RESEARCH STATEMENT

YUANYUAN FENG

My research interests lie broadly in applied mathematics including fluid dynamics, systems with memory, algorithms in machine learning, etc. My thesis advisor is Gautam Iyer and my thesis project [1] concerns mixing. We quantify the interaction between diffusion and mixing by studying the energy dissipation rate based on an assumption quantifying the mixing rate. Besides the project on mixing, I collaborated with Jian-Guo Liu and Lei Li and studied the Markov semigroups for two important algorithms from machine learning: stochastic gradient descent (SGD) and online principal component analysis (PCA) [2]. The third project I have worked on is time-fractional ODEs [3, 4]. Together with Lei Li, Jian-Guo Liu and Xiaoqian Xu, we study 1-D autonomous fractional ODEs $D_c^\gamma u = f(u), 0 < \gamma < 1$, where D_c^γ is the generalized Caputo derivative proposed by Li and Liu [5] based on a convolution group.

In the following statement, I will talk about the above projects in detail with discussion on the future plans at the end. In Section 1.1, I will introduce my work on mixing [1]. In Section 1.2, I will talk about the project on Markov semigroups for SGD and online PCA [2]. In Section 1.3, the work on fractional ODEs [3, 4] would be explained.

1. Research Summary

1.1. **Dissipation Enhancement by Mixing.** In incompressible fluids, stirring induces mixing by filamentation and facilitates the formation of small scales. Diffusion, on the other hand, efficiently damps small scales. Our aim is to quantify the interaction between diffusion and mixing by studying the energy dissipation rate based on an assumption quantifying the mixing rate. We study the passive scalar model

$$\begin{cases} \partial_t \theta + (u \cdot \nabla)\theta - \nu \Delta \theta = 0 \quad \text{in } \mathbb{T}^d, \text{ for } t > 0\,, \\ \theta(0) = \theta_0 \qquad\qquad\qquad\qquad \text{for } t = 0\,, \end{cases}$$

where u is incompressible, $\theta_0 \in L^2_0(\mathbb{T}^d)$. By saying passive scalar, we mean the tracer θ provides no feedback to the advecting velocity field u. Since u is divergence free, we have

$$\|\theta\|_{L^2} \leqslant e^{-\nu\lambda_1 t}\|\theta_0\|_{L^2}\,, \tag{1.1}$$

where $\lambda_1 > 0$ is the smallest non-zero eigenvalue of $-\Delta$ on $\mathbb{T}^d$. In [6], Poon proved that $\|\theta\|_{L^2}$ satisfies a double exponential lower bound

$$\|\theta(t)\|_{L^2} \geqslant \exp(-C\nu\gamma^t)\|\theta_0\|_{L^2} \tag{1.2}$$

for some constants $C > 0$ and $\gamma > 1$. In [1] (with G. Iyer), we quantify the decaying rate of $\|\theta\|_{L^2}$ by using the notion of *dissipation time* in [7] (see also [8, 9]). The dissipation time is defined as

$$\tau_d := \inf\left\{ t \,|\, \|\theta(t)\|_{L^2_0} \leqslant \frac{\|\theta_0\|_{L^2_0}}{e},\ \text{for all } \theta_0 \in L^2_0 \right\}.$$

Then (1.1) and (1.2) imply that $C|\ln \nu| \leqslant \tau_d \leqslant \frac{1}{\lambda_1 \nu}$. Constantin, Kiselev et. al. [10] proved that the dissipation time $\tau_d = o(\nu^{-1})$ if and only if $u \cdot \nabla$ has no H^1 eigenfunctions other than

constants. We call such incompressible flows relaxation enhancing flows. Since weakly mixing flows are relaxation enhancing flows, the dissipation time for weakly mixing flows is at least $o(\nu^{-1})$. In the particular case of shear flows a stronger estimate on the dissipation time can be obtained using Theorem 1.1 in [11], which guarantees that

$$\tau_d \leqslant C\frac{|\ln \nu|^2}{\nu^{1/2}}\,. \tag{1.3}$$

For us, we hope to generally obtain a rate of the dissipation time based on the "mixing rate" of u. We quantify the mixing rate of u by imposing a rate at which these convergences occur. Let $X\colon \mathbb{T}^d \to \mathbb{T}^d$ be the flow map of u defined by

$$\partial_t X = u(X(t),t) \qquad \text{and} \qquad X(0) = \mathrm{Id}\,. \tag{1.4}$$

Recall, (strongly) mixing maps are those for which the correlation $\langle f\circ X(t), g\rangle$ decays to 0 as $t\to\infty$ for all $f, g \in L^2_0$. Here $\langle\cdot,\cdot\rangle$ is used to denote the L^2_0 inner-product. We say that the vector field u is strongly 1, 1 mixing with rate function h if for all $f\in \dot{H}^1$, $g\in\dot{H}^1$ we have

$$|\langle f\circ X(t), g\rangle| \leqslant h(t)\|f\|_{H^1}\|g\|_{H^1}\,, \tag{1.5}$$

where $h : [0,\infty) \to [0,\infty)$ is some continuous, decreasing function that vanishes at ∞ and $\dot{H}^1$ represents the homogeneous Sobolev space of order 1.

The rate of the dissipation time is given by the following theorem. Here for simplicity, only strongly 1, 1 mixing case is discussed. Our method works for strongly α, β mixing with $\alpha > 0$, $\beta > 0$ as well.

Theorem 1.1. *If u is strongly 1, 1 mixing with rate function h, then the dissipation time is bounded by*

$$\tau_d \leqslant \frac{C}{\nu H_1(\nu)}\,. \tag{1.6}$$

Here C is a universal constant, and $H_1\colon (0,\infty)\to(0,\infty)$ is defined by

$$H_1(\mu) = \sup\Big\{\lambda \,\Big|\, \frac{\lambda\exp\left(4\|\nabla u\|_{L^\infty} h^{-1}(\frac{1}{2\lambda})\right)}{h^{-1}(\frac{1}{2\lambda})} \leqslant \frac{\|\nabla u\|^2_{L^\infty}}{2\mu}\Big\}\,, \tag{1.7}$$

where h^{-1} is the inverse function of h.

Based on Theorem 1.1, we can explicitly tell the rate of the dissipation time when the mixing rate function h is explicit. We discuss two special cases below.

- If the mixing rate function h is a power law, i.e. $h = \frac{c}{t^p}$, then
$$\tau_d \leqslant \frac{C}{\nu|\ln\nu|^p}\,.$$
- If the mixing rate function h is exponential, i.e. $h = c_1\exp(-c_2 t)$, then
$$\tau_d \leqslant \frac{C}{\nu^\delta}\,, \qquad \text{where} \qquad \delta \overset{\text{def}}{=} \frac{4\|\nabla u\|_{L^\infty}}{c_2 + 4\|\nabla u\|_{L^\infty}}\,.$$

The above two cases were also recently studied by Coti Zelati, Delgadino and Elgindi [12]. Here the authors show that if the mixing rate is given by a power law, i.e. $h = \frac{c}{t^p}$, then the dissipation time is bounded by

$$\tau_d \leqslant \frac{C}{\nu^{2/2+p}}\,. \tag{1.8}$$

Alternately, if the mixing rate is exponential, i.e. $h = c_1 \exp(-c_2 t)$, then [12] shows that the dissipation time is bounded by

$$\tau_d \leqslant C|\ln \nu|^2 . \tag{1.9}$$

In both these cases, the bounds provided by [12] are stronger than those provided by our theorem.

Recall that based on the work of Poon [6], the L^2 norm of the solution satisfies a double exponential lower bound (1.2). To the best of our knowledge, there are no incompressible smooth divergence free vector fields for which the lower bound (1.2) is attained. Moreover, recent work of Miles and Doering [13] suggests that the Batchelor length scale may limit the long term effectiveness of mixing forcing only a single-exponential energy decay. However, if we transform such problem from a continuous time setting into a discrete time setting, then we can prove the lower bound (1.2) is sharp and can be attained by some exponentially mixing map.

In the discrete time setting, we construct a so-called *pulsed diffusion* model, which is a mixing dynamical system interposed with diffusion,

$$\theta_{n+1} = e^{\nu\Delta} U \theta_n . \tag{1.10}$$

Here $U\colon L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is the Koopman operator associated with φ, which is defined by $Uf = f \circ \varphi$. And $\varphi : \mathbb{T}^d \to \mathbb{T}^d$ is a smooth volume preserving diffeomorphism. To connect with the continuous case, here φ can be chosen as $\varphi = X(1)$ where X is the flow map defined in (1.4).

Similar to the continuous case, here we also study the dissipation time of the map φ which is strongly 1, 1 mixing.

Theorem 1.2. *If φ is strongly 1, 1 mixing with rate function h, then the dissipation time satisfies the bound*

$$\tau_d \leqslant \frac{C}{\nu H_2(\nu)} . \tag{1.11}$$

Here C is a universal constant which can be chosen to be 34, and $H_2\colon (0,\infty) \to (0,\infty)$ is defined by

$$H_2(\mu) \stackrel{\text{def}}{=} \sup\Big\{ \lambda \,\Big|\, h\Big(\frac{1}{2\sqrt{\lambda\mu}}\Big) \leqslant \frac{1}{2\lambda} \Big\} . \tag{1.12}$$

As before, we compute the dissipation time τ_d in two useful cases.

- If the mixing rate function h is a power law, i.e. $h = \frac{c}{t^p}$, then

$$\tau_d \leqslant \frac{C}{\nu^{2/(2+p)}} . \tag{1.13}$$

- If the mixing rate function h is exponential, i.e. $h = c_1 \exp(-c_2 t)$, then

$$\tau_d \leqslant C|\ln \nu|^2 . \tag{1.14}$$

In the discrete case, our results (1.13) , (1.14) are consistent with the results (1.8) , (1.9) in [12]. We verify that *toral automorphisms* are nice candidates that satisfies strongly mixing with exponential rate function h. Recall a toral automorphism is a map of the form

$$\varphi(x) = Ax \pmod{\mathbb{Z}^d} ,$$

where $A \in SL_d(\mathbb{Z})$ is an integer valued $d \times d$ matrix with determinant 1. Maps of this form are known as "cat maps", and one particular example is when $d = 2$ and

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

We then explicitly calculate the dissipation time of pulsed diffusion driven by a toral automorphism, and find that

$$\|\theta_n\|_{L^2} \leqslant \exp(-C\nu\gamma^t)\|\theta_0\|_{L^2} \tag{1.15}$$

for some constants $C > 0$ and $\gamma > 1$. This matches with Poon's lower bound (1.2). Together with (1.2) and (1.15), we get the dissipation time τ_d for toral automorphism is of order $|\ln \nu|$. For general exponentially mixing diffeomorphisms, however, it is not known whether the dissipation time is still of order $|\ln \nu|$. According to our result in the discrete case (1.14), we can only get the dissipation time τ_d is of order $|\ln \nu|^2$. In the future, we hope to improve such bound (1.14) when the map mixes exponentially, or find an exponentially mixing map which gives a dissipation time of order $|\ln \nu|^2$.

1.2. **Markov Semigroups for SGD and Online PCA.** Many nonconvex optimization tasks involve finding desirable stationary points. The stochastic gradient decent algorithm and its variants enjoy favorable computational and statistical efficiency and are hence popular for these tasks. The optimization problem can be formulated as

$$\min_{x \in \mathbb{R}^d} f(x) := \mathbb{E} f(x, \xi), \tag{1.16}$$

where $f(x, \xi)$ is the stochastic loss function and ξ is sampled from some distribution $\mathcal{D}$. The stochastic gradient descent (SGD) is then to consider

$$x_{n+1} = x_n - \eta \nabla f(x_n; \xi_n), \tag{1.17}$$

where $\{\xi_n\}$ are i.i.d. random variables sampled from the distribution $\mathcal{D}$ and independent of $\{x_n\}$ and η is the learning rate. We expect that $\{x_n\}$ can lead to some approximation solution to the optimization problem (1.16). It is straightforward to observe that the iteration $\{x_n\}$ generated by (1.17) forms a discrete time, time-homogeneous Markov chain. Let $\mathbb{E}_{x_0}$ denote the expectation under the distribution of this Markov chain starting from x_0 and $\mu^n(\cdot; x_0)$ be the law of x_n. For a fixed test function $\varphi \in L^\infty(\mathbb{R}^d)$, we define

$$U^n(x_0) = \mathbb{E}_{x_0} \varphi(x_n) = \int_{\mathbb{R}^d} \varphi(y) \mu^n(dy; x_0). \tag{1.18}$$

Given the SGD (1.17), we find explicitly that

$$U^{n+1}(x) = \mathbb{E}(U^n(x - \eta \nabla f(x; \xi))) =: SU^n(x). \tag{1.19}$$

Then, $U^0 = \varphi$ and $\{S^n\}_{n \geq 0}$ forms a semigroup for the Markov chain. In [2], with J.-G. Liu and L. Li, we study the properties of the discrete semigroup. Properties including regularity preserving, L^∞ contraction are discussed. This semigroup is the dual of the semigroup for evolution of probability, while the latter is L^1 contracting and positivity preserving. Based on these properties of the semigroup, we show that the discrete semigroup can be approximated by some continuous semigroup in the weak sense, namely we find a diffusion process which solves an SDE and whose trajectory is close to the SGD trajectory in a weak sense.

To be specific, for $0 \leqslant n \leqslant \lfloor T/\eta \rfloor$, the dynamics of (1.17) can be approximated by the following SDE with weak accuracy $O(\eta^2)$,

$$dX_t = -\nabla(f(x) + \frac{1}{4}\eta|\nabla f(x)|^2)\,dt + \sqrt{\eta\Sigma}\,dW\,, \tag{1.20}$$

where

$$\Sigma = \mathrm{Var}(f(x,\xi))\,.$$

For online PCA, we proceed in a similar way and obtain analogous results. This diffusion approximation to SGD is also discussed in [14]. Our approach relies on semigroups. However, this approximation in general is only valid for $0 \leqslant n \leqslant \lfloor T/\eta \rfloor$ but not for long time. The difficulty is that the coefficients in (1.20) are unbounded and the behavior of the SDE is indeed hard to investigate. In the future, we aim to overcome this difficulty.

1.3. **Time-fractional 1-D Autonomous ODEs.** Fractional calculus in continuous time has been used widely in physics and engineering for memory effect, viscoelasticity, porous media, etc [15–17]. Given a smooth function φ, letting $n-1 < \gamma < n$, where n is an integer, the Caputo derivative of order γ is defined as

$$D_c^\gamma \varphi(t) = \frac{1}{\Gamma(n-\gamma)} \int_0^t \frac{\varphi^{(n)}(s)}{(t-s)^{\gamma+1-n}}\,ds\,, \quad t>0\,,$$

which is the fractional integral of $\varphi^{(n)}$ with order $n-\gamma$. In [3, 4], we use a slightly modified definition of the Caputo derivative in [5] to investigate the nonlinear fractional ODE

$$D_c^\gamma u = f(u),\ u(0) = u_0, \tag{1.21}$$

for $\gamma \in (0,1)$. Here f is locally Lipschitz with domain containing u_0. It is well-known that the solutions of 1-D autonomous ODEs with usual first order derivative are monotone, since the solution curves never cross zeros of f and $f(u)$ has a definite sign. One of our main results is that if $f \in C^1$ and f' is locally Lipschitz, the first order derivative of the solution to the fractional ODE (1.21) does not change sign and therefore the solution is monotone.

Similar to PDEs, the comparison principles are also very important in the analysis of the time-fractional PDEs. There are many versions of comparison principles proved in literature using various definitions of Caputo Derivatives. In [3], with J.-G. Liu et. al, we assume $f(t,\cdot)$ to be non-decreasing. In [18, Lemma 2.6], $f(t,\cdot)$ is assumed to be non-increasing, In [19, Theorem 2.3], there is no restriction on the monotonicity of $f(t,\cdot)$, but the function v is assumed to be C^1 so that the pointwise value of $D_c^\gamma v$ can be defined. Inspired by the work [19], we [4] establish a more generalized version of comparison principles with no monotonicity assumption on function $f(t,\cdot)$ and no C^1 assumption on v.

Theorem 1.3. *Suppose $f(t,u)$ is continuous and locally Lipschitz in u. Let $v(t)$ be a continuous function. If $D_c^\gamma v \leq f(t,v)$ in the distributional sense, and $D_c^\gamma u = f(t,u)$, with $v_0 \leq u_0$. Then, $v \leq u$ on the common interval. Similarly, if we have $D_c^\gamma v \geq f(t,v)$ as distributions and $v_0 \geq u_0$, then $v \geq u$ on the common interval.*

Besides comparision principles, we also study the asymptotic behaviors for a class of fractional ODEs (1.22).

$$D_c^\gamma u = Au^p,\ u(0) = u_0 > 0\,. \tag{1.22}$$

The reason for studying such a particular class of fractional ODEs is that typically *a priori* estimates of certain energies of the solution to a fractional PDE have form $D_c^\gamma E \leqslant AE^p$. By the comparison principles mentioned above, the energy E may be controlled by the solution

to fractional ODE (1.22). Hence studying the behavior of the solution to this fractional ODE (1.22) is important for the analysis of fractional PDEs.

In [3], based on an Osgood type blow-up criteria, we find relatively sharp bounds of the blow-up time T_b in the case $A > 0$, $p > 1$. These bounds indicate that as the memory effect becomes stronger ($\gamma \to 0$), if the initial value is big, the blow-up time T_b tends to zeros, while if the initial value is small, the blow-up time T_b tends to infity. In the case $A < 0$, $p < 1$, we show that the solution decays to zero more slowly compared with the usual derivative. Later on, in [4], we give a complete description regarding the asymptotic behavior of the solution curves to (1.22).

2. **Future Work**

In the future, I would like to continue working on problems about mixing and diffusion. Also I will continue to investigate the machine learning algorithms from math point of view.

- Mixing. As mentioned at the end of Section 1.1, it is not known whether the bounds we got so far are optimal. I will try to either improve the bounds or find some nice mixing map to verify those bounds are optimal. Also, another interesting topic in this area is to study the decay of H^1 norm of the solution. In Section 1.1, we see that the L^2 norm of the solution decays at least exponentially fast, and can be tremendously improved when the mixing map is sufficiently mixing. In such case, the behavior of H^1 norm of the solution is worth to study. And the long time behavior of the length scale $\|\theta\|_{H^1}^2/\|\theta\|_{L^2}^2$ also deserves to get explored.
- Machine learning algorithms. As I have already mentioned in Section 1.2, l hope to study the long-time asymptotic behavior of the SGD. In the case that the stochastic loss function $f(\cdot, \xi)$ is strongly convex, the dynamics of the SGD (1.17) is trapped and we may approximate the dynamics for long time. The difficulty is that the coefficients of the SDE (1.20) are unbounded and the behavior of the SDE is indeed hard to investigate. To overcome this difficulty, with J.-G. Liu, L. Li et. al, we are currently trying to use the formal asymptotic expansion to approximate the SGD for long time.

Besides what I mentioned above, I am glad to learn new tools and expand my research areas. As already stated, I am interested broadly in applied mathematics, and happy to work on any interesting problems arising in the interdisciplinary areas.

References

[1] Yuanyuan Feng and Gautam Iyer. Dissipation enhancement by mixing. *arXiv preprint arXiv:1806.03699*, 2018.
[2] Yuanyuan Feng, Lei Li, and Jian-Guo Liu. Semi-groups of stochastic gradient descent and online principal component analysis: properties and diffusion approximations. *Commun. Math. Sci*, 2018.
[3] Yuanyuan Feng, Lei Li, Jian-Guo Liu, and Xiaoqian Xu. Continuous and discrete one dimensional autonomous fractional odes. *Discrete Contin. Dyn. Syst. Ser. B*, 23(8):3109–3135, 2018.
[4] Yuanyuan Feng, Lei Li, Jian-Guo Liu, and Xiaoqian Xu. A note on one-dimensional time fractional ODEs. *Appl. Math. Lett.*, 83:87–94, 2018.
[5] Lei Li and Jian-Guo Liu. A generalized definition of caputo derivatives and its application to fractional odes. *SIAM Journal on Mathematical Analysis*, 50(3):2867–2900, 2018.
[6] Chi-Cheung Poon. Unique continuation for parabolic equations. *Comm. Partial Differential Equations*, 21(3-4):521–539, 1996.
[7] Albert Fannjiang and Lech Wołowski. Noise induced dissipation in Lebesgue-measure preserving maps on d-dimensional torus. *J. Statist. Phys.*, 113(1-2):335–378, 2003.
[8] Albert Fannjiang, Stéphane Nonnenmacher, and Lech Wołowski. Dissipation time and decay of correlations. *Nonlinearity*, 17(4):1481–1508, 2004.

[9] Albert Fannjiang, Stéphane Nonnenmacher, and Lech Wołowski. Relaxation time of quantized toral maps. *Ann. Henri Poincaré*, 7(1):161–198, 2006.
[10] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš. Diffusion and mixing in fluid flow. *Ann. of Math. (2)*, 168(2):643–674, 2008.
[11] Jacob Bedrossian and Michele Coti Zelati. Enhanced dissipation, hypoellipticity, and anomalous small noise inviscid limits in shear flows. *Arch. Ration. Mech. Anal.*, 224(3):1161–1204, 2017.
[12] M. Coti Zelati, M. G. Delgadino, and T. M. Elgindi. On the relation between enhanced dissipation time-scales and mixing rates. *ArXiv e-prints*, June 2018.
[13] Christopher J Miles and Charles R Doering. Diffusion-limited mixing by incompressible flows. *Nonlinearity*, 31(5):2346, 2018.
[14] Q. Li, C. Tai, and W. E. Stochastic modified equations and adaptive stochastic gradient algorithms. In *International Conference on Machine Learning*, pages 2101–2110, 2017.
[15] R. Gorenflo and F. Mainardi. *Fractional Calculus*. Springer, 1997.
[16] K. Diethelm and N. J. Ford. Analysis of fractional differential equations. *J. Math. Anal. Appl.*, 265(2):229–248, 2002.
[17] K. Diethelm. *The analysis of fractional differential equations: An application-oriented exposition using differential operators of Caputo type*. Springer, 2010.
[18] V. Vergara and R. Zacher. Optimal decay estimates for time-fractional and other nonlocal subdiffusion equations via energy methods. *SIAM J. Math. Anal.*, 47(1):210–239, 2015.
[19] J.D. Ramirez and A.S. Vatsala. Generalized monotone iterative technique for Caputo fractional differential equation with periodic boundary condition via initial value problem. *Int. J. Differ. Equations*, 2012, 2012.

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213.
E-mail address: yuanyuaf@andrew.cmu.edu