutator method in a general setting, and narrow down to our case shortly. Assume

$$\varepsilon \partial_t f + Bf = g,$$

where B is a skew-adjoint operator, $\varepsilon \leq 1$ and g are given. For a time-independent operator Q, we consider

$$\varepsilon \partial_t \int f \, \overline{Qf} \, dx \, dv = \int [B, Q] \, f \bar{f} \, dx \, dv + \int g \, \overline{Qf} \, dx \, dv + \int f \, \overline{Qg} \, dx \, dv$$

And by fundamental theorem of calculus we have

$$\begin{aligned} Re \int_0^T [B, Q] \, f \bar{f} \, dx \, dv \, dt \leq \sup_{t=0,T} \left| \int f \, \overline{Qf} \, dx \, dv \right| + \left| \int g \, \overline{Qf} \, dx \, dv \, dt \right| \\ + \left| \int f \, \overline{Qg} \, dx \, dv \, dt \right|. \end{aligned} \tag{2.2}$$

The idea is to find Q, bounded in some L^p spaces, such that the commutator of B and Q, $[B, Q]$, is positive-definite and gain extra derivatives. Hence by applying these conditions on (2.2) we get a desired bound on f.

This method was used for example by taking B to be of Schrödinger type, where the commutator appear naturally from the Hamilton vector field. Roughly speaking it involves constructing a proper symbol, which corresponds to Q, such that the Poisson bracket implies a spacetime bound on f by Gårding's inequality. See for example [30], [41], [72] and [97].

In this paper we fix B to be the kinetic transport operator,

$$\varepsilon \partial_t f + a(v) \cdot \nabla_x f = g, \tag{2.3}$$

and Q is a bounded multiplier operator. That is, we consider

$$\mathcal{F}_{\xi,\zeta}(Qf) := m(\xi,\zeta)\mathcal{F}_{\xi,\zeta}(f),$$

where m is bounded. So there is a tempered distribution $K(x, v)$ such that $Qf = K \star_{x,v} f$ with $\mathcal{F}_{\xi,\zeta}(K) = m$. In this case the commutator becomes

$$\begin{aligned}
&\int [a(v) \cdot \nabla_x, K \star_{x,v}] f \bar{f}\, dx\, dv \\
&= \int (a(v) - a(w)) \cdot \nabla_x K(x - y, v - w) f(y, w)\, dy\, dw f(x, v)\, dx\, dv.
\end{aligned}$$

When $a(v) = v$, it is simply the quadratic form with the multiplier $\xi \cdot \nabla_\zeta m$. We shall take an advantage of this simple formula and show that the velocity average of f would gain regularity $1/2$ in x when $a(v) = v$, and g is not singular.

The multiplier we select for this purpose is

$$m_0(\xi,\zeta) = \frac{\xi}{|\xi|}\cdot\frac{\zeta}{(1+|\zeta|^2)^{1/2}},$$

and the corresponding kernel

$$K_0 = R\cdot\nabla_v G_1^n,$$

where R is the Riesz potential and G_1^n is the Bessel potential of order 1 in dimension n. With this choice by Plancherel identity,

$$\begin{aligned}\int [v\cdot\nabla_x, K_0\star_{x,v}] f\bar{f}\,dx\,dv\,dt &= \int \xi\cdot\nabla_\zeta m_0 |\hat{f}|^2\,d\xi\,d\zeta\,dt\\ &= \int\int\left[\frac{1}{(1+|\zeta|^2)^{1/2}} - \frac{\left|\frac{\xi}{|\xi|}\cdot\zeta\right|^2}{(1+|\zeta|^2)^{3/2}}\right]|\xi||\hat{f}|^2\,d\zeta\,d\xi\,dt\\ &\geq \int\int\frac{|\xi|}{(1+|\zeta|^2)^{3/2}}|\hat{f}|^2\,d\zeta\,d\xi\,dt = \|f\|^2_{L^2\left([0,T],H^{1/2}(\mathbb{R}^n_x,H^{-3/2}(\mathbb{R}^n_v)\right)}.\end{aligned}$$

From classical Fourier theory (see for example [98]), K_0 is bounded on L^p spaces for all $1<p<\infty$. With this the right hand side of (2.2) is bounded as long as f is in $L^\infty([0,T],L^2(\mathbb{R}^n_x\times\mathbb{R}^n_v))$ and the dual space of g. For convenience, let us denote the conjugate index of p by p', that is, $\frac{1}{p}+\frac{1}{p'}=1$. From the discussion above, we have shown:

Theorem 1. *Let $\varepsilon\leq 1$. If $f\in L^\infty([0,T],(L^2\cap L^p)(\mathbb{R}^n_x\times\mathbb{R}^n_v))$ solves (2.3) with $a(v)=v$ for some $g\in L^1\left([0,T],L^{p'}(\mathbb{R}^n_x\times\mathbb{R}^n_v)\right)$, where $1<p<\infty$, then for all $\phi\in H^{3/2}(\mathbb{R}^n_v)$, $\rho_\phi\in L^2\left([0,T],H^{1/2}(\mathbb{R}^n_x)\right)$, and*

$$\|f\|^2_{L^2_tH^{1/2}_xH^{-3/2}_v}\leq C\left(\|f\|^2_{L^\infty_tL^p_{x,v}}+\|g\|^2_{L^1_tL^{p'}_{x,v}}\right),$$

where C is independent of ε.

Remark 1. *By Wigner transform, this result with $p = 2$ connects to the local smoothing effect for Schrödinger equation.*

Remark 2. *The exchange of regularity between x and v variables is visible through the calculation of commutator, which shares its similarity with the hypoellipticity phenomenon. Very roughly speaking, it is a phenomenon that the degenerate directions can be recovered by commutators, which was developed systematically by Hörmander [61] for Fokker-Planck type of operators. For the hypoellipticity of kinetic transport equations we refer to [21].*

The difference here is that we added a homogeneous zero multiplier m_0 as a buffer, which takes on the impact from the transport operator. So the request for extra regularity in v goes to the test function ϕ, unlike the results in [21], which asked for extra regularity in v for f.

Notice here the requirement of test functions can be adapted to the L^∞_c, same as classical averaging results. This is because the product $f\phi$ with $\phi \in L^\infty_c(\mathbb{R}^n_v)$ still satisfies the kinetic transport equation, and the same procedure would give $f\phi \in L^2\left([0,T], H^{1/2}(\mathbb{R}^n_x, H_v^{-3/2}(\mathbb{R}^n_v))\right)$. Now because of the compact support of the integration, we can take a smooth function identically one inside the integral domain. Our main results will require the test functions to be in L^∞_c, and this argument can be found later in the proof of Theorem 2 in Section 2.4.

Our setting is reminiscent of the multiplier method in [50]. It was used to prove moment and trace lemmas for kinetic equations. For them, the dispersive nature of solutions was acquired by integrating along characteristics in physical space, while here we utilize the technique in frequency domain and so it results in gain of regularity.

For the rest of this paper we are going to extend this method to the general transport equation (2.1) with the variable coefficient $a(v)$ and a singular source term $(-\Delta_v)^{\alpha/2}g$, which introduce difficult technical issues. The commutator method pairing with m_0 will be the main mechanism for our proofs. The advantage of this approach is that the integrability of f and g can be of assistance to each other. This is the feature that distinguishes our results from others in the literature, and provides averaging results for a new type of mixed integrability assumptions, which fits nicely for the conditions that the kinetic formulation of scalar conservation law naturally attain.

This paper is organized as follows. We shall present our main theorems in Section 2.2, and an example of application to scalar conservation laws in Section 2.3. Finally proofs of theorems are in Section 2.4.

2.2 Main results

2.2.1 Our main velocity averaging result

We present averaging lemmas for (2.1) derived by the commutator method. To have dispersion in Fourier space for the kinetic transport operator $a(v) \cdot \nabla_x$, one need conditions on the variable coefficients $a(v)$. Indeed, there is no gain of regularity if a is only constant for example.

In this section, we assume $a(v) \in Lip(\mathbb{R}^n)$ with conditions:

$$a(v) \text{ one-to-one, and } J_{a^{-1}} \in L^{\gamma}, \tag{2.4}$$

where $J_{a^{-1}} = det(Da^{-1})$. The assumptions quantify the nonlinearity of $a(v)$ with index γ, and allow us to control the integrability of functions after the change of variables $v \mapsto w = a(v)$.

Our proof involves regularization of equation (2.1) through various embeddings. The interaction between embedding and the singular term $(-\Delta_v)^{\alpha/2} g$ will affect the resulting gain of regularity, and this introduce several exponents and indices in the formulas which we collect below,

$$d_1 = \max\left\{n\left(\frac{1}{p_2}+\frac{1}{q_2}-\ell\right),0\right\}, \quad d_2 = \max\left\{n\left(\frac{2}{p_2}-\ell\right),0\right\}, \ \ell = \frac{\gamma-2}{\gamma-1}, \tag{2.5}$$

$$d_3 = \max\left\{n\left(\frac{1}{p_1}+\frac{1}{q_1}-1\right),0\right\}, \quad d_4 = \max\left\{n\left(\frac{2}{p_1}-1\right),0\right\}. \tag{2.6}$$

Our result is as follows,

Theorem 2. *Given $\alpha \geq 0$, $T > 0$ and $0 < \varepsilon \leq 1$. Let $a \in Lip(\mathbb{R}^n)$ satisfy (2.4) with $\gamma \geq 2$. Let $f \in L^\infty([0,T], L^{p_1}(\mathbb{R}^n_x, L^{p_2}(\mathbb{R}^n_v)))$ solve (2.1) for some $g \in L^1([0,T], L^{q_1}(\mathbb{R}^n_x, L^{q_2}(\mathbb{R}^n_v)))$, with p_1, p_2, q_1, $q_2 \in [1,\infty]$. Then for any ball $B_R(x_0) \subset \mathbb{R}^n_x$ and $\phi \in C^\infty_c(\mathbb{R}^n_v)$, one has that $\rho_\phi(t,x) \in L^2([0,T], H^s(B_R(x_0)))$ for all $s < S$, with*

$$\|\rho_\phi\|^2_{L^2([0,T],H^s_x(B_R(x_0)))} \leq C\left(\|f\|^2_{L^\infty([0,T],L^{p_1}(\mathbb{R}^n_x,L^{p_2}(\mathbb{R}^n_v)))} + \|g\|^2_{L^1([0,T],L^{q_1}(\mathbb{R}^n_x,L^{q_2}(\mathbb{R}^n_v)))}\right),$$

where $S = \frac{1}{2}\left\{(1-d_2)\theta - d_4\right\}$ *with* $\theta = \left[\min\left\{\frac{1-(d_3-d_4)}{\alpha+1+(d_1-d_2)}, 1\right\}\right]$, *where* d_i *are defined in (2.5) and (2.6) for* $i = 1,2,3,4$ *and* C *only depends on* R, p_1, q_1 *and* $Lip(a)$.

Remark 3. *The restriction* $\gamma \geq 2$ *can be relaxed, but with a different formula for* $S = \frac{1}{2}\left\{\left[1 - n\left(\frac{2}{p_2} + \frac{2}{\gamma} - 1\right)\right]\tilde{\theta} - d_4\right\}$ *when* $1 \leq \gamma < 2$, *where* $\tilde{\theta} = \min\left\{\frac{1-(d_3-d_4)}{\alpha+1+n\left(\frac{1}{q_2}-\frac{1}{p_2}\right)}, 1\right\}$.

Remark 4. *If* $f \in L^\infty([0,T], B^0_{p_1,2}(\mathbb{R}^n_x, L^{p_2}(\mathbb{R}^n_v)))$ *and* $g \in L^1([0,T], B^0_{q_1,2}(\mathbb{R}^n_x, L^{q_2}(\mathbb{R}^n_v)))$, *the end point* $s = S$ *can be included when* $p_1, p_2, q_1, q_2 \in (1, \infty)$.

Remark 5. *Because of the quadratic form in our method, our result always bounds the velocity average in* L^2, *and the bound has the same weight on the norms of* f *and* g, *independent of* p_1, p_2, q_1, q_2.

When $a(v) = v$, one has that $\gamma = \infty$. In this case, we have a simpler formula for Theorem 2 when f and g are in the dual space of each other:

Corollary 1. *Given* $\alpha \geq 0$, $T > 0$ *and* $0 < \varepsilon \leq 1$. *If* f *belongs to the space* $L^\infty([0,T], L^{p_1}(\mathbb{R}^n_x, L^{p_2}(\mathbb{R}^n_v)))$ *and solves (2.1) with* $a(v) = v$ *for some* $g \in L^1\left([0,T], L^{p_1'}(\mathbb{R}^n_x, L^{p_2'}(\mathbb{R}^n_v))\right)$, *where* $p_1, p_2 \in [2, \infty]$. *Then for any* $\phi \in C^\infty_c(\mathbb{R}^n_v)$, $\rho_\phi \in L^2([0,T], H^s(\mathbb{R}^n_x))$ *for all* $s < \frac{1}{2(\alpha+1)}$.

Here we conclude with some relations between our result and previous literature.

- First of all, let us point out that our velocity averaging result is independent of small ε. This could have applications to the compactness of solutions for rescaled kinetic equations, which frequently appear in the discussions of hydrodynamic limits. For more in this direction we refer to for example [57] and [93].

Moreover, since our argument doesn't perform a Fourier transform in time variable, this method has possible extensions for time discretized kinetic equations or stochastic cases.

As there is already a huge literature on averaging lemmas, and under some situations the results were proven optimal, we would like to give the readers an idea on when our method becomes effective, and what are the potential advantages our result could provide.

For the rest of this subsection, we will compare the regularity in x of our result, with the theorems in [11], [40] and [105]. Because our resulting space has a different integrability from previous results except for the L^2 case, our method may render a more appropriate tool under certain circumstances. We will also point out the regions where one theorem can imply the other, through embedding or interpolation. The interpolation is applied between the resulting space of ρ_ϕ and the assumption space of f, because ρ_ϕ has the same integrability in x as f.

Notice some theorems we quote here apply to more general conditions in the original statements, but for simplicity we shall only state the parts that concern our discussion, and restrict to the special case $a(v) = v$. We also assume for convenience that f and g are compactly supported in x and v, and $\phi \in C_c^\infty$ for this entire discussion.

Let us begin with the classical averaging result in [40], where the different integrabilities for f and g and $\alpha > 0$ are available.

Theorem 3. *[40] If $f \in L^p(\mathbb{R}_t \times \mathbb{R}^n_x \times \mathbb{R}^n_v)$ and $g \in L^q(\mathbb{R}_t \times \mathbb{R}^n_x \times \mathbb{R}^n_v)$, satisfying (2.1) with $a(v) = v$, then $\rho_\phi \in B^s_{r,\infty}(\mathbb{R}_t \times \mathbb{R}^n_x)$ where $s = \frac{1}{\bar{p}}\left(\alpha + \frac{1}{\bar{p}} + \frac{1}{\underline{q}}\right)^{-1}$, $\bar{p} = \max\{p, p'\}$, $\underline{q} = \min\{q, q'\}$, and $\frac{1}{r} = \frac{s}{q} + \frac{1-s}{p}$. Moreover, if $p = q \in (1, \infty)$, $\rho_\phi \in B^s_{r,t}(\mathbb{R}_t \times \mathbb{R}^n_x)$ where $t = \max\{p, 2\}$.*

Under the assumption of Theorem 3, we start our discussions for the cases when $p = q$.

- *When $p = q = 2$, both Theorem 2 and 3 reach the same regularity $H^{\frac{1}{2(1+\alpha)}}$.*

- *When $p = q \in (1, 2)$, the result by Theorem 3 implies Theorem 2:*

  Indeed, Theorem 3 reaches $B^{\frac{1}{p'(1+\alpha)}}_{p,2}$, while Theorem 2 gives H^s for all $s < S = \frac{1}{2(1+\alpha)}\left[1 - n(2+\alpha)\left(\frac{2}{p} - 1\right)\right]$. By embedding theorem for Besov spaces, $B^{\frac{1}{p'(1+\alpha)}}_{p,2} \subset H^{\tilde{s}}$ with $\tilde{s} = \frac{1}{p'(1+\alpha)} + n\left(\frac{1}{2} - \frac{1}{p}\right)$, which is larger or equal to S for all $n \geq 1$ and $p < 2$.

- *When $p = q \in (2, \infty)$, the result by Theorem 2 has more differentiability but less integrability than Theorem 3. And when $n = 1$ and $\alpha = 0$, Theorem 2 implies Theorem 3:*

  Theorem 3 reaches $B^{\frac{1}{p(1+\alpha)}}_{p,p}$, while Theorem 2 have $H^{\frac{1}{2(1+\alpha)}}$. Our result has more differentiability but less integrability as $p > 2$. By embedding $H^{\frac{1}{2(1+\alpha)}} \subset$

$B^{\tilde{s}}_{p,2}$, where $\tilde{s} = \frac{1}{2(1+\alpha)} + n\left(\frac{1}{p} - \frac{1}{2}\right)$. And $\tilde{s} < \frac{1}{p(1+\alpha)}$ except when $n = 1$ and $\alpha = 0$, where the equality holds.

Because of the quadratic form in our method, one sees the more favorable type of conditions for our method is when $p \geq 2$ and $\frac{1}{p} + \frac{1}{q} = 1$. We therefore compare Theorem 2 and 3 under this assumption:

- *Under the assumption of Theorem 3 with* $\frac{1}{p}+\frac{1}{q} = 1$ *and* $p \in (2, \infty)$*, the result by Theorem 2 has more differentiability but less integrability in* x*. Moreover, Theorem 2 implies Theorem 3 when* $\alpha = 0$ *by interpolation, or when* $0 \leq \alpha < \frac{1}{n}$ *and* $2 < p < \frac{2n}{n(1+\alpha)-1}$ *by embedding:*

  Under these conditions, Theorem 2 results in $H^{\frac{1}{2(1+\alpha)}}(\mathbb{R}^n_x)$, while Theorem 3 reaches $B^{\frac{1}{p(1+\alpha)}}_{r,\infty}(\mathbb{R}^n_x)$, where $\frac{1}{r} = \frac{1}{p(1+\alpha)}\left(1 - \frac{2}{p}\right) + \frac{1}{p}$. By the interpolation between $H^{\frac{1}{2(1+\alpha)}}$ and L^p, we have $W^{\frac{1}{p(1+\alpha)^2},r} \subset B^{\frac{1}{p(1+\alpha)^2}}_{r,r}$. This shows when $\alpha = 0$, Theorem 2 implies Theorem 3.

  In the other hand, by embedding $H^{\frac{1}{2(1+\alpha)}} \subset B^{\tilde{s}}_{r,2}$, where $\tilde{s} = \frac{1}{2(1+\alpha)} + n\left(\frac{1}{r} - \frac{1}{2}\right)$. Even with the dimension dependence, there are regions that embedding gives a better regularity than interpolation. For example when $n = 1$, $\tilde{s} \geq \frac{1}{p(1+\alpha)^2}$ when $p \leq 2 + \frac{2}{\alpha}$. We compare $\tilde{s}$ with the regularity obtained by Theorem 3. In general for each fixed n, $\tilde{s} \geq \frac{1}{p(1+\alpha)}$ when $p \leq \frac{2n}{n(1+\alpha)-1}$, which is compatible with $p > 2$ only when $\alpha < \frac{1}{n}$. Hence Theorem 2 implies Theorem 3 when $0 \leq \alpha < \frac{1}{n}$ and $2 < p < \frac{2n}{n(1+\alpha)-1}$.

We now compare our result with [11] and [105], where mixed norm conditions in general dimensions were considered for the stationary transport equation

$$v \cdot \nabla_x f = g. \tag{2.7}$$

We shall take $\varepsilon = 0$, in order to compare our theorem with results for (2.7).

Theorem 4. *[105] For $1 < p < \frac{n}{n-1}$, if $f \in B^0_{p,q}(\mathbb{R}^n_x, L^{p_2}(\mathbb{R}^n_v))$ and $g \in B^0_{p,q}(\mathbb{R}^n_x, L^{q_2}(\mathbb{R}^n_v))$ satisfy (2.7), then $\rho_\phi \in B^S_{P,q}(\mathbb{R}^n_x)$, where $S = -n + 1 + \frac{1}{p_2'}\left[1 + \frac{1}{q_2} - \frac{1}{p_2}\right]^{-1}$ and $P = \left[\frac{1}{p} - \frac{n-1}{n}\right]^{-1}$.*

Theorem 5. *[11] When $\frac{4}{3} \le p \le 2$, if $f, g \in L^p(\mathbb{R}^n_x, L^2(\mathbb{R}^n_v))$ satisfy (2.7), then $\rho_\phi \in W^{s,p}(\mathbb{R}^n)$ for all $s < S$, where $S = \frac{1}{2}$ when $n = 1, 2$, and $S = \frac{1}{2}\left(3 - \frac{4}{p}\right) + \frac{n}{4(n-1)}\left(\frac{4}{p} - 2\right)$ when $n \ge 3$.*

For the comparison with Theorem 4, we take $q = 2$ for an easier discussion with our H^s result. And since Theorem 4 allows general integrabilities in v, let us consider $p_2 = q_2' \ge 2$, which is the most favorable condition for our method.

- *Under the assumption of Theorem 4 with $n = 1$, $q = 2$ and $p_2 = q_2' \ge 2$. Both Theorem 2 and 4 reach the same regularity when $p = 2$. And Theorem 4 implies Theorem 2 when $p \neq 2$:*

  Here Theorem 4 reaches $B^{1/2}_{p,2}$, while Theorem 2 has $H^{1/p'}$ when $p \le 2$ and $H^{1/2}$ when $p > 2$, as mentioned in Remark 4. When $p = 2$, the two results are exactly the same. When $p < 2$, the spaces $B^{1/2}_{p,2}$ and $H^{1/p'}$ have the same scaling, and $B^{1/2}_{p,2} \subset H^{1/p'}$ by embedding. At last for $p > 2$, $H^{1/2} \subset B^{1/2}_{p,2}$.

Notice for $n \geq 2$, Theorem 4 no longer applies to $p > 2$, same as Theorem 5. The restriction $p < 2$ is not the best situation for our method, but the comparison is still interesting under these mixed norm conditions.

- *Under the assumption of Theorem 4 with $n \geq 2$ (which forces $1 < p < 2$), $q = 2$ and $p_2 = q_2' \geq 2$, our result implies Theorem 4:*

  In this case Theorem 4 gets $B^{3/2-n}_{P,2}$ with $P = \left[\frac{1}{p} - \frac{n-1}{n}\right]^{-1}$, and our method reaches $H^{\frac{1}{2}\left[1-\frac{2n}{p}+n\right]}$ as mentioned in Remark 4. Our result has more differentiability but less integrability. Moreover, by the embedding $H^{\frac{1}{2}\left[1-\frac{2n}{p}+n\right]} \subset B^{\tilde{s}}_{P,2}$, where $\tilde{s} = \frac{1}{2}\left[1 - \frac{2n}{p} + n\right] + n\left(\frac{1}{P} - \frac{1}{2}\right) = \frac{3}{2} - n$.

- *Under the assumption of Theorem 5, the result by Theorem 2 has more integrability but less differentiability than Theorem 5. Furthermore, Theorem 5 implies Theorem 2 when $n = 1$ and 2, but the implication does not hold for $n \geq 3$:*

  Under this assumption, we again have H^s with $s < \frac{1}{2}\left[1 - \frac{2n}{p} + n\right]$. For both $n = 1$ and 2, $W^{1/2,p}_x \subset H_x^{\frac{1}{2}\left[1-\frac{2n}{p}+n\right]}$ by Sobolev embedding. As for $n \geq 3$, $W^{s,p}_x \subset H^{\tilde{s}}_x$ where $s = \frac{1}{2}\left(3 - \frac{4}{p}\right) + \frac{n}{4(n-1)}\left(\frac{4}{p} - 2\right)$ and $\tilde{s} = \frac{1}{2}\left(3 - \frac{4}{p}\right) + \frac{n}{4(n-1)}\left(\frac{4}{p} - 2\right) + n\left(\frac{1}{2} - \frac{1}{p}\right)$. Notice $\tilde{s} < \frac{1}{2}\left[1 - \frac{2n}{p} + n\right]$ for all $p < 2$ and $n \geq 3$, so Theorem 5 cannot imply Theorem 2 in this case.

2.2.2 On the non-degeneracy conditions

The assumption (2.4) we imposed for Theorem 2 is different from the classical conditions on $a(v)$ in the previous literature, called the non-degeneracy condition:

Definition 1. *$a \in Lip(\mathbb{R}^n, \mathbb{R}^m)$ satisfies the* ***non-degeneracy condition of order*** *$\nu \in (0, 1]$, if there exists $c_0 > 0$ such that for all compact set $D \subset \mathbb{R}^n$,*

$$\mathcal{L}^n(\{v \in D : |a(v) \cdot \sigma - \tau| \leq \alpha/2\}) \leq c_0 \alpha^\nu, \tag{2.8}$$

for all $\sigma \in \mathbb{S}^{m-1}$ and $\tau \in \mathbb{R}$, where $\mathcal{L}^n$ is the Lebesgue measure in $\mathbb{R}^n$.

Our assumption (2.4) is stronger than (2.8) with $\nu = 1 - \frac{1}{\gamma}$. Indeed, when $n = m$, the assumption $J_{a^{-1}} \in L^\gamma_v$ implies (2.8) with $\nu = 1 - \frac{1}{\gamma}$, but the other direction holds only when $n = \nu = 1$. When $n > 1$, (2.8) only gives restrictions on the pre-images of bands. And when $\nu < 1$, one can construct a Lipschitz function a_ν on $\mathbb{R}$ satisfying (2.8), and a sequence of measurable sets $\mathcal{O}^i$ such that $\frac{|a_\nu^{-1}(\mathcal{O}^i)|}{|\mathcal{O}^i|^\nu} \to \infty$ as $i \to \infty$, which shows $J_{a^{-1}} \notin L^\gamma$. An example of construction can be found in Appendix A.

The dimension of interests is $n \leq m$ for applications, especially when $n = 1$ for scalar conservation laws. In an attempt to weaken the assumption to non-degeneracy condition with general $n \leq m$ cases, we do a different change of variables $v \mapsto \lambda = a(v) \cdot \frac{\xi}{|\xi|}$, where ξ is the frequency variable of x, and our method can recover the traditional result in L^2 for $\nu = 1$.

Theorem 6. *Given $n \le m$, $\alpha \ge 0$, $T > 0$ and $0 < \varepsilon \le 1$. Assume $a \in Lip(\mathbb{R}^n, \mathbb{R}^m)$ satisfies the non-degeneracy condition (2.8) with $\nu = 1$. Let $f \in L^\infty([0,T], L^2(\mathbb{R}^m_x \times \mathbb{R}^n_v))$ solve (2.1) for some $g \in L^1([0,T], L^2(\mathbb{R}^m_x \times \mathbb{R}^n_v))$, then for any $\phi \in C_c^\infty$, one has $\rho_\phi(t,x) \in L^2\left([0,T], H^{\frac{1}{2(\alpha+1)}}(B_R(x_0))\right)$, and*

$$\|\rho_\phi\|^2_{L^2\left([0,T], H^{\frac{1}{2(\alpha+1)}}(\mathbb{R}^m_x)\right)} \le C\left(\|f\|^2_{L^\infty([0,T],L^2(\mathbb{R}^m_x \times \mathbb{R}^n_v))} + \|g\|^2_{L^1([0,T],L^2(\mathbb{R}^m_x \times \mathbb{R}^n_v))}\right),$$

where C only depends on c_0 and $Lip(a)$.

This L^2 theorem recovers the same regularity $H^{\frac{1}{2(\alpha+1)}}$ in x as in [38] and [40]. Even though this regularity result is not new, we provide a different approach for proving this theorem. As we mentioned in the discussion after Corollary 1, some interesting features which are also inherited by Theorem 3 include:

- Potential applications to hydrodynamic limits as our results are independent of ε.
- The absence of Fourier transform in time variable which enables potential extensions of our method for time-discretized or stochastic kinetic equations.

Remark 6. *We were unable to obtain a L^p statement as we did in Theorem 2. This is because the natural multiplier for the alternate proof here is not a Calderon-Zygmund operator, and we lose bounds in general L^p spaces. In fact, when $a(v) = v$, the corresponding multiplier would be in the form of $S\left(\frac{\xi}{|\xi|} \cdot \zeta\right)$, where ζ is the frequency variable of v. If S is smooth, the inverse Fourier transform of this type of "directed multiplier" in two-dimension is in the form of $\frac{x \cdot v}{|x|^3} \tilde{S}\left(\frac{x^\perp \cdot v}{|x|}\right)$, which is not bounded on $L^p_{x,v}$.*

Remark 7. *We use the non-degeneracy condition as a constraint on the measures of pre-images of intervals. We extend this condition from intervals to general measurable sets, so that this is equivalent to a constraint on the determinant of Jacobian matrices and a proof similar to the one of Theorem 2 follows. But when* $\nu < 1$, *the extension from intervals to measurable sets fails (see the counterexample in the Appendix) and hence the strategy is not applicable directly here.*

2.3 An example of future perspective: Regularizing effects for measure-valued solutions to scalar conservation law

Among several potential applications of the new method for averaging lemmas presented here, this section focuses on the regularity of so-called measure-valued solutions of conservation laws and in particular scalar conservation laws.

Scalar conservation laws can be viewed as a simplified model of hyperbolic systems which still captures some of the basic singular structure. They read

$$\begin{cases} \partial_t u + \sum_{i=1}^n \partial_{x_i} A_i(u) = 0, \\ u(t=0,x) = u_0(x), \end{cases} \tag{2.9}$$

where $u(t,x) : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}$ is the scalar unknown and $A : \mathbb{R} \to \mathbb{R}^n$ is a given flux.

The concept of measure-valued solutions to hyperbolic systems such as (2.9) had already been introduced in [36]. It has recently seen a significant revival of interest as measure-valued solutions offer a more statistical description of the dynamics, see in particular [46, 47].

It is convenient to define measure-valued solution through the kinetic formulation of (2.9), which also allows for a straightforward application of our results. A scalar function $u(t,x) \in L^\infty(\mathbb{R}_+,\ L^1(\mathbb{R}^n)$ corresponds to a measure-valued solution if there exists $f(t,x,v) \in L^\infty(\mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R})$ with the constraint

$$u(t,x) = \int_{\mathbb{R}} f(t,x,v)\,dv, \quad -1 \leq f \leq 1, \tag{2.10}$$

and if f solves the kinetic equation

$$\partial_t f + a(v) \cdot \nabla_x f = \partial_v m, \tag{2.11}$$

for $a(v) = A'(v)$ and any finite Radon measure m. If u is obtained as a weak-limit of a sequence u_n then f includes some information on the oscillations of u_n since it can directly be obtained from the Young measure μ of the sequence

$$f(t,x,v) = \int_0^v \mu(t,x,dz).$$

The system (2.10)-(2.11) is hence immediately connected to the notion of kinetic formulation for scalar conservation laws introduced in the seminal article [77] and extended to isentropic gas dynamics in [78]. If u is an entropy solution to (2.9), then one may define

$$f(t,x,v) = \begin{cases} 1 \quad \text{if } 0 \leq v \leq u(t,x), \\ -1 \quad \text{if } u(t,x) \leq v < 0, \\ 0 \quad \text{otherwise}, \end{cases} \tag{2.12}$$

and f solves the kinetic equation (2.11) with the additional constraint that $m \geq 0$ which corresponds to the entropy inequality.

We refer for example to [92] for a thorough discussion of kinetic formulations and their usefulness, such as recovering the uniqueness of the entropy solution first obtained in [71].

The use of kinetic formulations has proved effective in particular in obtaining regularizing effects for scalar conservation laws. In one dimension and for strictly convex flux, Oleinik [88] proved early that entropy solutions are regularized in BV. In more than one dimension and for more complex flux that are still non-linear in the sense of (2.8) with $\nu = 1$, a first regularizing effect had been obtained in [77] yielding $u \in W^{s,p}$ for all $s < 1/3$ and some $p > 1$.

Such regularizing effects actually do not use the sign of m and for this reason hold for any weak solution to (2.9) with bounded entropy production. Among that wider class a counterexample constructed in [34] proves that solutions cannot in general be expected to have more than 1/3 derivative. The optimal space $(B^{\infty}_{1/3,3})_{x,loc}$ was eventually derived in [55]. Whether a higher regularity actually holds for entropy solutions (instead of only bounded entropy production) remains a major open problem though.

It had been observed in [66] that the regularizing effect for the kinetic formulation relies in part in the regularity of the function f defined by (2.12): For example such an f belongs to $L^\infty(\mathbb{R}_+ \times \mathbb{R}^n,\ BV(\mathbb{R}))$. Unfortunately such additional regularity is lost for measure-valued solutions since we only have $f \in L^1 \cap L^\infty$

by (2.10).

A priori, one may hence only apply the standard averaging result from [40] directly on (2.11). Assuming non-degeneracy of the flux, *i.e.* (2.8) with $\nu = 1$, we may apply Theorem 3 for any $\alpha > 1$, $g \in L^1$ and $f \in L^2$ (the optimal space for this theorem). One then deduces that if u corresponds to a measure-valued solution with f compactly supported in v then $u \in B^s_{5/3,2}$ for any $s < 1/5$.

However we are then making no use of the additional integrability of f. Instead one may also apply our new result Theorem 2 to (2.11) with

Corollary 2. *Let f satisfy (2.10) and solve (2.11) for some finite Radon measure m and some $a : \mathbb{R}^n \to \mathbb{R}^n$ with (2.4) for $\gamma = \infty$. Assume moreover that $f \in L^\infty([0,\ T],\ L^1(\mathbb{R}^n \times \mathbb{R}^n))$ and is compactly supported in velocity. Then $u \in L^2([0,\ T],\ H^s(\mathbb{R}^n))$ for any $s < 1/4$.*

In dimension 1, Corollary 2 directly applies to measure-valued solutions and improve the regularity from almost $B^{1/5}_{5/3,2}$ in x to almost $H^{1/4}$. In higher dimensions, as we observed, we cannot directly replace (2.4) with (2.8). Therefore a better understanding of the regularity of measure-valued solutions is directly connected to further investigations of what should replace (2.4) if $a : \mathbb{R}^m \to \mathbb{R}^n$ with $m < n$.

2.4 Proofs

2.4.1 Proof of Theorem 2

2.4.1.1 Main proof

The proof contains mainly three steps as follows.

Step 1: *Preparations: localization, regularization and change of variables.*

As the result is local, we assume f is compactly supported in x for convenience. Fix a compactly supported function $\phi(v) \in W_v^{\alpha,\infty}$. Without loss of generality, assume $supp(\phi) \subseteq B(0,1)$. Consider $f\phi$, which satisfies

$$\varepsilon\partial_t(f\phi) + a(v)\cdot\nabla_x(f\phi) = (-\Delta_v)^{\alpha/2}g\phi.$$

We denote the Fourier transform of f in x by $\tilde{f}$. Fix a smooth function $\Phi(v)$ with $supp(\Phi) \subseteq B(0,1)$. Consider $F_{s_1} = (\tilde{f}\phi)\star_v\Phi_{|\xi|^{-s_1}}$, where $\Phi_{|\xi|^{-s_1}}(v) = |\xi|^{ns_1}\Phi(v|\xi|^{s_1})$ with $s_1 \geq 0$ to be decided later. Notice

$$supp(F_{s_1}) \subseteq \overline{supp(\phi) + supp(\Phi_{|\xi|^{-s_1}})} \subseteq \overline{B(0, 1+|\xi|^{-s_1})} \subset \overline{B(0,2)}$$

is of compact support for all $|\xi| \geq 1$. And it satisfies

$$\varepsilon\partial_t F_{s_1} + ia(v)\cdot\xi F_{s_1} = ((-\Delta_v)^{\alpha/2}\tilde{g}\phi)\star_v\Phi_{|\xi|^{-s_1}} + Com^1, \tag{2.13}$$

where $F_{s_1} = (\tilde{f}\phi)\star_v\Phi_{|\xi|^{-s_1}}$ and the commutator term

$$Com^1(v) = i\int(a(v)-a(w))\cdot\xi\tilde{f}(w)\phi(w)\Phi_{|\xi|^{-s_1}}(v-w)\,dw.$$

Note the usage of localization in v will be more clear in the last step of our proof.

By change of variables $v \mapsto v' = a(v)$, (2.13) can be rewritten as

$$\varepsilon \partial_t h + i v' \cdot \xi h = k^1 + k^2 \tag{2.14}$$

in the sense of distribution, where h, k^1 and k^2 are defined as follows:

$$\int h(v')\psi(v')\,dv' = \int F_{s_1}(v)\psi(a(v))\,dv.$$

$$\int k^1(v')\psi(v')\,dv' = \int \left[((-\Delta_v)^{\alpha/2}\tilde{g}\phi) \star_v \Phi_{|\xi|^{-s_1}}\right](v)\psi(a(v))\,dv,$$

and

$$\int k^2(v')\psi(v')\,dv' = \int Com^1(v)\psi(a(v))\,dv.$$

Step 2: *Commutator method with m_0 on h.* Consider a smooth radial bump function $\chi(\xi)$ with support on $\frac{1}{2} < |\xi| < 2$, such that $\sum_{k\in\mathbb{Z}} \chi(2^{-k}\xi) \equiv 1$, for all $\xi \neq 0$. For each $k \in \mathbb{N}$, we apply commutator method with m_0 on $h(v')\chi\left(2^{-k}\xi\right)$ and get

$$\begin{aligned}
&\int |\mathcal{F}_{\zeta'}(h)|^2(\zeta')\frac{2^k}{(1+|\zeta'|^2)^{3/2}}\,d\zeta'\,dt\,\chi(2^{-k}\xi)d\xi \\
&\lesssim \int \xi\cdot\nabla_{\zeta'} m_0(\xi,\zeta')|\mathcal{F}_{\zeta'}(h)|^2\,\chi(2^{-k}\xi)d\xi\,d\zeta'\,dt \\
&= \int \bar{h}(v')(\frac{1}{i}\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n \star_{v'} h)\,dv'\chi(2^{-k}\xi)\,d\xi|_{t=0}^{t=T} \\
&\qquad + Re\int \bar{h}(v')\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n \star_{v'} \left[(k^1+k^2)\right]\,dv'\chi(2^{-k}\xi)\,d\xi\,dt \\
&:= A_k
\end{aligned} \tag{2.15}$$

We estimate A_k and get

Lemma 1. *Denote* $\mathcal{F}_x^{-1}(\chi(2^{-k}\xi)\tilde{f}\phi)$ *by* f_k *and* $\mathcal{F}_x^{-1}(\chi(2^{-k}\xi)\tilde{g}\phi)$ *by* g_k*. Let* $p_1, p_2, q_1, q_2 \in (1, \infty]$*. Then for each fixed* $k \in \mathbb{N}$ *,*

$$\begin{aligned} |A_k| &\lesssim 2^{kd_4+ks_1d_2}\|f_k\|^2_{L_x^{p_1}L_v^{p_2}}|_{t=0}^{t=T} \\ &\quad + 2^{kd_3+ks_1d_1+k\alpha s_1}\int \|f_k\|_{L_x^{p_1}L_v^{p_2}}\|g_k\|_{L_x^{q_1}L_v^{q_2}}\,dt \\ &\quad + 2^{kd_4+ks_1d_2+k(1-s_1)}\int \|f_k\|^2_{L_x^{p_1}L_v^{p_2}}\,dt. \end{aligned} \tag{2.16}$$

where $d_1 = \max\left\{n\left(\frac{1}{p_2}+\frac{1}{q_2}-\ell\right), 0\right\}$, $d_2 = \max\left\{n\left(\frac{2}{p_2}-\ell\right), 0\right\}$, *with* $\ell = \frac{\gamma-2}{\gamma-1}$, *and* $d_3 = \max\left\{n\left(\frac{1}{p_1}+\frac{1}{q_1}-1\right), 0\right\}$, $d_4 = \max\left\{n\left(\frac{2}{p_1}-1\right), 0\right\}$.

To minimize the order of ξ of the sum in (2.16), we choose

$$s_1 = \min\left\{\frac{1-(d_3-d_4)}{\alpha+1+(d_1-d_2)}, 1\right\},$$

and so the highest order is $1-S$, where $S = s_1(1-d_2)-d_4$.

Divide the whole inequality (2.15) with $2^{k(1-S+\delta)}$ for any small $\delta > 0$, then we attain

$$\begin{aligned} &\int |\mathcal{F}_{\zeta'}(h)|^2\frac{|\xi|^{(S-\delta)}}{(1+|\zeta'|^2)^{3/2}}\,d\zeta'\,dt\,\chi(2^{-k}\xi)d\xi \\ &\lesssim 2^{-k\delta}\left[\|f_k\|^2_{L_x^{p_1}L_v^{p_2}}|_{t=0}^{t=T} + \int \|f_k\|_{L_x^{p_1}L_v^{p_2}}\|g_k\|_{L_x^{q_1}L_v^{q_2}}\,dt + \int \|f_k\|^2_{L_x^{p_1}L_v^{p_2}}\,dt\right], \end{aligned} \tag{2.17}$$

for all $k \in \mathbb{N}$.

The same inequality can be obtained even if any of p_1, p_2, q_1, q_2 is equal to 1, because the additional logarithm appears from the weak boundedness of Calderon-Zygmund operator would not affect the argument.

Sum over $k \in \mathbb{N}$ for (2.17), we get

$$\begin{aligned} &\int \chi_0(\xi)|\xi|^s\bar{h}(v)G_3^n(v-w)h(w)\,dw\,dv\,dt\,d\xi \\ &\qquad \lesssim \|f\|^2_{L^\infty\left([0,T],L_x^{p_1}L_v^{p_2}\right)} + \|g\|^2_{L^1\left([0,T],L_x^{q_1}L_v^{q_2}\right)}, \end{aligned}$$

with $s < S = (1-d_2)\min\left\{\frac{1-(d_3-d_4)}{\alpha+1+(d_1-d_2)}, 1\right\} - d_4$, and $\chi_0(\xi) := \sum_{k\in\mathbb{N}} \chi(2^{-k}\xi)$.

The last step is to translate the quadratic form of h back to a norm of velocity average of f.

Step 3: *Derive result back to f.*

With the change of variables again we have

$$\int\int\left|\int F_{s_1}(v)\psi(a(v))\,dv\right|^2 |\xi|^s\,d\xi\,dt$$
$$= \int\int\left|\int h(v')\psi(v')\,dv'\right|^2 |\xi|^s\,d\xi\,dt < \infty,$$

for all $\psi \in H^{3/2}$. By the assumptions that ϕ and Φ are compactly supported in v, one can show

Lemma 2. *There exists $\psi \in H^{3/2}$ such that*

$$\int_0^T \int_{|\xi|\geq 1}\left|\int \tilde{f}\phi\,dv\right|^2 |\xi|^s\,d\xi\,dt \lesssim \int\int\left|\int F_{s_1}(v)\psi(a(v))\,dv\right|^2 |\xi|^s\,d\xi\,dt < \infty$$

for all $s < S$.

This concludes our proof.

Remark 8. *Note that $m(\xi,\zeta)$ to be homogeneous zero in ζ is essential for the commutator to be positive-definite after interacting with the transport operator. In fact, if consider $m(\xi,\zeta) = \frac{\xi}{|\xi|}\cdot\frac{\zeta}{(1+|\zeta|^2)^{\beta/2}}$ with $\beta > 1$,*

$$\xi\cdot\nabla_\zeta m = \frac{|\xi|\left[(1+|\zeta|^2) - \beta|\frac{\xi}{|\xi|}\cdot\zeta|^2\right]}{(1+|\zeta|^2)^{\beta/2+1}}.$$

When ζ is parallel to ξ and $|\zeta|$ is large, it is negative and the argument doesn't work.

The regularization recollects the regularization process in [37]. Here the convolution with $\Phi_{|\xi|^{-s_1}}$, along with the multiplier m_0, show explicitly the interaction between the regularity in x and v.

2.4.1.2 Proof of Lemma 1

Before estimating A_k, let us first show the relation of functions connected through change of variables.

Proposition 1. *Let $a \in Lip(\mathbb{R}^n)$. If $J_{a^{-1}} \in L^\gamma$, the change of variables is bounded from L^p to $L^{(p'\gamma')'}$. Precisely, if $\int \ell(v')\psi(v')\,dv' = \int L(v)\psi(a(v))\,dv$, then*

$$\|\ell\|_{L^{(p'\gamma')'}_{v'}} \lesssim \|L\|_{L^p_v}.$$

Proof. By Hölder's inequality,

$$\int |\psi(a(v))|^{p'}\,dv = \int |\psi(v')|^{p'} J_{a^{-1}}(v')\,dv' \le \|J_{a^{-1}}\|_{L^\gamma} \left(\int |\psi(v')|^{p'\gamma'}\,dv' \right)^{1/\gamma'}.$$

So

$$\begin{aligned}\|\ell\|_{L^{(p'\gamma')'}_{v'}} &= \sup_{\|\psi\|_{L^{p'\gamma'}_{v'}}=1} \left| \int \ell\psi \right| = \sup_{\|\psi\|_{L^{p'\gamma'}_{v'}}=1} \left| \int L(v)\psi(a(v)) \right| \\ &\le \sup_{\|\psi\|_{L^{p'\gamma'}_{v'}}=1} \|L\|_{L^p_v} \|\psi(a(v))\|_{L^{p'}_v} \lesssim \|L\|_{L^p_v}.\end{aligned}$$

□

Remark 9. *If a is one-to-one and $a \in Lip(\mathbb{R}^n, \mathbb{R}^m)$, where $n < m$, the area formula gives*

$$\int |\psi(a(v))|^{p'}\,dv = \int |\psi(v')|^{p'} |J_{a^{-1}}(v')|\,d\mathcal{H}^n(v'),$$

with $|J_{a^{-1}}| = (det(Da^{-1})(Da^{-1})^T)^{1/2}$, and $\mathcal{H}^n$ is a Hausdorff measure of dimension n. This relation would put ℓ in Hausdorff measurable spaces, which are not compatible with our arguments with Fourier analysis in the whole space.

We now use Proposition 1 to estimate A_k term by term when $d_i > 0$ for all $i = 1, 2, 3, 4$. The other cases follow similar calculations.

- For the first term: By the Cauchy-Schwarz inequality, and that $R \cdot \nabla_v G_1^n$ is Calderon-Zygmund operator:

$$\begin{aligned}\int\int \bar{h}\left(\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n \star_{v'} h\right) dv'\,\chi(2^{-k}\xi)d\xi|_{t=0}^{t=T} \\ \leq \|\mathcal{F}_x^{-1}(h\chi(2^{-k}\xi))\|^2_{L^2_{xv'}}|_{t=0}^{t=T},\end{aligned} \tag{2.18}$$

Denote $\mathcal{F}_x^{-1}(\chi)$ by S. By Proposition 1, for each fixed t,

$$\begin{aligned}\|\mathcal{F}_x^{-1}(h\chi(2^{-k}\xi))\|_{L^2_{xv'}} &\lesssim \|S_{2^{-k}} \star_x f_k \star_v \Phi_{2^{-ks_1}}\|_{L^2_x L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}} \\ &\lesssim 2^{kn\left(\frac{1}{p_1}-\frac{1}{2}\right)+ks_1\left(\frac{1}{p_2}-\frac{\gamma-2}{2\gamma-2}\right)}\|f_k\|_{L_x^{p_1}L_v^{p_2}}.\end{aligned}$$

Plug this back into (2.18) and we have

$$\begin{aligned}\int\int \bar{h}\left(\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n \star_{v'} h\right) dv'\,\chi(2^{-k}\xi)d\xi|_{t=0}^{t=T} \\ \lesssim 2^{kn\left(\frac{2}{p_1}-1\right)+ks_1\left(\frac{2}{p_2}-\frac{\gamma-2}{\gamma-1}\right)}\|f_k\|^2_{L_x^{p_1}L_v^{p_2}}|_{t=0}^{t=T}.\end{aligned}$$

- For the second term:

$$\begin{aligned}&\int \bar{h}\left(\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n \star_{v'} k^1\right) dv'\,\chi(2^{-k}\xi)d\xi\,dt \\ &\lesssim \int \|S_{2^{-k}} \star_x f_k \star_v \Phi_{2^{-ks_1}}\|_{L^2_x L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}} \\ &\qquad \|2^{k\alpha s_1}S_{2^{-k}} \star_x g_k \star_v ((-\Delta_v)^\beta\Phi)_{2^{-ks_1}}\|_{L^2_x L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\,dt,\end{aligned}$$

which is of the order of

$$2^{kn\left(\frac{1}{p_1}+\frac{1}{q_1}-1\right)+ks_1n\left(\frac{1}{p_2}+\frac{1}{q_2}-\frac{\gamma-2}{\gamma-1}\right)+k\alpha s_1}\int \|f_k\|_{L_x^{p_1}L_v^{p_2}}\|g_k\|_{L_x^{q_1}L_v^{q_2}}\,dt.$$

- For the last term:

$$\begin{aligned}&\int \bar{h}\left(\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n\star_{v'}k^2\right)dv'\,\chi(2^{-k}\xi)d\xi\,dt\\ &\lesssim \int \|S_{2^{-k}}\star_x f_k\star_v\Phi_{2^{-ks_1}}\|_{L_x^2L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\\ &\qquad\|S_{2^{-k}}\star_x Com^1\star_v\Phi_{2^{-ks_1}}\|_{L_x^2L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\,dt.\end{aligned}$$

Because Φ is compactly supported, $\Phi_{2^{-ks_1}}(v-w)$ forces $|v-w|\lesssim 2^{-ks_1}$. Moreover since a is Lipschitz, $|a(v)-a(w)|\lesssim 2^{-ks_1}$.

$$\begin{aligned}&\|S_{2^{-k}}\star_x Com^1\star_v\Phi_{2^{-ks_1}}\|_{L_x^2L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\\ &=\left\|\int 2^k(a(v)-a(w))\cdot(f\star_x(\nabla_xS)_{2^{-k}})(w)\phi(w)\right.\\ &\qquad\left.\Phi_{2^{-ks_1}}(v-w)\,dw\right\|_{L_x^2L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\\ &\lesssim 2^{k-ks_1}\|f\star_x(\nabla_xS)_{2^{-k}}|\star_v\Phi_{2^{-ks_1}}\|_{L_x^2L_v^{\frac{2(\gamma-1)}{(\gamma-2)}}}\\ &\lesssim 2^{kns_1\left(\frac{1}{p_2}-\frac{(\gamma-2)}{2(\gamma-1)}\right)+kn\left(\frac{1}{p_1}-\frac{1}{2}\right)+k(1-s_1)}\|f\|_{L_x^{p_1}L_v^{p_2}}.\end{aligned}$$

Hence

$$\begin{aligned}&\int \bar{h}\left(\frac{\xi}{|\xi|}\cdot\nabla_{v'}G_1^n\star_{v'}k^3\right)dv'\,\chi(2^{-k}\xi)d\xi\,dt\\ &\lesssim 2^{kn\left(\frac{2}{p_1}-1\right)+ks_1n\left(\frac{2}{p_2}-\frac{\gamma-2}{\gamma-1}\right)+k(1-s_1)}\int \|f\|^2_{L_x^{p_1}L_v^{p_2}}\,dt.\end{aligned}$$

Combining all estimates,

$$
\begin{aligned}
|A_k| &\lesssim 2^{kn\left(\frac{2}{p_1}-1\right)+ks_1n\left(\frac{2}{p_2}-\frac{\gamma-2}{\gamma-1}\right)}\|f_k\|^2_{L^{p_1}_x L^{p_2}_v}\Big|_{t=0}^{t=T} \\
&\quad + 2^{kn\left(\frac{1}{p_1}+\frac{1}{q_1}-1\right)+ks_1n\left(\frac{1}{p_2}+\frac{1}{q_2}-\frac{\gamma-2}{\gamma-1}\right)+k\alpha s_1}\int \|f_k\|_{L^{p_1}_x L^{p_2}_v}\|g_k\|_{L^{q_1}_x L^{q_2}_v}\,dt \\
&\quad + 2^{kn\left(\frac{2}{p_1}-1\right)+ks_1n\left(\frac{2}{p_2}-\frac{\gamma-2}{\gamma-1}\right)+k(1-s_1)}\int \|f_k\|^2_{L^{p_1}_x L^{p_2}_v}\,dt.
\end{aligned}
$$

2.4.1.3 Proof of Lemma 2

Choose two smooth functions ψ_1 and ψ_2 such that $\psi_1(a(v)) \equiv 1$ on $v \in B(0,1)$, and $\psi_2(v) \equiv 1$ on $v \in B(0,2)$. We put ψ_1 in the place of ψ and plug in ψ_2 as an auxiliary function at no cost since it's 1 on the support of ϕ. Then

$$
\begin{aligned}
\infty &> \int\int_{|\xi|\geq 1}\left|\int F_{s_1}(v)\psi_2(v)\,dv\right|^2|\xi|^s\,d\xi\,dt \\
&= \int\int_{|\xi|\geq 1}\left|\int \mathcal{F}_\zeta(\tilde{f}\phi)(\zeta)\mathcal{F}_\zeta(\Phi)(\zeta|\xi|^{-s_1})\mathcal{F}_\zeta(\psi_2)(\zeta)\,d\zeta\right|^2|\xi|^s\,d\xi\,dt \\
&= \int\int_{|\xi|\geq 1}\left|\int (\tilde{f}\phi)(\Phi_{|\xi|^{-s_1}}\star_v\psi_2)\,dv\right|^2|\xi|^s\,d\xi\,dt.
\end{aligned}
$$

Because $\psi_2 \equiv 1$ on $B(0,2)$ and $|v-w| \leq |v|+|w| \leq 1+|\xi|^{-s_1} \leq 2$ when $|\xi| \geq 1$,

$$
\begin{aligned}
(\Phi_{|\xi|^{-s_1}}\star_v\psi_2)(v) &= \int |\xi|^{ns_1}\Phi(w|\xi|^{s_1})\psi_2(v-w)\,dw \\
&= \int |\xi|^{ns_1}\Phi(w|\xi|^{s_1})\,dw = \|\Phi\|_{L^1_v} \quad \text{for all } |v|\leq 1.
\end{aligned}
$$

So finally we reach

$$
\int_0^T\int_{|\xi|\geq 1}\left|\int \tilde{f}\phi\,dv\right|^2|\xi|^s\,d\xi\,dt < \infty
$$

for all $s < S$.

2.4.2 Proof of Theorem 6

This proof is essentially the same as Theorem 2, but with a different change of variable. After Step 1, instead of $v \mapsto v' = a(v)$, we make $v \mapsto \lambda = a(v) \cdot \frac{\xi}{|\xi|}$ for each fixed ξ. For convenience, let us denote $\epsilon = |\xi|^{-s_1}$. So parallel to (2.14), we have

$$\partial_t h_\epsilon + i\lambda|\xi|h_\epsilon = k^1_\epsilon + k^2_\epsilon \tag{2.19}$$

in the sense of distribution, where $h_\epsilon, k^1_\epsilon, k^2_\epsilon, k^3_\epsilon$ are defined as following:

$$\int F_\epsilon(v)\psi\left(a(v)\cdot\frac{\xi}{|\xi|}\right)dv = \int h^\xi_\epsilon(\lambda)\psi(\lambda)\,d\lambda.$$

$$\int k^1_\epsilon(\lambda)\psi(\lambda)\,d\lambda = \int \left[((-\Delta_v)^{\alpha/2}\tilde{g}\phi)\star_v \Phi_\epsilon\right](v)\psi\left(a(v)\cdot\frac{\xi}{|\xi|}\right)dv,$$

and

$$\int k^2_\epsilon(\lambda)\psi(\lambda)\,d\lambda = \int Com^1(v)\psi\left(a(v)\cdot\frac{\xi}{|\xi|}\right)dv.$$

The subscript ϵ is to emphasize the dependence on ξ.

Thanks to the non-degeneracy condition with $\nu = 1$, this change of variables preserves L^p norm:

Proposition 2. *Let a be Lipschitz and satisfy (2.8) with $\nu = 1$. Let $\psi : \mathbb{R} \to \mathbb{R}$. Then for all $\sigma \in \mathbb{S}^m$, $1 \le p \le \infty$,*

$$\|\psi(a(v)\cdot\sigma)\|_{L^p_v} \le c_0\|\psi\|_{L^p_\lambda}.$$

And hence if $\int L(v)\psi\left(a(v)\cdot\sigma\right)dv = \int \ell^\sigma(\lambda)\psi(\lambda)\,d\lambda$, then

$$\|\ell^\sigma\|_{L^p_\lambda} \lesssim \|L\|_{L^p_v}.$$

Consider

$$\int \bar{h_\epsilon^\xi}(\lambda)\frac{1}{i}(\partial_\lambda G_1^1)(\lambda-\alpha)h_\epsilon^\xi(\alpha)\,d\alpha\,d\lambda.$$

Then similar estimations and procedures lead to

$$\int |\xi|^{1/(\alpha+1)}\bar{h_\epsilon^\xi}(\lambda)G_3^1(\lambda-\alpha)h_\epsilon^\xi(\alpha)\,d\alpha\,d\lambda\,dt\,d\xi<\infty.$$

One can conclude the result from here by following Step 3 in the proof of Theorem 2.

Notice here everything is in one dimension for each fixed ξ. And because of the L^2 setting, it is valid to do calculation in the level of (v,ξ).

The last thing to check is Proposition 2.

Proof of Proposition 2. When $p=\infty$, the result is straightforward. For $1\le p<\infty$. (2.8) implies for any interval I, we have

$$m(\{v\in B(0,1):a(v)\cdot\sigma\in I\})\le c_0m(I).$$

By a standard approximation from intervals to general measurable sets, one has for any measurable set A,

$$m(\{v\in B(0,1):a(v)\cdot\sigma\in A\})\le c_0m(A).$$

From this we see the relation between the distribution functions of $\psi(a(v)\cdot\sigma)$ and ψ:

$$\begin{aligned}d_{\psi(a(v)\cdot\sigma)}(s)&=m(\{v\in B:a(v)\cdot\sigma\in\{\lambda:|\psi(\lambda)|>s\}\})\\&\le c_0m(\{\lambda:|\psi(\lambda)|>s\})=c_0d_\psi(s).\end{aligned}$$

Therefore

$$\begin{aligned}\|\psi(a(v)\cdot\sigma)\|_{L^p_v} &= p^{1/p}\left(\int_0^\infty \left[d_{\psi(a(v)\cdot\sigma)}(s)^{1/p}s\right]^p \frac{ds}{s}\right)^{1/p} \\ &\le p^{1/p}\left(\int_0^\infty \left[c_0^{1/p}d_\psi(s)^{1/p}s\right]^p \frac{ds}{s}\right)^{1/p} = c_0\|\psi\|_{L^p_\lambda}.\end{aligned} \tag{2.20}$$

And by duality,

$$\begin{aligned}\|\ell^\sigma\|_{L^p_\lambda} = \sup_{\|\psi\|_{L^{p'}_\lambda}=1}\left|\int \ell^\sigma\psi\right| &= \sup_{\|\psi\|_{L^{p'}_\lambda}=1}\left|\int L(v)\psi\left(a(v)\cdot\sigma\right)\right| \\ &\le \sup_{\|\psi\|_{L^{p'}_\lambda}=1}\|L\|_{L^p_v}\left|\left|\psi\left(a(v)\cdot\sigma\right)\right|\right|_{L^{p'}_v} \\ &\le c_0\sup_{\|\psi\|_{L^{p'}_\lambda}=1}\|L\|_{L^p_v}\|\psi\|_{L^{p'}_\lambda} = c_0\|L\|_{L^p_v},\end{aligned}$$

where the first inequality is due to the Hölder's inequality, and second by (2.20). This concludes our proof for Proposition 2 and hence Theorem 6.

Acknowledgement

The authors would like to thank M. Machedon for sharing his insight on our method.

Chapter 3: Unmixing property on a 2-dimensional torus

3.1 Introduction

Fluid mixing not only provides rich mathematical problems, especially in the dynamical systems and partial differential equations, but also is an important topic with the applications in many fields, such as chemistry, engineering, and atmospheric and oceanic science.

There is abundant literature in fluid mechanics studying mixing phenomena. A fluid system is often described by partial differential equations solved by tracers, under the assumption that the interaction between the tracers and the flow is negligible. For instance, the transport equation of a divergence free vector field is used to describe the passive scalar mixing [5, 75]. The study in relaxation enhancing considered the transport equation with a diffusion term [29]. Mixing for the stochastic 2-dimensional Navier-Stokes equation is also discussed in [24].

We consider the passive scalar mixing behavior of the incompressible flow f on the n-dimensional torus $\mathbb{T}^n = [-\pi, \pi]^n$, described by the divergence free transport equation,

$$\begin{cases} \partial_t f + \mathbf{a} \cdot \nabla f = 0, \\ f(x, 0) = f_0, \end{cases} \tag{3.1}$$

where $\nabla \cdot \mathbf{a} = 0$ and $f_0 \in L^2(\mathbb{T}^n)$.

The well-posedness of the Cauchy problem (3.1) has been carefully studied and many sufficient conditions have been discovered. After the classical Cauchy-Lipschitz theory, the condition on $\mathbf{a}$ was extended to $W^{1,1}_{loc,x}$ by [37] and later further to BV class by [6]. For the 2-dimensional case an even weaker condition is required, owing to the natural Hamiltonian structure. The Sobolev or BV class condition can be replaced by various assumptions on the direction of $\mathbf{a}$ [20, 60] or the weak Sard condition introduced in [3].

Assume the Cauchy problem (3.1) corresponding to a vector field $\mathbf{a}$ has a unique solution f for all $f_0 \in L^2(\mathbb{T}^2)$, then we say $\mathbf{a}$ is **mixing** if

$$\lim_{t\to\infty} \int_{\mathbb{T}^n} \Phi(x)\, \theta(x,t)\, dx = 0 \tag{3.2}$$

for all f_0 and $\Phi \in L^2(\mathbb{T}^2)$ with $\theta(x,t) := f(x,t) - \int_{\mathbb{T}^n} f_0(x,t)\, dx$.

Besides the above definition for mixing, other definitions were proposed in the quest of a proper norm for quantifying the mixing phenomenon. For example, [84] proposed $H^{-1/2}$ as a mix norm, which was later extended to H^{-s} for all $s > 0$ and was proved to be equivalent to (3.2) in [75]. That is, it was shown that for all $s > 0$,

$$\|\theta\|_{H^{-s}} \to 0 \quad \text{as } t \to 0, \tag{3.3}$$

is equivalent to (3.2) for all $s > 0$. Another example of mixing scales can be found in [23], related to set rearrangements. This mixing scale is not equivalent to H^{-s}, yet still closely related; see for instance [4, 108].

This work focuses on the **unmixing** property of the incompressible flows on a

2-dimensional torus. We take the advantage of the Hamiltonian structure and show a perturbation result of unmixing. The key observation is that the quantities near the local maximums or minimums of the Hamiltonian (stream) function cannot escape and therefore the corresponding vector field is not mixing. We shall show that this phenomenon (hence the unmixing property) is preserved under a small perturbation of the sup-norm.

This chapter is organized as follows: main results are placed in Section 3.2 by first giving a perturbation result for unmixing property in a deterministic setting in Section 3.2.1; a simple consequence of the main theorem in a probabilistic setting shall be presented in Section 3.2.2. This probabilistic result offers an interesting observation for the unmixing property from the perspective of Fourier analysis.

3.2 Main results

3.2.1 Deterministic result

Definition 2. *We say $\psi \in H^1(\mathbb{T}^2)$ is a **stream function** of the vector field $\mathbf{a} = (a_1, a_2) \in L^2(\mathbb{T}^2)$ on an open set $\mathcal{O}$, if on $\mathcal{O}$*

$$\mathbf{a} = \nabla^{\perp}\psi := (\partial_2\psi, -\partial_1\psi) \tag{3.4}$$

in the sense of distribution.

Because of the Hamilton equations, the stream function has the following important property:

Proposition 3. *Let ψ be the stream function corresponding to $\mathbf{a}$ on $\mathcal{O}$. If $\phi \in C^1$ and $(supp\,\phi) \cap \mathcal{O}^c$ is measure zero for almost all $t \in [0,T]$, then $\int_{\mathbb{T}^2} \phi(\psi(x)) f(t,x)\,dx$ stays constant for all $t \in [0,T]$, where f solves (3.1).*

This proposition shows that ψ is constant on each characteristic of the corresponding vector field a.

Proof of Proposition 3. By (3.4), we have

$$\partial_t \int_{\mathbb{T}^2} \phi(\psi(x)) f(t,x)\,dx = -\int_{\mathbb{T}^2} \phi(\psi(x)) div_x(\mathbf{a}f)\,dx$$
$$= \int_{\mathbb{T}^2} \phi'(\psi(x)) \nabla_x \psi \cdot \mathbf{a} f\,dx = 0 \quad (3.5)$$

□

As it is assumed in our case that ψ is time-independent, this statement can be reworded as that the characteristics are retained in level sets of ψ. We will use this proposition to show that a vector field is unmixing if the corresponding ψ has a local maximum or minimum. Furthermore, this phenomenon can be preserved under a sup-norm perturbation:

Theorem 7. *Let $\mathbf{a}^0$ be a vector field on $\mathbb{T}^2$ and ψ^0 be a continuous stream function of $\mathbf{a}^0$ on some open subset $\mathcal{O} \subset \mathbb{T}^2$ with a local maximum or minimum inside $\mathcal{O}$. Then there exists $\varepsilon > 0$ such that the following statement holds:*

If $\|\psi - \psi^0\|_{L^\infty(\mathbb{T}^2)} < \varepsilon$ and the Cauchy problem (3.1) corresponding to $\mathbf{a} := \nabla^\perp \psi$ is well-posed, then $\mathbf{a}$ is unmixing. (3.6)

Proof of Theorem 7. Without loss of generality, we assume ψ^0 has a local maximum M at the origin in $\mathcal{O}$. Let ψ be a stream function with

$$||\psi - \psi^0||_\infty < \varepsilon := \frac{\alpha M}{4} \tag{3.7}$$

and (3.1) corresponding to $\mathbf{a} := \nabla^\perp \psi$ is well-posed. We shall show that $\mathbf{a}$ is unmixing, by constructing an initial condition f_0 such that (3.2) fails when $\Phi = \psi \in L^2(\mathbb{T}^2)$.

Denote $U^n = \{\psi^0 > M(1 - n\alpha)\}$, for all $n = 1, 2, ..., 6$, where $\alpha > 0$ is fixed such that $U^6 \subset \mathcal{O}$. Let $\tilde{V} = \{M(1 - 4\alpha) \geq \psi^0 > h\} \subset U^5 \setminus U^4$, where h is chosen so that $|\tilde{V}| = |U^1|$. We define

$$f_0 := f_0^+ - f_0^- = \mathcal{X}_{U^1} - \mathcal{X}_{\tilde{V}}.$$

Note that $\int_{\mathbb{T}^2} f_0 \, dx = 0$. Denote f^+ the solution of (3.1) with the initial condition f_0^+, and f^- with f_0^-.

Let $\phi(x)$ be a C^1 function such that $\phi(x) = 0$ if $x > M(1 - \frac{3}{2}\alpha)$, and $\phi(x) > 0$ if $x \leq M(1 - \frac{3}{2}\alpha)$. Then by Proposition 3,

$$\int \phi(\psi) f^+(t, x) \, dx \equiv \int \phi(\psi) f_0^+ = 0$$

for all t, which implies that *supp* $f^+ \subset U^2$ for all $t \in [0, \infty)$. With the same argument, one has *supp* $f^- \subset U^6 \setminus U^3$ for all t.

Because of (3.7), $\psi > M(1 - \frac{7}{4}\alpha)$ on U^2 and $\psi < M(1 - \frac{11}{4}\alpha)$ on $U^6 \setminus U^3$. We therefore derive

$$\int_{\mathbb{T}^2} \psi f = \int_{\mathbb{T}^2} \psi(f^+ - f^-) > M\alpha |U^1| > 0$$

for all t and the proof is concluded.

□

3.2.2 Simple consequence in a probabilistic setting

With above result for unmixing, we would like use it to give a quantitative statement about how many vector fields are unmixing. One way to do so is by introducing a probability into the problem. We take a classical approach, and consider the vector field $\mathbf{a}$ to be a random Fourier series. The study of random Fourier series can be traced at least back to 1930s; see for example [89]. We refer to [70] for more references.

We consider the vector fields in the following form:

$$\mathbf{a} = (a_1, a_2) = (\gamma_0^1, \gamma_0^2) + \sum_{k\in\mathbb{Z}^2\setminus\{0\}} k^{\perp} \frac{i\gamma_k}{(1+|k|)^{\theta}} e^{ik\cdot x}, \tag{3.8}$$

where for all $k \in \mathbb{Z} \setminus \{0\}$, $k^{\perp} \cdot k = 0$. As we concern only real vector fields, the condition $\bar{\gamma}_k = \gamma_{-k}$ will always be assumed in this discussion and we shall assume

$$\begin{aligned}&\gamma_0^1, \gamma_0^2, \{\gamma_k\}_{k\in\mathbb{H}} \text{ are independent random variables, where } \mathbb{H} := \\ &\{(x,y) \in \mathbb{Z}^2 \setminus \{0\} : y > 0 \text{ or } x > 0 \text{ when } y = 0\}.\end{aligned} \tag{3.9}$$

Combining Theorem 7 and some additional conditions on the distributions of the Fourier coefficients, we show that the probability of unmixing is positive:

Corollary 3. *Let $\theta > 3$, and γ_0^1, γ_0^2, $\{\gamma_k\}_{k\in\mathbb{H}}$ be a sequence of independent and identically distributed random variables with zero expectation, where*

$$\mathbb{H} := \left\{(x,y) \in \mathbb{Z}^2 : y > 0 \textit{ or } x > 0 \textit{ when } y = 0\right\}.$$

If there exists a number $\epsilon_0 > 0$ such that

$$\mathbb{E}(\chi_{\gamma_k\in(\alpha,\beta)}) > 0 \quad \textit{for all } -\epsilon_0 < \alpha < \beta < \epsilon_0, \tag{3.10}$$

then there exists a set S of unmixing vector fields such that $P(S) > 0$.

Notice the conditions of independence, zero mean, and decay rate θ of Fourier coefficients guarantee that $\mathbf{a}$ in the form of (3.8) is H^1 and hence (3.1) is well-posed by the DiPerna-Lions theory almost surely.

The assumption (3.10) on the other hand guarantees the *contribution from different frequencies*, which is crucial for this result. In fact, if $\gamma_k \equiv 0$ for all $k \in \mathbb{Z}^2 \backslash \{0\}$, then with probability one the flow on two-dimensional torus is ergodic, which is the weakest notion of mixing in the ergodic sense. (3.10) excludes this special case, and ensures the infinite Fourier series $\sum_{k\in\mathbb{Z}^2\backslash\{0\}} k^\perp \frac{i\gamma_k}{(1+|k|)^\theta} e^{ik\cdot x}$ has a non-negligible influence on our system. This effect therefore makes the positive probability of unmixing possible.

Proof of Corollary 3. We first check that $\mathbf{a} \in H^1$ almost surely:

$$\mathbb{E}(||\mathbf{a}||_{H^1}^2) \lesssim \mathbb{E}\left(\int \sum_{i,j} \frac{\gamma_{k_i}\bar{\gamma_{k_j}}}{(1+|k_i|)^{(\theta-2)}(1+|k_j|)^{\theta-2}} e^{i(k_i-k_j)\cdot x}\right)$$
$$\lesssim \sum_{i,j} \mathbb{E}\left(\frac{\gamma_{k_i}\bar{\gamma_{k_j}}}{(1+|k_i|)^{(\theta-2)}(1+|k_j|)^{\theta-2}}\right).$$

By the assumption that $\mathbb{E}(\gamma_k) = 0$ and γ_k, γ_ℓ are independent when $k \neq \pm\ell$, only those terms with duplicated index would survive. Therefore,

$$\mathbb{E}(||\mathbf{a}||_{H^1}^2) \lesssim \sum_i \mathbb{E}\left(\frac{|\gamma_{k_i}|^2}{(1+|k_i|)^{2(\theta-2)}}\right) < \infty.$$

Therefore, $\mathbf{a} \in H^1$ almost surely.

We now write down an explicit set of unmixing vector fields. Consider

$$\mathbf{a}^0 = \frac{1}{B2^{\theta-1}} \begin{pmatrix} -\sin y \\ \\ \sin x \end{pmatrix} = \sum_{|k|=1} \begin{pmatrix} k_2 \\ \\ -k_1 \end{pmatrix} \frac{iB^{-1}}{(1+|k|)^\theta} e^{ik\cdot x},$$

where B is chosen so that $\frac{1}{B} < \epsilon_0$. We choose a stream function to be

$$\psi^0(x,y) = \frac{1}{B2^{\theta-1}}(\cos x + \cos y).$$

Fix the open set $\mathcal{O} = [-\pi/2, \pi/2]^2$. By Theorem 7, there exists $\varepsilon > 0$ such that (3.6) holds. Define a set of vector fields:

$$S_L := \Bigg\{ \mathbf{a} \text{ in the form of (3.8)} : \pi(|\gamma_0^1| + |\gamma_0^2|) + \sum_{|k|=1} \frac{|\gamma_k - \frac{1}{B}|}{2^\theta} + \sum_{1<|k|\le L} \frac{|\gamma_k|}{(1+|k|)^\theta} < \varepsilon/2 \text{ and } \sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta} < \varepsilon/2 \Bigg\}, \quad (3.11)$$

where $L \in \mathbb{N}$ will be chosen later. We claim that every $\mathbf{a} \in S_L$ is unmixing.

In fact, for all $\mathbf{a} \in S$, we can define a corresponding stream function $\psi : \mathbb{T}^2 \to \mathbb{R}$ by:

$$\psi(x,y) = \Phi(x,y) + \sum_{k\in\mathbb{Z}^2\setminus\{0\}} \frac{\gamma_k}{(1+|k|)^\theta} e^{ik\cdot x}, \quad (3.12)$$

where

$$\Phi(x,y) = \begin{cases} \gamma_0^1 y - \gamma_0^2 x, & \text{for } (x,y) \in \mathcal{O} \\ \tilde{\Phi}(x,y), & \text{for } (x,y) \in \mathcal{O}^c \end{cases}$$

where $\tilde{\Phi}$ is defined smoothly such that Φ is continuous and its absolute value decreases to 0 when x and y go to $\pm\pi$.

By the definitions of ψ^0, ψ and S_L,

$$|\psi - \psi^0| \le \pi(|\gamma_0^1| + |\gamma_0^2|) + \sum_{|k|=1} \frac{|\gamma_k - \frac{1}{B}|}{2^\theta} + \sum_{k\in\mathbb{Z}^2\setminus\{0\}, |k|\ne 1} \frac{|\gamma_k|}{(1+|k|)^\theta} < \varepsilon.$$

So $\mathbf{a}$ is unmixing by Theorem 7.

The last thing is to show $P(S_L) > 0$ for some large enough $L \in \mathbb{N}$.

$$P(S_L) = P\left(\pi(|\gamma_0^1| + |\gamma_0^2|) + \sum_{|k|=1} \frac{|\gamma_k - \frac{1}{B}|}{2^\theta} + \sum_{1<|k|\le L} \frac{|\gamma_k|}{(1+|k|)^\theta} < \varepsilon/2\right) \cdot P\left(\sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta} < \varepsilon/2\right) := Q_1 \cdot Q_2, \quad (3.13)$$

for any integer $L > 1$.

Q_1 is always positive thanks to (3.10). For Q_2, notice that

$$\mathbb{E}\left(\sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta}\right) \lesssim \sum_{|k|>L} \frac{1}{(1+|k|)^\theta} \lesssim \int_L^\infty \frac{1}{(1+r)^\theta} r dr \lesssim L^{2-\theta}.$$

Hence by choosing L large enough, we can make $\mathbb{E}\left(\sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta}\right) < \varepsilon/2$, which implies that $Q_2 > 0$. (Otherwise $P\left(\sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta} \ge \varepsilon/2\right) = 1$ and $\mathbb{E}\left(\sum_{|k|>L} \frac{|\gamma_k|}{(1+|k|)^\theta}\right) \ge \varepsilon/2$, which is a contradiction.) Therefore, $P(S_L) = Q_1 \cdot Q_2 > 0$ for a large enough L and the proof is concluded.

□

Chapter 4: Modeling: Memory effects on animal migrations

4.1 Introduction

The interaction between dynamic landscapes and animal movements has been an important research topic in biology, particularly with regard to the process of migration. For instance, insufficient spatiotemporal change in the distribution of resources may 'short-circuit' migration in some seasons [14]. In other cases, the age structure of an animal population can create new migratory patterns in response to environmental changes [100]. From theoretical work, we know that gathering of nonlocal information is beneficial for resource uptake in dynamic landscapes [44].

Among the long list of factors that one could consider as a variable in this rich topic, the effects of spatial memory on animal movements in dynamic landscapes has attracted considerable recent attention. Many works have demonstrated the essential role of memory in animal migration patterns [1,22,45]. A variety of models have been proposed to explore this memory effect, some of which have been quite complex [16,94]. Memory and environmental persistence are both clearly connected with migratory movement [17]. However, even with abundant existing results, the underlying memory mechanism and its relation with animal movement remain unclear.

The goal of this work is to obtain a better understanding of the effects of memory on animal migration patterns. For this purpose, we propose a memory-driven movement model, consisting of a stochastic transport equation, evolution equations for the memory and fitness, and an eikonal equation with a potential depending on the animal's perception and memory. Our model explicitly describes a wide range of different memory mechanisms, and the corresponding migration patterns can be directly observed by numerical simulations.

Migration patterns have long been known to follow seasonal changes in the environment and it is natural to expect that such periodic changes in the environment are the main factor contributing to such migrations [14, 43]. We thus test our model under a simple, idealized time-periodic environment to investigate memory effects on the migration patterns.

The use of the eikonal equation was inspired by the Hughes model for pedestrians [62, 63]. There are many works in the Hughes model from both analytical [7, 49] and numerical aspects [27, 102]. The Hughes model contains a conservation law for pedestrian flow, and an eikonal equation with a potential depending on the density of pedestrians. In our case, the potential of the eikonal equation depends on the animal's memory and perception. The article [106] combined both the conservation law of pedestrians and an eikonal equation with memory to discuss the memory effect for pedestrian flows.

One advantage of using the eikonal equation is that it provides a natural interpretation for an animal's decision-making process under the context of optimal control theory. This view of optimal individual-level movement strategies complements mathematical theory on optimal population-level movement that has sought to identify the best movement strategies for different resource landscapes in an evolutionary context using invasibility criteria (e.g., [26, 73]). As an optimal control problem at the individual level, an animal's migratory journey consists of a series of movements in which the animal relocates to the region with the best resources by choosing an optimal path that minimizes a certain cost function. The cost function therefore offers an easy way of introducing environmentally based preferences in the individual's movement. A similar concept of utilizing a cost function for memory-based movements can also be seen in [69]. Another advantage of the eikonal equation is that efficient algorithms are available; see for example [28, 96, 107]. These algorithms help accelerate our computations and make our numerical simulations much less expensive.

This paper is organized as follows. Our model is introduced in Section 2. Its application to the migration behaviors under periodic environments is in Section 3. A discussion for model components and time scales of memory can be found in Section 4. Some examples of simulations under more complicated environments are presented in Section 5. Finally, the conclusion is in Section 6.

4.2 Mathematical model

4.2.1 Overview of the model

For this section we shall construct a model for an individual's movement, which depends on its health status, the local environment conditions, and its memory for the global environment. The dynamics follow the following assumptions:

- The animal tries to move to, or stay in, the places with the most resources that it remembers.
- An animal's desire to move depends on its fitness and the condition of the animal's current location. We assume one would be less likely to move if it is in good health, or its surrounding is full of resources.
- The movement has a small stochastic effect for the explorations for the local environment.

The dynamics are recorded by the individual's position $X(t)$ for time $t \in \mathbb{R}^+$. The first two important factors that affect our dynamics are the individual's fitness and the environment condition. We consider the fitness $P(t)$, and the environment $E(t,x)$ on $\mathbb{R}^+ \times \mathbb{R}^n$. The value of $E(t,x)$ indicates the condition of environment at time t and location x. The larger the value is, the more resources (or fewer predators) are available for the individual. P is therefore evolving according to the condition of local environment,

$$\frac{dP}{dt} = E(t, X(t))(\overline{P} - P(t)),$$

where $\overline{P} > 0$ is the optimal fitness status an animal can have. Our description for P and E are simplified. As our main interest for this model is the effect of memory on movement, we only keep those parts necessary to our focus.

We model the dynamics with the above hypotheses by the following stochastic differential equation:

$$dX = \sigma dW_t + \chi(P(t), E(t, x))v \, dt,$$

where $\sigma > 0$, W_t is the Brownian motion, and $\chi(P, E) := (\overline{P} - P)e^{-E}$ is called the desire function, which modifies the magnitude of velocity. Consistent with our second assumption of the dynamics, the value of χ is close to 0 when P is close to $\overline{P}$, or when E is large.

The velocity v would be chosen according to the information in memory and perception. We shall introduce our model for memory and perception in Section 2.2, and clarify the choice of velocity in Section 2.3.

4.2.2 Mechanism of memory

While a memory mechanism could be quite complicated (e.g. [16]), here we extract only some basic features that we consider important for our purpose. The assumptions are as follows.

- A memory system consists of multiple channels of memory.

- Each channel of memory fades over time with a rate depending on the intensity of the memory. The stronger the memory intensity is, the slower it would be forgotten. The weaker, the faster.

- Each channel is updated independently over time with new information gathered by the individual within its perception range.

We assume all memory channels operate on the same principle but with different decay and update rates. We first clarify the evolution of each channel, and finish this subsection with a description of a whole memory system.

4.2.2.1 Evolution of one memory channel.

One memory channel is modeled by a memory function $M(t,x)$ on $\mathbb{R}^+ \times \mathbb{R}^n$. The value reflects how the individual remembers the situation of environment at time t and point x.

The evolution of memory contains two terms, one is losing information, another is gaining. Each channel is characterized by two positive indices, the decay rate d and update rate u. We assume the two rates are in the same order, otherwise the channel would fail to capture information correctly over time.

What the second assumption above suggests is a nonlinear term for the fading memory. For the desired behavior we choose the function $-sgn(M)\sqrt{|M|}$. (In fact every function in the form $-sgn(M)|M|^s$, $0 < s < 1$ will do.) $-sgn(M)$ guarantees positive memory decays and negative memory increases.

Comparing to the linear function M, which has its slope identically 1, the function $\sqrt{|M|}$ possesses the characteristic that when $|M|$ is large, its slope is smaller than 1, while when $|M|$ is small, it is larger than 1. This matches our description that when the intensity of memory $|M|$ is large, the change of the forgetting rate is slower than when the strength is small. Moreover, another feature of using $-sgn(M)\sqrt{|M|}$, is any memory with finite intensity shall return to zero within finite time.

The memory update is assumed to depend on the individual's perception of the actual environment. To introduce this factor, we define a perceptual kernel $K(x, y) = k(|x - y|)$, where k is a positive function on $\mathbb{R}$, decreases to zero within a finite distance, and with maximum 1. The magnitude of $K(x, y)$ represents the percentage of information for $E(x, t)$ that an animal can gather when standing at location y.

Combining the above discussion, the evolution of M is governed by the following equation:

$$\partial_t M(t, x) = \tau^{-1} \left[-d \cdot sgn(M)\sqrt{|M|} + u \cdot K(X(t), x)(E(t, x) - M(t, x)) \right], \quad (4.1)$$

where τ is the time scale of this channel. The introduction of τ is for convenience for later discussion when multiple channels are present.

The perception can also be included in our definition as one memory channel, simply by taking its time scale close to zero. Indeed, when τ tends to zero, the memory is forgotten and updated almost immediately. In this case, the corresponding memory function works just like visual perception, which receives instant information for nearby landscapes, but with almost no persistence. As a result, this channel attains almost the same value as the environment function within its perceptual range.

4.2.2.2 Description of a memory system

We call a collection of independent memory channels a *memory system*. Assume we have m channels, $M_i(t,x)$ where $i = 1, ..., m$. Each channel is tagged with a decay rate d_i, an update rate u_i, a time scale τ_i and a perceptual kernel K_i. And each M_i is governed by the following evolution equation:

$$\partial_t M_i(t,x) = \tau_i^{-1}\left[-d_i \cdot sgn(M_i)\sqrt{|M_i|} + u_i \cdot K_i(X(t),x)(E(t,x) - M_i(t,x))\right], \tag{4.2}$$

for $i = 1, 2, ..., m$.

4.2.3 Choice of velocity and optimal control

An animals' decision-making is modeled in the context of optimal control theory, with a cost function depending on the memory and perception of its environment. Precisely, we consider the Hamilton-Jacobian-Bellman equation:

$$\partial_t \psi = \frac{|\nabla_x \psi|^2}{2\lambda} - exp(-H(t,x)), \tag{4.3}$$

where $\lambda > 0$ is a fixed parameter, and $H(t,x) = \sum_{i=1}^m w_i(t, X(t), x) M_i(t,x)$, with weight functions w_i with $(\sum_{i=1}^m w_i)(t,y,x) \equiv 1$, for all t, x, y. The value of $H(t,x)$

represents how an individual evaluates the location x at time t, using the information it gathered and stored in its memory system. The protocol for environment assessment is encoded in the weight functions.

A classical argument in optimal control theory [13] shows that the solution ψ in (4.3) is the value realizing the minimum over of every route $\mathbf{x}(s)$, starting from x at time $s = t$ to $s = T$ by the value function:

$$\psi(t, x) = \inf_{\mathbf{x}(s), \mathbf{x}(t)=x} C(\{\mathbf{x}(s), T\}), \tag{4.4}$$

where

$$C(\{\mathbf{x}(s), T\}) = \int_t^T \left[exp(-H(s, \mathbf{x}(s))) + \frac{\lambda}{2}|\mathbf{x}'(s)|^2 \right] ds,$$

with a fixed time horizon $T > 0$. The cost functional consists of the evaluation of environment and the kinetic energy, which penalizes high speed. For completeness, the derivation from (4.4) to (4.3) assuming $\psi \in C^1$ can be found in Appendix B.

As our setting does not carry a specific finite time horizon, we take time horizon T to infinity. This leads us from (4.3) to the eikonal equation:

$$\frac{|\nabla_x \psi|^2}{2\lambda} = exp(-H(t, x)). \tag{4.5}$$

We assume that the individual would choose to move along the path that minimizes the cost function. The velocity of our choice is therefore the gradient of ψ:

$$v = -\lambda^{-1} \nabla_x \psi,$$

where ψ solves (4.5)

4.3 Migration behaviors under periodic environments

We shall now simulate our general model under a simple periodic environment with two types of evaluation functions H. Our goal is to see which settings allow the animal to successfully follow resources and generate a periodic migration pattern.

4.3.1 Simple time-periodic environment

We assume there are two potential habitats, modeled by two disjoint circular regions A and B, see Figure 1. The location with positive value of E (*good resources area*) is alternating between A and B with a fixed duration T. E is assumed uniformly negative (*poor resources area*) outside the single good region. For our interest in memory effect, we also assume A and B are far enough from each other so that the animals cannot see both of them in the same time. That is, we assume

$$d(A, B) > \sup \{|x - y| : x, y \in supp(K)\}, \tag{4.6}$$

where K is the perceptual kernel.

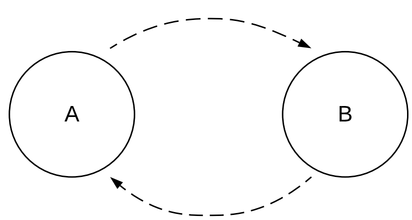

Figure 4.1: Time-periodic Setting: The location of the good resources alternates between A and B with duration T.

4.3.2 Two simple memory models

4.3.2.1 Memory model I: One single memory channel.

We first consider a memory system with only one memory channel M, with H in the following form:

$$H = KE + M, \tag{4.7}$$

where K is a perceptual kernel. This form means when a place x is close to where the individual stands, the evaluation mainly depends on what it sees. For distant places, it depends mainly on memory.

To encourage the first migration from A to B, we initiated the memory function M with positive values in both A and B, and zero otherwise. We also set the decay rate d small and update rate u large. While memory model I appears reasonable, it cannot produce a periodic migration pattern under the simple periodic environment as one expected, see Figure 4.2.

After the individual's first return for A, the value of M was updated negative in both A and B. The individual thus explores the other places that haven't been visited before, instead of returning to A or B.

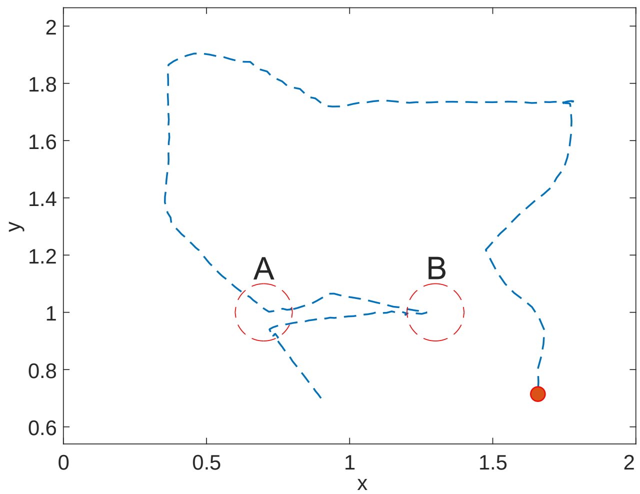

Figure 4.2: Trajectory for Memory Model I. The blue dashed line represents the trajectory and the red dot is the location of the individual at the end of the experiment. In this case, the individual was not able to repeat the migratory process because its memory structure was mismatched to the dynamics of the resource landscape.

4.3.2.2 Memory model II: Long and short-term memory.

Because memory model I is too simple to produce a periodic movement in a periodic environment, we increase the complexity and introduce the concept of short-term memory.

Memory model II contains two memory channels, including the long-term memory $M_\ell(t, x)$ and short-term memory $M_s(t, x)$. We assume M_s has larger decay and update rates than M_ℓ, so that it takes longer time to update and forget for information in M_ℓ, while M_s responds to changes quickly, and fades easily.

In this model, we define H as:

$$H(t, x) = M_s + M_\ell. \tag{4.8}$$

With (4.8), the individual makes a decision depending more on its local environment when it is in an extreme condition. Otherwise, it tends to rely more on the long-term memory. The following experiment shows a successful result.

Under the same simple periodic environment, memory model II successfully produced the desired migration patterns, see Figure 4.3. Observe that the individual will leave an exhausted region after a bit of explorations because of M_s, and return to A or B according to M_ℓ.

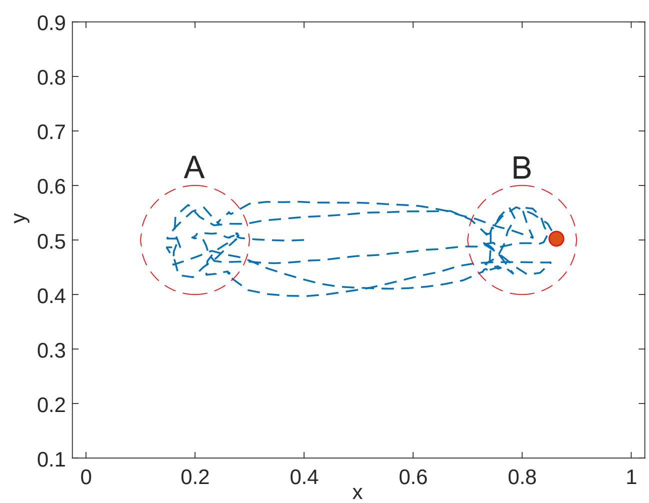

Figure 4.3: Trajectory for Memory Model II. In this case, a periodic migration pattern is successfully produced.

4.4 Discussion

4.4.1 Remarks on model components

Our main model components include the position, memory, and fitness. The eikonal equation is also important as a policy that utilizes the information in memory to make travel decisions. To investigate memory effects on animal migrations, the position and memory are indispensable in our model.

The fitness P has two roles in this work. It not only provides an universal measurement for different experiments, but also becomes an index to indicate when an animal would have the desire to move. Recall that we assume an animal would not want to move when the value of P is large.

Note that one could easily increase the complexity of memory and fitness models, by adding more assumptions or even introducing more functions to describe them. Here we intended to keep our model as simple as possible, and only considered essential features.

The eikonal equation, on the other hand, can be replaced by any other reasonable policy. Even though it is not the only option, the existence of efficient algorithms for the eikonal equation accelerates the numerical simulations. This advantage makes the eikonal equation a practical choice for us here.

4.4.2 Comparison between memory model I and II: Time scales of memory channels

Both Memory Model I and II have two memory channels, but with different time scales. In fact, the perception in Memory Model I can be seen as a channel with its scale close to zero. This observation combining with the experimental outcomes in Section 3.2, shows that the time scales play a decisive role on whether periodic dynamics can be produced. We shall demonstrate the relation between time scales and dynamic patterns with simulation results.

Consider $0 < \tau_2 < \tau_1$, where τ_1 is the time scale for the long-term memory, and τ_2 for the short-term. We again perform experiments with the same environment introduced in Section 3.1, and initiate M_ℓ with positive value in both A and B. Figure 4.4 shows the simulation results for time scales in different orders.

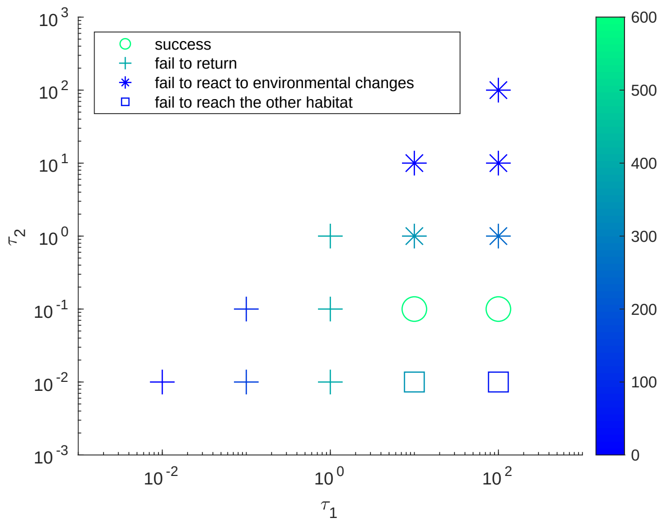

Figure 4.4: This graph shows the outcomes for several combinations of time scales. The different symbols correspond to different qualitative outcomes, whereas the color bar on the right hand side indicates the time step at which the periodic dynamic breaks.

We see from Figure 4.4 that there are roughly three different issues that could prevent us from having periodic dynamics:

1. ***The individual does not return to habitats if τ_1 is not large enough.***

   The larger τ_1 is, the more enduring the long-term memory would be. With a rather small τ_1, the individual would lose a positive long-term memory of both habitats A and B relatively quickly. Hence the individual ends up wandering around, instead of returning to A or B directly. See Figure 4.5 as an example.

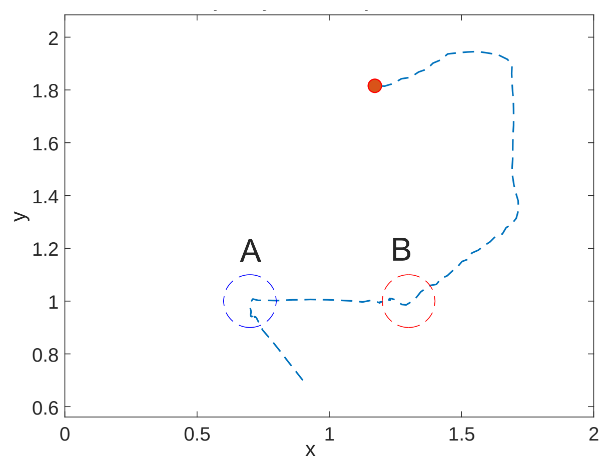

Figure 4.5: Trajectory when $\tau_1 \sim \tau_2 \sim 0.1$. In this case the animal does not return to A directly after visiting B.

2. ***The individual does not leave an exhausted habitat, if τ_2 is not small enough.***

   The smaller τ_2 is, the faster the short-term memory is updated. If the short-term memory is not updated fast enough, the individual cannot respond to the environmental change rapidly. Therefore, in this case the individual never leaves its current habitat; see for example Figure 4.6.

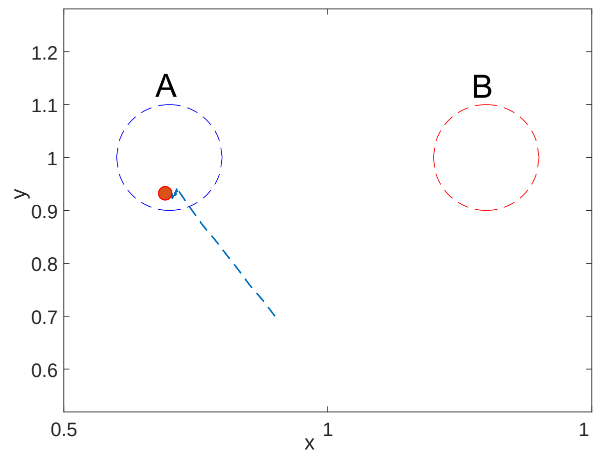

Figure 4.6: Trajectory when $\tau_1 \sim \tau_2 \sim 10$. The individual never leaves A because of the large τ_2.

3. ***The individual could have an early return and never reach the other habitat, when τ_2 is too small while τ_1 is large.***

   We mentioned in the second case that τ_2 needs to be small enough for the fast update of short-term memory, but there is also a lower bound for τ_2. The purpose of this lower bound is to make sure that the short-term memory has a high enough strength, so that the individual remembers the previous habitat is exhausted at least until it moves past the middle point of A and B. Otherwise, an early return could happen and the migration would not be successful.

   See Figure 4.7 as an example. The individual starts to leave A when the resources in A become exhausted, but the individual forgets that A lacks of resources before it moves past the middle point of its journey. Because the individual has a positive long-term memory of both habitats and it is closer to A, the individual chooses to return before reaching B.

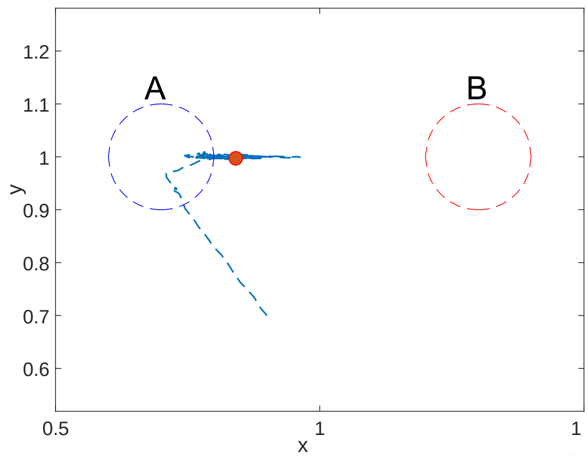

Figure 4.7: Trajectory when $\tau_1 \sim 100$, $\tau_2 \sim 0.01$. Because of the small τ_2 the information of M_s has been lost before the individual moves past the middle point of A and B. Hence the return to A happens early and the individual never reaches B.

4.5 Examples of further experiments

Beyond the simple time-periodic environment introduced in Section 3.1, several different environments could also be tested for further experiments. We give examples in the following:

- **Three Habitats.** We could test our model under a time-periodic environment with three habitats A, B and C. For one example of results, see Figure 4.8.

- **Changing habitats.** In this example we again have two habitats A and

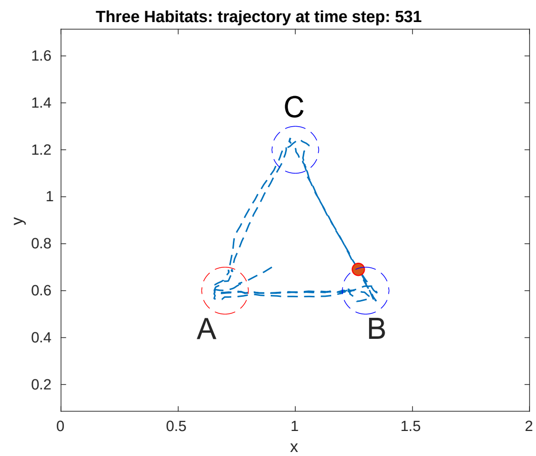

Figure 4.8: Here we have periodic dynamics, but the migration dynamics and environmental change do **not** have the same period. This can be observed from the movie: https://umd.box.com/s/01nvkkmfbn76y4fnn8kw0unwueac7ncs

B, but with A shrinking and B growing. At the end of experiments, A disappears entirely. See Figure 4.9.

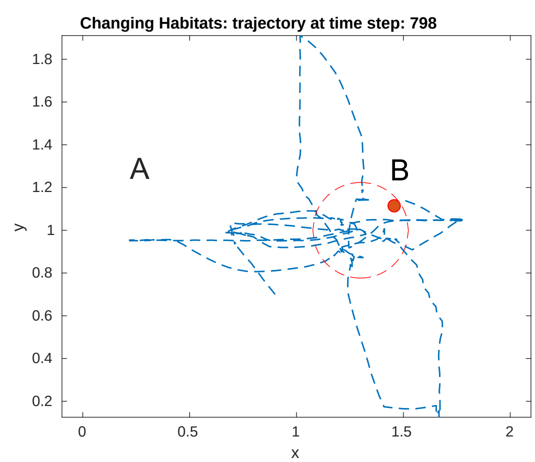

Figure 4.9: This is a result with an relatively small τ_1. The trajectory shows that the individual does not return to the area of A after the disappearance of that habitat. The movie can be found in: https://umd.box.com/s/ejkucrm3ghnxr2loazu65k5lcma99bsk

- **Intermediate Habitats/Dangerous Locations.** Here we add two bad areas in the middle of A and B, and two intermediate habitats above the bad regions.

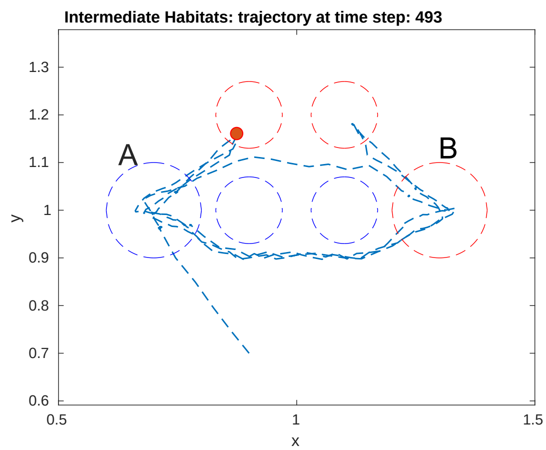

Figure 4.10: We see from the trajectory that the individual always avoids the bad areas, and sometimes chooses to reside in the intermediate habitats over A and B. Those intermediate habitats are good enough to retain the migrants. This type of phenomena has been observed in nature, see for example the article [100]. The movie of simulation can be found in: https://umd.box.com/s/z47d0nqw30t3m2azvb6ey7h8pmd8sg9m

- **Two Habitats with Random Seasonal Changes.** In this example we again assume the good habitat alternates between A and B. But different from before, each duration that resources stay in A or B is a random variable. The random variable is positive and uniformly distributed with the mean T and variance σ^2.

  We say the individual succeeds one journey, if it reaches one habitat from the other before the destination becomes exhausted. If the location of good resources changes n times across an experiment, it is considered there are totally n possible journeys.

In our experiments the time scales of long and short-term memory are fixed, and T is chosen such that the individual can succeed all possible journeys when the environment is time-periodic with T as the fixed duration for both habitats.

Recall from Section 4.2, we showed there is only a small region of appropriate time scales that the individual can successfully produce periodic dynamics under a time-periodic environment. When the duration of resources changes, the appropriate time scales change accordingly. The appropriate time scales should be smaller for a shorter duration, while larger when the duration is longer.

Now the duration of resources staying in one habitat is random each time, the appropriate time scales for each possible journey can be different. Every time our prior fixed time scales locate outside of the appropriate region corresponding to a certain duration in the experiment, the corresponding possible journey fails. For instance, if one of the duration of resources is really short, the individual could miss the corresponding possible journey because the fixed τ_2 is not small enough to respond to the fast environmental change, (which is the second case discussed in Section 4.2).

Therefore, the larger the variance in environmental duration is, the more likely that migratory journeys will fail. To visualize this tendency, we ran a series of experiments. All experiments have a total of 10 possible journeys and the time scales are fixed ($\tau_1 \sim 10$, $\tau_2 \sim 0.1$). T is set at 60 time steps, such that all 10 possible journeys are successful when there is no variance. A small ($\sigma^2 \sim 10$), moderate ($\sigma^2 \sim 100$) and large variance ($\sigma^2 \sim 1000$) case are then considered. We run 10 trials for each case. There are on average 9.5 successful journeys for the small variance case, 7.7 for moderate variance, and only 3.3 for large variance; see Figure 4.11. Roughly speaking, it becomes harder for the individual to follow the resources when the variance in the resource duration is increased. From this point of view, **environmental persistence** is very important for a memory-based migration to have a periodic pattern.

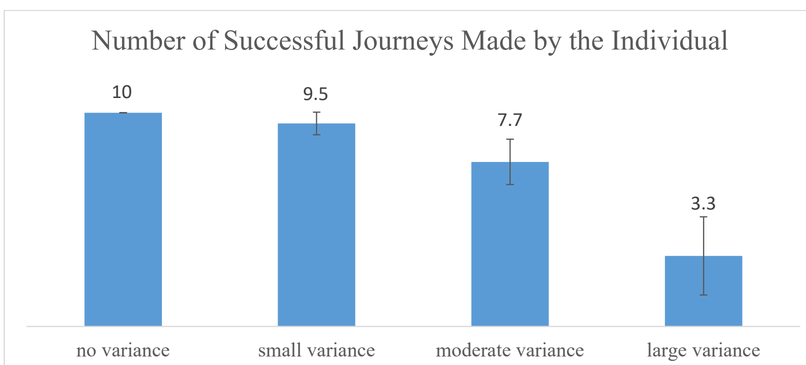

Figure 4.11: This chart indicates how many journeys on average the individual succeeds for the entire experiment time. A tendency is shown from this chart that the larger the variance in environmental duration is, the less successful journeys the individual has.

4.6 Conclusion

In this paper we develop a model for memory-based migrations of one individual over a broad range of memory mechanisms. Through numerical simulations under a simple time-periodic environment, periodic migration patterns are successfully recovered. Furthermore, we discover that in order to produce a periodic movement, the individual must be able to gather and carry enough information from both short and long-term memory, and capable of discriminating which information is more important with appropriate time scales.

While periodic movements can be recovered, the memory systems in our model do not include any intrinsic, a priori periodicity. The resulting periodic migration patterns are developed by the individual as its adaptation to periodic environmental changes.

Here we have considered the dynamics of one individual. For future research, it would be interesting to extend this memory-based model to a model for several individuals. Information sharing behaviors have been observed in many different species and shown beneficial for foraging efficiency [82]. In this future extension, the communications between individuals will be introduced, so that the interplay between information exchanges, individual memory, and group dynamics can be discussed.

Appendix A: Example for the non-degeneracy condition

We say $a(v) \in Lip(\mathbb{R})$ satisfies (2.8) with $\nu \in (0,1]$ on intervals if

$$|\{v : a(v) \in I\}| \leq C|I|^{\nu}, \quad \text{for all intervals } I, \tag{A.1}$$

And $a(v)$ satisfies the non-degeneracy condition on open sets with $\nu \in (0,1]$:

$$|\{v : a(v) \in \mathcal{O}\}| \leq C|\mathcal{O}|^{\nu}, \quad \text{for all open set } \mathcal{O}. \tag{A.2}$$

Here we give an example to show (A.1) cannot imply (A.2) with the same ν when $\nu = 1/2$. In fact the construction can be adapted to produce examples for all $\nu < 1$. Notice (A.1) and (A.2) are equivalent when $\nu = 1$.

Define $a : \left[0, \sum_{i=0}^{\infty} \frac{1}{3^i}\right] \to \left[0, \sum_{i=0}^{\infty} \frac{1}{3^{2i}}\right] \subset \mathbb{R}$ as follows:

$$\text{on } [0,1] = D_1, \quad a(v) = a_1(v) = 1 - (1-v)^2,$$

$$\text{on } \left[1, 1 + \frac{1}{3}\right] = D_2, \quad a(v) = a_2(v) = 1 + \frac{1}{3^2} a_1((v-1)3)$$

$$\vdots$$

The general formula is

$$a(v) = a_n(v) = \sum_{i=0}^{n-2} \frac{1}{3^{2i}} + \frac{1}{3^{2(n-1)}} a_1\left(\left(v - \sum_{i=0}^{n-2} \frac{1}{3^i}\right) 3^{n-1}\right) \quad \text{on } \left[\sum_{i=0}^{n-2} \frac{1}{3^i}, \sum_{i=0}^{n-1} \frac{1}{3^i}\right] = D_n.$$

We shall prove that a satisfies condition (A.1) with $\nu = 1/2$, but it fails (A.2) with the same ν.

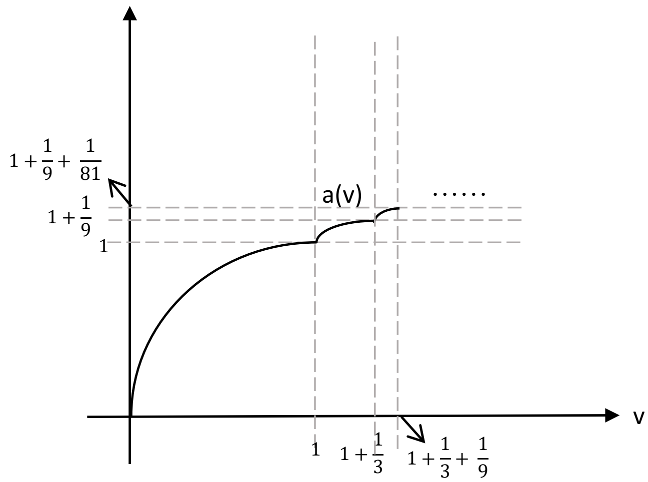

Figure A.1: graph of $a(v)$

Proposition 4. *There exists $C > 0$ such that for any interval I,*

$$|a^{-1}(I)| = |\{v : a(v) \in I\}| \leq C|I|^{1/2}. \tag{A.3}$$

Proof. Consider an interval $I = \left[\sum_{i=0}^{n-1} \frac{1}{3^{2i}} - p_2, \sum_{i=0}^{n-1} \frac{1}{3^{2i}} - p_1\right] = [c, d]$ inside some $a(D_n)$, where $0 \leq p_1 < p_2 \leq \frac{1}{3^{n-1}}$. So $|I| = p_2 - p_1$. Denote the pre-image of c and d by v_2 and v_1 respectively. Then we have for each $k = 1, 2$,

$$a_n(v_k) = \sum_{i=0}^{n-2} \frac{1}{3^{2i}} + \frac{1}{3^{2(n-1)}} a_1 \left((v_k - \sum_{i=0}^{n-2} \frac{1}{3^i}) 3^{n-1} \right) = \sum_{i=0}^{n-1} \frac{1}{3^{2i}} - p_k.$$

So

$$a_n^{-1} \left(\sum_{i=0}^{n-1} \frac{1}{3^{2i}} - p_1 \right) = v_k = \sum_{i=0}^{n-1} \frac{1}{3^i} - \sqrt{p_k}.$$

We therefore have

$$|a^{-1}(I)| = \sqrt{p_2} - \sqrt{p_1} \leq \sqrt{p_2 - p_1} = |I|^{1/2}.$$

If $I = [c, d] \subset a(\cup_{i=m_1}^{m_2} D_i)$, separate I into three sub-intervals: $I = I_1 \cup I_2 \cup I_3$, where $I_1 = \left[c, \sum_{i=0}^{m_1-1} \frac{1}{3^{2i}}\right]$, $I_2 = \left[\sum_{i=0}^{m_1-1} \frac{1}{3^{2i}}, \sum_{i=0}^{m_2-2} \frac{1}{3^{2i}}\right]$ and $I_3 = \left[\sum_{i=0}^{m_2-2} \frac{1}{3^{2i}}, d\right]$.

The above case applies to I_1 and I_3, so $|a^{-1}(I_1)| \le |I_1|^{1/2}$ and $|a^{-1}(I_3)| \le |I_3|^{1/2}$.

For I_2, we have

$$|I_2| = \sum_{m_1}^{m_2-2} \frac{1}{3^{2i}} = \frac{9}{8}\frac{1}{3^{2m_1}}\left[1 - \left(\frac{1}{9}\right)^{m_2-m_1-1}\right].$$

And

$$\begin{aligned} |a^{-1}(I_2)|^2 &= \left(\sum_{m_1}^{m_2-2} \frac{1}{3^i}\right)^2 = \frac{9}{4}\frac{1}{3^{2m_1}}\left[1 - \left(\frac{1}{3}\right)^{m_2-m_1-1}\right]^2 \\ &\le \frac{9}{4}\frac{1}{3^{2m_1}}\left[1 - 2\left(\frac{1}{9}\right)^{m_2-m_1-1} + \left(\frac{1}{9}\right)^{m_2-m_1-1}\right] \\ &= 2|I_2|. \end{aligned}$$

So

$$|a^{-1}(I_2)| \le 2^{1/2}|I_2|^{1/2}.$$

Notice that this inequality is still true when m_2 goes to infinity, so there are no issues near the right end point.

Combining the three inequalities we get

$$|a^{-1}(I)| = \sum_{i=1}^{3} |a^{-1}(I_i)| \le 2^{1/2}\sum_{i=1}^{3} |I_i|^{1/2} \le 6^{1/2}\left(\sum_{i=1}^{3} |I_i|\right)^{1/2} = 6^{1/2}|I|^{1/2}.$$

□

Proposition 5. *There exists a sequence of set $\mathcal{O}^m$ such that*

$$\frac{|a^{-1}(\mathcal{O}^m)|}{|\mathcal{O}^m|^{1/2}} \to \infty \qquad \text{as } m \to \infty.$$

Proof. Let

$$\mathcal{O}^m = \cup_{n=1}^{m} I_n,$$

where $I_n = \left[\sum_{i=0}^{n-1} \frac{1}{3^{2i}} - \frac{1}{3^{2(m-1)}}, \sum_{i=0}^{n-1} \frac{1}{3^{2i}}\right]$ for all $1 \leq n \leq m$.

So

$$|I_n| = \frac{1}{3^{2(m-1)}} \quad \text{for all } 1 \leq n \leq m,$$

and

$$|a^{-1}(I_n)| = |I_n|^{1/2} = \frac{1}{3^{m-1}} \quad \text{for all } 1 \leq n \leq m.$$

Therefore,

$$\frac{|a^{-1}(\mathcal{O}^m)|}{|\mathcal{O}^m|^{1/2}} = \frac{\frac{m}{3^{m-1}}}{\left(\frac{m}{3^{2(m-1)}}\right)^{1/2}} = \sqrt{m} \to \infty \qquad \text{as } m \to \infty.$$

□

Appendix B: Derivation of Hamilton-Jacobi-Bellman equation

Proposition 6. *Let*

$$\psi(t,x) = \inf_{\mathbf{x}(s),\mathbf{x}(t)=x} C(\{\mathbf{x}(s),T\}),$$

where

$$C(\{\mathbf{x}(s),T\}) = \int_t^T \left[exp(-H(s,\mathbf{x}(s))) + \frac{\lambda}{2}|\mathbf{x}'(s)|^2 \right] ds.$$

Assume the value function ψ is C^1 in (x,t), then ψ solves

$$\partial_t \psi = \sup_p \left\{ -p \cdot \nabla_x \psi - exp(-H) - \frac{\lambda}{2}|p|^2 \right\}.$$

Moreover, the optimal trajectory is the one starting with velocity $-\lambda^{-1}\nabla_x\psi$, in which case gives the Hamilton-Jacobian-Bellman equation:

$$\partial_t \psi = \frac{|\nabla_x \psi|^2}{2\lambda} - exp(-H(t,x)).$$

Proof. For a $h > 0$, for any vector p, we consider the line segment $\ell(s) = x+p(s-t)$ from $s = t$ to $s = t+h$. Connecting ℓ and any path $\mathbf{x}(s)$ from the point $x+ph$ at time $s = t+h$ to $s = T$, we get a path $\tilde{\mathbf{x}}$ starting from x at time $s = t$ and end at time $s = T$. By the definition of ψ,

$$\psi(t,x) \le \inf_{p,\tilde{\mathbf{x}}} C(\tilde{\mathbf{x}}) = \int_t^{t+h} \left[exp(-H(s,\ell(s))) + \frac{\lambda}{2}|p|^2 \right] ds + \psi(t+h, x+ph)$$

So

$$\frac{\psi(t+h,x+ph) - \psi(t,x)}{h} \ge -\frac{1}{h}\int_t^{t+h} \left[exp(-H(s,\ell(s))) + \frac{\lambda}{2}|p|^2 \right]$$

Taking $h \to 0$, we derive

$$\partial_t \psi \geq -p \cdot \nabla_x \psi - exp(-H(t,x)) - \frac{\lambda}{2}|p|^2.$$

This inequality holds for every p, hence

$$\inf_p \left\{ \partial_t \psi + p \cdot \nabla_x \psi + exp(-H(t,x)) + \frac{\lambda}{2}|p|^2 \right\} \geq 0.$$

In fact the equality holds as zero is realized when p is chosen as the velocity of the optimal trajectory at $s = t$. Let $\mathbf{x}_{op}$ be the optimal trajectory from $s = t$ to $t + h$, then

$$\psi(t,x) = \int_t^{t+h} \left[exp(-H(s, \mathbf{x}_{op}(s))) + \frac{\lambda}{2}|\mathbf{x}'_{op}|^2 \right] ds + \psi(t+h, \mathbf{x}_{op}(t+h)).$$

So

$$\frac{\psi(t+h, \mathbf{x}_{op}(t+h)) - \psi(t,x)}{h} = -\frac{1}{h} \int_t^{t+h} \left[exp(-H(s, \mathbf{x}_{op}(s))) + \frac{\lambda}{2}|\mathbf{x}'_{op}|^2 \right] ds$$

Taking $h \to 0$,

$$\partial_t \psi(t,x) = -\mathbf{x}'_{op} \cdot \nabla_x \psi - exp(-H(t,x)) - \frac{\lambda}{2}|\mathbf{x}'_{op}|^2,$$

for some vector $\mathbf{x}'_{op}$.

We rewrite

$$\inf_p \left\{ \partial_t \psi + p \cdot \nabla_x \psi + exp(-H(t,x)) + \frac{\lambda}{2}|p|^2 \right\} = 0$$

to

$$\partial_t \psi = \sup_p \left\{ -p \cdot \nabla_x \psi - exp(-H(t,x)) - \frac{\lambda}{2}|p|^2 \right\}.$$

Notice inside the parentheses is a quadratic form in p,

$$\begin{aligned}
\partial_t \psi &= \sup_p \left\{ -p \cdot \nabla_x \psi - exp(-H(t,x)) - \frac{\lambda}{2}|p|^2 \right\} \\
&= \sup_p \left\{ -\frac{\lambda}{2} \left(p + \frac{\nabla_x \psi}{\lambda} \right)^2 + \frac{|\nabla_x \psi|^2}{2\lambda} - exp(-H(t,x)) \right\} \\
&= \frac{|\nabla_x \psi|^2}{2\lambda} - exp(-H(t,x)),
\end{aligned}$$

which is realized when $p = -\lambda^{-1} \nabla_x \psi$. $\square$

Appendix C: Implementation for our model

Consider $\Omega = [0,1] \times [0,1]$ and the final time $T > 0$. We discretize $[0,T] \times \Omega$ uniformly for $N_t \times N^2$ increments.

1. Update fitness P with implicit scheme if E is positive, explicit if negative:

$$P(t_{k+1}) = \begin{cases} \frac{P(t_k)+E(t_k,X(t_k),Y(t_k))\overline{P}\Delta t}{1+E(t_k,X(t_k),Y(t_k))\Delta t} \\ \\ P(t_k) + E(t_k, X(t_k), Y(t_k))(\overline{P} - P(t_k))\Delta t. \end{cases}$$

   Stop if $P(t_k) \leq \underline{P}$.

2. Update memory M_ℓ, M_s:

$$M_\ell(t_{k+1}, x_i, y_j) = M_\ell(t_k, x_i, y_j) + \\ \Delta t\, \tau \left\{ -d\sqrt{|M_\ell(t_k, x_i, y_j)|}|M_\ell(t_k, x_i, y_j)| + Perception \right\}, \quad \text{(C.1)}$$

   where $Perception = u\, K(X(t_k), Y(t_k), x_i, y_j)[E(t_k, x_i, y_j) - M_\ell(t_k, x_i, y_j)]$.

   If M_ℓ changes sign, put it zero. Same formula for M_s.

3. Update evaluation H:

$$H[t_{k+1}, x_i, y_j] = \frac{|M_s[t_k, x_i, y_j]|}{(1 + |M_s[t_k, x_i, y_j]|)} M_s[t_k, x_i, y_j] + \\ \frac{1}{(1 + |M_s[t_k, x_i, y_j]|)} M_\ell[t_k, x_i, y_j]. \quad \text{(C.2)}$$

4. Solve eikonal equation $|\nabla_x \psi^b| = \sqrt{2\lambda exp(-H)}$ with boundaries $\{x : H(x) = b\}$ for $\inf_\Omega(H) \leq b \leq \sup_\Omega(H)$. Then take $\psi = \psi^{b_0}$, where $b_0 = \min\left\{b : \psi^b(X,Y) + exp(-b)\right\}$.

5. Update Position (X,Y):

$$\Delta X = \begin{cases} \left[\psi(t_k, x_{i+1}, y_j) - \psi(t_k, x_{i-1}, y_j)\right]/(2\Delta x) \text{ if } \frac{1}{N} < X < 1 - \frac{1}{N} \\ -\psi(t_k, x_1, y_j)/(2\Delta x) \text{ if } 0 \leq X \leq \frac{1}{N} \\ \psi(t_k, x_{N-1}, y_j)/(2\Delta x) \text{ if } \frac{1}{N} \leq X \leq 1. \end{cases}$$

Similar for ΔY. Let (r_x, r_y) be a random variable with standard normal distribution in 2 dimension.

$$\begin{cases} X(t_{k+1}) = X(t_k) - \chi(P, (E(t, X(t_k), Y(t_k))))\Delta X \Delta t + \sigma r_x \sqrt{\Delta t} \\ Y(t_{k+1}) = Y(t_k) - \chi(P, (E(t, X(t_k), Y(t_k))))\Delta Y \Delta t + \sigma r_y \sqrt{\Delta t}. \end{cases}$$

Repeat the procedure until $t = T$.

We make two remarks about the implementation. First, as $sgn(M)\sqrt{|M|}$ is not smooth and the finite difference method is unstable near zero, whenever M changes sign, we put it as zero for the new step.

Second, to contain the experiment inside a bounded domain Ω for all time, we solve the eikonal equation for multiple level sets inside Ω, instead of $\partial\Omega$, and choose the path with the smallest cost. Otherwise the dynamics will eventually escape as the cost of moving around forever will eventually become larger than one fixed exit cost.

Bibliography

[1] B. Abrahms, E. L. Hazen, E. O. Aikens, M. S. Savoca, J. A. Goldbogen, S. J. Bograd, M. G. Jacox, L. M. Irvine, D. M. Palacios, B. R. Mate, *Memory and resource tracking drive blue whale migrations*, Proceedings of the National Academy of Sciences, 116(12), (2019), 5582-5587.

[2] V.I. Agoshkov, *Spaces of functions with differential-difference characteristics and the smoothness of solutions of the transport equation*, Dokl. Akad. Nauk SSSR, 276(6), (1984), 1289-1293.

[3] G. Alberti, S. Bianchini, G. Crippa, *A uniqueness result for the continuity equation in two dimensions*, Journal of the European Mathematical Society 16, no. 2 (2014), 201-234.

[4] G. Alberti, G. Crippa, A. L. Mazzucato, *Exponential self-similar mixing and loss of regularity for continuity equations*, Comptes rendus mathematique 352, no. 11 (2014), 901-906.

[5] G. Alberti, G. Crippa, and A. Mazzucato, *Exponential self-similar mixing by incompressible flows*, Journal of the American Mathematical Society, 32(2), (2019), 445-490.

[6] L. Ambrosio, *Transport equation and Cauchy problem for BV vector fields*, Invent. Math., 158 (2004), 227-260.

[7] D. Amadori, P. Goatin, M. D. Rosini, *Existence results for Hughes' model for pedestrian flows*, Journal of Mathematical Analysis and applications 420, no. 1, (2014), 387-406.

[8] A.A. Arsenev, *Global existence of a weak solution of Vlasov's system of equations*, U.S.S.R. Comp. Math. and Math. Phys. 15, (1975), 131-141.

[9] D. Arsénio, L. Saint-Raymond, *Compactness in kinetic transport equations and hypoellipticity*, J. Funct. Anal., 261(10), (2011), 3044-3098.

[10] D. Arsénio, N. Masmoudi, *A new approach to velocity averaging lemmas in Besov spaces*, J. Math. Pures Appl.(9) 101, no. 4, (2014), 495-551.

[11] D. Arsénio, N. Masmoudi, *Maximal gain of regularity in velocity averaging lemmas, Analysis and PDE*, Vol. 12, no. 2, (2019), 333-388.

[12] E. Audusse, B. Perthame, *Uniqueness for scalar conservation laws with discontinuous flux via adapted entropies*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics 135.2 (2005), 253-265.

[13] M. Bardi, I. Calpuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations* Birkhauser, Cambridge, MA, (1997).

[14] H. L. Bartlam-Brooks, P. S. Beck, G. Bohrer, S. Harris, *In search of greener pastures: Using satellite images to predict the effects of environmental change on zebra migration*, Journal of Geophysical Research: Biogeosciences 118, no. 4, (2013), 1427-1437.

[15] J. Bedrossian, N. Masmoudi, C. Mouhot. *Landau damping in finite regularity for unconfined systems with screened interactions*, Communications on Pure and Applied Mathematics 71, no. 3 (2018), 537-576.

[16] D. A. Bennett, W. Tang. *Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone*, International Journal of Geographical Information Science 20, no. 9 (2006), 1039-1066.

[17] J. M. Berbert, W. F. Fagan, *How the interplay between individual spatial memory and landscape persistence can generate population distribution patterns*, Ecological Complexity, 12, (2012), 1-12.

[18] M. Bézard, *Régularité L^p précisée des moyennes dans les équations de transport*, Bull. Soc. Math. France, 122(1) (1994), 29-76.

[19] F. Bouchut, L. Desvillettes, *Averaging lemmas without time Fourier transform and application to discretized kinetic equations*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics 129, no. 1 (1999), 19-36.

[20] F. Bouchut, L. Desvillettes, *On two-dimensional Hamiltonian transport equations with continuous coefficients*, Differential and Integral Equations 14, no. 8 (2001), 1015-1024.

[21] F. Bouchut, *Hypoelliptic regularity in kinetic equations*, J. Math. Pures Appl. (9) 81 (11) (2002), 1135–1159.

[22] C. Bracis, T. Mueller, *Memory, not just perception, plays an important role in terrestrial mammalian migration*, Proc. R. Soc. B 284 (2017), 20170449.

[23] A. Bressan, *A lemma and a conjecture on the cost of rearrangements*, Rendiconti del Seminario Matematico della Universita di Padova 110 (2003), 97-102.

[24] J. Bricmont, A. Kupiainen, R. Lefevere, *Exponential mixing of the 2D stochastic Navier-Stokes dynamics*, Communications in mathematical physics, 230(1), (2002), 87-132.

[25] M. Burger, Y. Dolak-Struss, C. Schmeiser, *Asymptotic analysis of an advection-dominated chemotaxis model in multiple spatial dimensions*, Commun. Math. Sci. 6, no. 1, (2006), 1-28.

[26] R. S. Cantrell, C. Cosner, and Y. Lou, *Evolution of dispersal and the ideal free distribution*, Mathematical Biosciences and Engineering 7 (2010), 17-36.

[27] E. Cartee, A. Vladimirsky, *Anisotropic Challenges in Pedestrian Flow Modeling*, Communications in Mathematical Sciences 16/4 (2018), 1067-1093.

[28] A. Chacon, A. Vladimirsky, *Fast two-scale methods for eikonal equations*, SIAM Journal on Scientific Computing 34, no. 2 (2012), A547-A578.

[29] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš, *Diffusion and mixing in fluid flow*, Ann. of Math. (2), 168(2), (2008), 643-674.

[30] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, T. Tao, *Viriel, Morawetz, and interaction Morawetz inequalities.*

[31] G. Crippa, C. De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math., 616, (2008), 15-46.

[32] C. Dafermos, *Hyperbolic conservation laws in continuum physics*, Grundlehren der Mathematischen Wissenschaften, 325, Springer-Verlag, Berlin, (2000).

[33] A. Debussche, J. Vovelle, *Scalar conservation laws with stochastic forcing*, J. Funct. Anal. 259, (2010), 1014-1042.

[34] C. De Lellis, M. Westdickenberg, *On the optimality of velocity averaging lemmas*, Ann. Inst. H. Poincaré Anal. Non Linéaire 20, no. 6, (2003), 1075-1085.

[35] R. DeVore, G. Petrova, *The averaging lemma*, J. Amer. Math. Soc. 14 (2001), no. 2, 279-296.

[36] R.J. DiPerna, *Measure-valued solutions to conservation laws*, Arch. Rational Mech. Anal., 88 (1985), 223-270.

[37] R.J. DiPerna, P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. 98 (1989), 511-547.

[38] R.J. DiPerna, P.-L. Lions, *Global weak solutions of Vlasov-Maxwell systems*, Comm. Pure Appl. Math. 42 (1989), 729-757.

[39] R. J. DiPerna, P.-L. Lions, *On the Cauchy problem for Boltzmann equations: global existence and weak stability*, Ann. of Math. .2/ 130:2 (1989), 321-366.

[40] R. J. DiPerna, P.-L. Lions, Y. Meyer, *L^p regularity of velocity averages*, Ann. Inst. H. Poincaré Anal. Non Linéaire 8 (1991), no. 3-4, 271-287.

[41] S. Doi, *On the Cauchy problem for Schrödinger type equations and the regularity of the solutions*, J. Math. Kyoto Univ. 34, no. 2, (1994), 319-328.

[42] M. Escobedo, S. Mischler, M. Valle, *Homogeneous Boltzmann equation in quantum relativistic kinetic theory*, Vol. 4 of Electronic Journal of Differential Equations, Monograph, Southwest Texas State University, San Marcos, TX, (2003).

[43] W. F. Fagan, M. A. Lewis, M. Auger-Méthé, T. Avgar, S. Benhamou, G. Breed, L. LaDage, U. E. Schlägel, W.-W. Tang, Y.P. Papastamatiou, J. Forester, *Spatial memory and animal movement*, Ecology letters 16, no. 10 (2013), 1316-1329.

[44] W. F. Fagan, E. Gurarie, S. Bewick, A. Howard, R. S. Cantrell, C. Cosner, *Perceptual ranges, information gathering, and foraging success in dynamic landscapes*, The American Naturalist, 189(5), (2017), 474-489.

[45] W. F. Fagan, *Migrating whales depend on memory to exploit reliable resources*, Proceedings of the National Academy of Sciences 116.12 (2019), 5217-5219.

[46] U.S. Fjordholm, S. Ulrik, R. Käppeli, S. Mishra, E. Tadmor, *Construction of approximate entropy measure-valued solutions for hyperbolic systems of conservation laws*, Found. Comput. Math. 17, no. 3, (2017), 763-827.

[47] U.S. Fjordholm, S. Mishra, E. Tadmor, *On the computation of measure-valued solutions*, Acta Numer. 25,(2016), 567-679.

[48] D. Fortin, H. L. Beyer, M. S. Boyce, D. W. Smith, T. Duchesne, J. S. Mao, *Wolves influence elk movements: behavior shapes a trophic cascade in Yellowstone National Park*, Ecology 86, no. 5 (2005), 1320-1330.

[49] M. Di Francesco, P. A. Markowich, J.-F. Pietschmann, M.-T. Wolfram, *On the Hughes' model for pedestrian flow: The one-dimensional case*, Journal of Differential Equations 250, no. 3 (2011): 1334-1362.

[50] I. Gasser, P. Markowich, B. Perthame, *Dispersion and Moments Lemma revisited*, J. Differential Equations 156 (1999), 254-281.

[51] F. Golse, P.-L. Lions, B. Perthame, R. Sentis, *Regularity of the moments of the solution of a transport equation*, J. Funct. Anal. 76, no. 1, (1988), 110-125.

[52] F. Golse, F. Poupaud, *Limite fluide des équations de Boltzmann des semiconducteurs pour une statistique de Fermi-Dirac*, Asymptotic Anal. 6 (1992), 135-160.

[53] F. Golse, L. Saint-Raymond, *Hydrodynamic limits for the Boltzmann equation*, Riv. Mat. Univ. Parma 4 (2005), 1-144.

[54] F. Golse, L. Saint-Raymond, *The Navier-Stokes limit of the Boltzmann equation for hard potentials*, J. Math. Pures Appl. 91 (2009), 508-552.

[55] F. Golse, B. Perthame, *Optimal regularizing effect for scalar conservation laws*, Rev. Mat. Iberoam., 29, no. 4, (2013), 1477-1504.

[56] F. Golse, L. Saint-Raymond, *Velocity averaging in L^1 for the transport equation*, C. R. Acad. Sci. Paris Ser. I Math, 334 (2002), 557-562.

[57] F. Golse, *Fluid dynamic limits of the kinetic theory of gases, From particle systems to partial differential equations*, Springer, Berlin, Heidelberg, (2014), 3-91.

[58] H. Grad, *On the kinetic theory of rare gases*, Comm. Pure Appl. Math. 2 (1949), 331-407.

[59] D. Han-Kwan, *L^1 averaging lemma for transport equations with Lipschitz force fields*, Kinet. Relat. Models 3 (4), (2010), 669-683.

[60] M. Hauray, *On two-dimensional hamiltonian transport equations with L^p_{loc} coefficients*, In Annales de l'IHP Analyse non linéaire, vol. 20, no. 4, (2003), 625-644.

[61] L. Hörmander, *Hypoelliptic second order differential equations*, Acta Mathematica 119.1 (1967), 147-171.

[62] R. L. Hughes, *A continuum theory for the flow of pedestrians*, Transportation Research Part B: Methodological 36(6), (2002), 507-535.

[63] R. L. Hughes, *The flow of human crowds*, Annu. Rev. Fluid Mech. 35, (2003), 169-182.

[64] G. Iyer, A. Kiselev, X. Xu, *Lower bounds on the mix norm of passive scalars advected by incompressible enstrophy-constrained flows*, Nonlinearity 27, no. 5 (2014), 973.

[65] P.-E. Jabin, B. Perthame, *Compactness in Ginzburg-Landau energy by kinetic averaging*, Comm. Pure Appl. Math. 54 (9) (2001), 1096-1109.

[66] P.-E. Jabin, B. Perthame, *Regularity in kinetic formulations via averaging lemmas, A tribute to J. L. Lions*, ESAIM Control Optim. Calc. Var. 8 (2002), 761-774.

[67] P.-E. Jabin, L. Vega, *Averaging lemmas and the X-ray transform*, C. R. Math. Acad. Sci. Paris, 337(8) (2003), 505-510.

[68] P.-E. Jabin, L. Vega, *A real space method for averaging lemmas*, J. Math. Pures Appl. (9) 83 (2004), no. 11, 1309-1351.

[69] B. R. Jesmer, J. A. Merkle, J. R. Goheen, E. O. Aikens, J. L. Beck, A. B. Courtemanch, M. A. Hurley, D. E. McWhirter, H. M. Miyasaki, K. L. Monteith, M. J. Kauffman, *Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals*, Science 361, no. 6406 (2018), 1023-1025.

[70] J.-P. Kahane, *Some random series of functions*, Cambridge University Press, (1993).

[71] S. N. Kruzkov, *First order quasilinear equations in several independent variables*, Math USSR Sb., no. 10, (1970), 217-243.

[72] K. Kunihiko, *The Cauchy problem for Schrödinger type equations with variable coefficients*, J. Math. Soc. Japan 50 (1998), no. 1, 179-202.

[73] K.-Y. Lam, Y. Lou, *Evolution of conditional dispersal: Evolutionarily stable strategies in spatial models*, Journal of Mathematical Biology 6 (2014), 851-877.

[74] F. Léger, *A new approach to bounds on mixing* Mathematical Models and Methods in Applied Sciences 28, no. 05 (2018), 829-849.

[75] Z. Lin, J.-L. Thiffeault, C. R. Doering, *Optimal stirring strategies for passive scalar mixing*, J. Fluid Mech., 675 (2011), 465-476.

[76] P.-L. Lions, B. Perthame, *Lemmes de moments, de moyenne et de dispersion*, C. R. Acad. Sc., Série I, 314, (1992), 801-806.

[77] P.-L. Lions, B. Perthame, E. Tadmor, *A kinetic formulation of multidimensional scalar conservation laws and related equations*, J. Amer. Math. Soc. 7 (1994), no. 1, 169-191.

[78] P.-L. Lions, B. Perthame, E. Tadmor, *Kinetic formulation of the isentropic gas dynamics and p-systems*, Comm. Math. Phys., 163 (1994), 415-431.

[79] P.-L. Lions, B. Perthame, P. E. Souganidis, *Stochastic averaging lemmas for kinetic equations*, Séminaire Laurent Schwartz—EDP et applications (2013), 1-17.

[80] P.-L. Lions, B. Perthame, and P. E. Souganidis. *Scalar conservation laws with rough (stochastic) fluxes: the spatially dependent case. Stochastic Partial Differential Equations*, Analysis and Computations 2.4 (2014), 517-538.

[81] E. Lunasin, Z. Lin, A. Novikov, A. Mazzucato, and C. R. Doering, *Optimal mixing and optimal stirring for fixed energy, fixed power, or fixed palenstrophy flows*, Journal of Mathematical Physics 53, no. 11 (2012), 115611.

[82] R. Martínez-García, J. M. Calabrese, T. Mueller, K. A. Olson, C. López, *Optimizing the search for resources by sharing information: Mongolian gazelles as a case study*, Physical Review Letters, 110(24), (2013), 248106.

[83] N. Masmoudi, M. L. Tayeb, *Diffusion limit of a semiconductor Boltzmann-Poisson system*, SIAM Journal on Mathematical Analysis 38.6 (2007), 1788-1807.

[84] G. Mathew, I. Mezić, and L. Petzold, *A multiscale measure for mixing*, Physica, D 211, (2005), 23-46.

[85] P. Moorcroft, M. A. Lewis, *Mechanistic home range analysis*, Princeton University Press, (2006).

[86] C. Mouhot, C. Villani, *On Landau damping*, Acta Math., 207, (2011), 29-201.

[87] A. Okubo, *Diffusion and ecological problems: mathematical models*, Springer-Verlag, Berlin, Heidelberg et New York, (1980).

[88] O. A. Oleinik, *Discontinuous solutions of nonlinear differential equations*, Transl. Amer. Math. Soc. Ser., 2 26, (1957) 95-172.

[89] R. E. A. C. Paley, N. Wiener, and A. Zygmund, *Notes on random functions*, Mathematische Zeitschrift, 37(1) (1933), 647-668.

[90] E. Y. Panov *On existence and uniqueness of entropy solutions to the Cauchy problem for a conservation law with discontinuous flux*, Journal of Hyperbolic Differential Equations 6.03 (2009), 525-548.

[91] B. Perthame, *Higher moments for kinetic equations: the Vlasov-Poisson and Fokker-Planck cases*, Math. Methods in the Applied Sc., 13, (1990), 441-452.

[92] B. Perthame, *Kinetic formulation of conservation laws. Oxford Lecture Series in Mathematics and Its Applications*, 21. Oxford University Press, Oxford, (2002).

[93] L. Saint-Raymond, *Hydrodynamic limits of the Boltzmann equation*, Springer Science and Business Media, no. 1971, (2009).

[94] U. E. Schlägel, M. A. Lewis, *Detecting effects of spatial memory and dynamic information on animal movement decisions*, Methods in Ecology and Evolution 5, no. 11 (2014), 1236-1246.

[95] C. Seis, *Maximal mixing by incompressible fluid flows*, Nonlinearity 26, no. 12 (2013), 3279.

[96] J.A. Sethian, *A Fast Marching Level Set Method for Monotonically Advancing Fronts*, Proc. Natl. Acad. Sci., 93, 4, (1996), 1591-1595.

[97] G. Staffilani, D. Tataru, *Strichartz estimates for a Schrödinger operator with nonsmooth coefficients*, Comm. Partial Differential Equations 27 (2002), no. 7-8, 1337-1372.

[98] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, no. 30. Princeton University Press, Princeton, N.J., (1970).

[99] E. Tadmor, T. Tao, *Velocity averaging, kinetic formulations, and regularizing effects in quasi-linear PDEs*, Comm. Pure Appl. Math. 60 (2007), no. 10, 1488-1521.

[100] C.S. Teitelbaum, S. J. Converse, W. F. Fagan, K. Böhning-Gaese, R. B. O'Hara, A. E. Lacy, and T. Mueller, *Experience drives innovation of new migration patterns of whooping cranes in response to global change*, Nature communications 7, no. 1 (2016), 1-7.

[101] C. M. Topaz, A. L. Bertozzi, *Swarming Patterns in a Two-Dimensional Kinematic Model for Biological Groups*, SIAM J. Appl. Math. 65 (2004), no. 1, 152-174.

[102] M. Twarogowska, P. Goatin, R. Duvigneau, *Macroscopic modeling and simulations of room evacuation*, Applied Mathematical Modelling 38, no. 24 (2014), 5781-5795.

[103] C. Villani, *A review of mathematical topics in collisional kinetic theory*, Handbook of mathematical fluid dynamics 1, no. 71-305, (2002), 3-8.

[104] C. Villani, *Landau damping*, Lecture notes, to appear in Panoramas et Synthèse, (2011).

[105] M. Westdickenberg, *Some new velocity averaging results*, SIAM J. Math. Anal. 33 (2002), no. 5, 1007-1032.

[106] Y. Xia, S. C. Wong, C.-W. Shu, *Dynamic continuum pedestrian flow model with memory effect*, Physical Review E 79.6 (2009), 066113.

[107] H. Zhao, *A fast sweeping method for eikonal equations*, Math. Comp., 74 (2004), 603-627.

[108] C. Zillinger, *On geometric and analytic mixing scales: comparability and convergence rates for transport problems*, (2018), preprint, arXiv:1804.11299.