Singularity Formation: Synergy in Theoretical, Numerical and Machine Learning Approaches.

Thesis by
Yixuan Wang

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended April 16th, 2026

Yixuan Wang
ORCID: 0000-0001-7305-5422

ACKNOWLEDGEMENTS

I am finishing up my thesis with extreme emotions, humility, and gratitude. What a journey! I spent some of my prime years towards my PhD, and it has been extremely rewarding: tons of ups and downs, but what is important is that it has been fun all along.

First and foremost, I express my sincere gratitude to the thesis and candidacy committee, along with other faculty at Caltech. Prof. Tom Hou, my supervisor, with great admiration, I grew and thrived under his training. Tom is a giant scholar, but even before that, he is a wonderful person: patient with the whole group, caring and devoting tons of his time, working extremely hard by himself, always there for us. It's through Tom that I learned how to devote myself to research and be a good person. I extend my thanks to the extended Tom's group, including his kind wife Dr. Chang, all the PhD students I have interacted with, and Mr. Choi. Prof. Anima Anandkumar, I have collaborated with her at Caltech most extensively besides Tom. Her vision and passion ignite me and guide me through my early stages with machine learning. Prof. Houman Owhadi, my chair and the very first person I knew at Caltech, has kindly provided guidance and interactions with his research group. Prof. Franca Hoffman, inspires me greatly through her outreach and caring for the community, where I have built the SIAM student chapter at Caltech myself. Prof. Kewen Wu, who will join online, my undergrad cohort, every conversation with him has been stimulating and insightful as a wonderful junior researcher and motivates me to stay eager and humble for learning. Prof. Andrew Stuart, who I had the pleasure to collaborate with, will unfortunately be away for my defense, has helped throughout my PhD. The help of CMS staff makes my stay in the department even nicer. I am forever in debt to Diana and Sydney, who have left. Diana really made Steele House her proper house and cared about each individual; everything about her is emotional. Then I would also like to thank Joy, Minzhi, and Jolene for their help.

I have to mention my amazing collaborators, many of whom are also close and dear friends. Prof. Ziming Liu, of course, what KAN I say but so many words: it is just amazing being the best friend all along, and thanks for being a role model. Prof. Zongyi Li, your humility and kindness always physically inform me. Prof. Van Tien Nguyen, thanks for being so encouraging and guiding me through the blowup world. Changhe Yang, the uniquely devoted collaborator, your dedication pays off

greatly, and it was extremely rewarding on board the grand project you carried. Prof. Yifan Chen, we have known each other for so long and quite deeply; our relationship has always been multiscale, but thanks for watching out for me. Prof. Jiajie Chen, you are simply the strongest PDE person, period; thanks for being patient as a big brother. Tao Zhou, it was so much fun meeting you at a conference, becoming good friends, and working on good things together, like what academia should be. Prof. Jonathan Siegel, your insights and kindness help make our paper happen. Dr. Valentin Duruisseaux, you are a dedicated and extremely talented collaborator, but you are also such a good friend and tennis buddy. I would also like to thank all the kind people writing me academic reference letters, although I decided to pursue math with AI at scale in the industry instead: Prof. Tom Hou, Prof. Anima Anandkumar, Prof. Andrew Stuart, Prof. Javier Serrano-Gomez, Prof. Hatem Zaag, Prof. Tristan Buckmaster, Prof. Arthur Jacot, Prof. Alexandru Ionescu, and Prof. Philippe Rigollet. The majority of math professors are just so nice to junior folks like me.

Caltech is great only because of its people. I would first like to thank my best friends in the department, without whom I would say, without exaggeration, that I would not have pulled through my PhD. Pau, my very first friend, we just have to pat each other on the shoulder to know that we are there. Matthieu, we have so much in common: the movie nights, concert nights, national parks, squash sessions, deep talks, and academic discussions; we are everything. Peicong, despite being a younger one in the group, you helped me so much emotionally. We experienced a lot of things together and had so much fun, and I really appreciated it. Han, thanks for playing tennis, hanging out at concerts and others, but most importantly, for bringing in deep thoughts about life, about the journey. It is just a shame that despite many math discussions, I did not end up having a single paper with my best friends at Caltech. Then, I would say that my life beyond research also provided me with extreme joy and emotional support. I would like to thank the team, busy beaver, and my tennis buddies: Xijin, Haowen, Aaron, Daniel, Blake, Analiese, Sam, Steven, Sean, Max, Frid, Edo, and Robert. I was often not the best player, to put it mildly, but playing and being on the court is the most important thing. Concert is a big part of life that makes me energetic, and I appreciate the company of my good friends: Tianshu, Berthy, Alex, and Jingyuan. Movies extend my and everybody's life, and my foodie friends, including Mansour and Steven, take it to another level. Cardio dance, especially Mariam, empowers me and gives me a little community, including the kind lady at the postal office who kindly let me mail my tax just before closing.

It is just so easy to get scared and lost without a community. I cherish in particular my Chinese heritage, with my international horizon, just like the favorite dish I cook, Avocado Mapo Tofu, sometimes with a Parmesan twist. I would like to thank the Chinese community here, including Xinyi, Yulu, Ding, Hongkai, Ray, Bohan, Qiren, Chris, Xiaozhou, and Jim.

I am extremely thankful to my friends all over the world: Nina, Luna, Roger, Anita, Vincent, Shawn, and Yihan. But especially friends from my college, Peking, the greatest place for an undergrad due to its comprehensive vision and greatest people, including the ones in the math department: Weiyuan, Zexuan, Ran, Jiasen, Jiayi, Jialei, Shangqin, Mengxi, and Yibo. I wanted to be emotional and write a paragraph to each one of you folks, but I will do (and have done) more on a postcard. You all have shaped me and witnessed me. Of course, I was grateful for the training and teachers there as well, including my undergrad advisor Prof. Ruo Li.

Finally, I thank my dearest family. I am extremely proud, but may not have expressed it due to the East Asian shyness, to have the best parents in the world. My dad Tao Wang is a silent protector for me, and my mom Fang Liu, oh my mom, is just so caring and kind. They are always there for me, as if I am the center of their universe. They are never too busy, and nothing about me is too difficult for them. I regretted that in my adolescence, we had some fights, but overall, we have a very lovely and healthy relationship. I want to look up to them so that in the future, I could be just as marginally good a parent. I am dedicating this thesis to my grandmom, Xia Lan. As I grew up with the four of us in a family, she is the closest person in my childhood. Ms. Lan, she grew up in troubled times in China, so she is not lucky to experience the world as I am fortunate to now. Her vision, open-mindedness, and insights are nonetheless profound and shaped me.

Moving forward, as I have always been, I will embark on and enjoy a love-hate, hopefully love on the upper side, relationship with Math, at DeepMind. It is a creative time, so do creative things. I am looking forward to the new paradigm of AI+math (PDEs): embracing the scale of things and hopefully witnessing grokking and phase-transition therein. I express my sincere gratitude to the senior people who helped me take this giant leap of faith, Dr. Yongji Wang, Prof. Chingyao Lai, Prof. Javier Serrano-Gomez, Prof. Tristan Buckmaster, and my team leader Dr. Ray Jiang.

ABSTRACT

This thesis develops numerical and theoretical approaches for understanding and analyzing singularity formation in Partial Differential Equations (PDEs). The singularity formation in the Navier-Stokes Equation (NSE) is famously challenging as one of the seven Clay Prize problems. Unlike simpler equations such as the Nonlinear Heat (NLH) or Keller-Segel (KS) equations, where formal asymptotics near blowup are better understood, the intrinsic complexity of NSE makes quantitative analytical treatment difficult, if not impossible, without numerical guidance.

Building on numerical insights, Chapter 2 3 and 4 introduce a robust analytical framework to simplify and systematize pen-and-paper proofs for singular PDEs. We present a novel approach based on enforcing vanishing modulation conditions for perturbations around approximate blowup profiles, complemented by singularly weighted energy estimates. Blowups are proven with a clear notion of stability, with rates automatically inferred, without the need to know the asymptotics a priori or explicit spectral information of the linearized operator. We demonstrate the efficacy of our method on PDEs with complicated asymptotics, such as NLH and the Complex Ginzburg-Landau (CGL) equation, and address the open problem of singularity formation in the 3D KS equation with logistic damping. We also provide a roadmap for extending our techniques to singularities involving multiple scales.

In Chapter 5 and 6, we develop and refine numerical approaches that facilitate deeper insights into singularity formation. We demonstrate that machine learning methods significantly enhance our capability to identify and characterize potential blowup solutions with high precision. We improve on existing Physics-Informed Neural Network (PINN) and Neural Operator (NO) frameworks. Moreover, we present a novel machine learning paradigm, the Kolmogorov-Arnold Network (KAN) architecture, whose interpretability and excellent scaling properties are achieved through learnable nonlinearities inspired by the Kolmogorov-Arnold representation theorem.

Chapter 7 introduces Exponential Multiscale Finite Element Method (ExpMsFEM), developed to efficiently solve challenging multiscale PDEs beyond elliptic problems, such as the Helmholtz equation. We construct adaptive local bases, proving exponential convergence theoretically and demonstrating superior computational performance in practice. Like KAN, ExpMsFEM exemplifies how insights from theory can guide the design of high-performance solvers with theoretical guarantees.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Thomas Y Hou, Van Tien Nguyen, and Yixuan Wang. "L^2-based stability of blowup with log correction for semilinear heat equation". In: *Archive for Rational Mechanics and Analysis* 250.3 (2026), p. 28. DOI: 10.1007/s00205-026-02191-7.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

[2] Jin Lee, Ziming Liu, Xinling Yu, Yixuan Wang, Haewon Jeong, Murphy Yuezhen Niu, and Zheng Zhang. "KANO: Kolmogorov-Arnold neural operator". In: *The Fourteenth International Conference on Learning Representations*. 2026. URL: https://openreview.net/forum?id=2QmiKXfsIr.
Y.W. contributed to the conception and analysis of the project.

[3] Spyros Rigas, Dhruv Verma, Georgios Alexandridis, and Yixuan Wang. "Initialization schemes for Kolmogorov-Arnold networks: An empirical study". In: *The Fourteenth International Conference on Learning Representations*. 2026. URL: https://openreview.net/forum?id=dwNXKkiP51.
Y.W. contributed to the conception and analysis of the project.

[4] Jiajie Chen, Thomas Y Hou, Van Tien Nguyen, and Yixuan Wang. "On the stability of blowup solutions to the complex Ginzburg-Landau equation in R^d". In: *Annals of PDE* 11.2 (2025), p. 29. DOI: 10.1007/s40818-025-00223-1.
Y.W. contributed to the conception, analysis, and writing of the project.

[5] Thomas Hou, Yixuan Wang, and Changhe Yang. "Nonuniqueness of Leray-Hopf solutions to the unforced incompressible 3D Navier-Stokes Equation". In: *arXiv preprint arXiv:2509.25116* (2025).
Y.W. contributed to the conception, analysis, and writing of the project.

[6] Jiaqi Liu, Yixuan Wang, and Tao Zhou. "Finite time blowup for Keller-Segel equation with logistic damping in three dimensions". In: *arXiv preprint arXiv:2504.12231* (2025).
Y.W. contributed to the conception, analysis, and writing of the project.

[7] Ziming Liu, Andrew M Stuart, and Yixuan Wang. "Second order ensemble Langevin method for sampling and inverse problems". In: *Communications in Mathematical Sciences* 23.5 (2025), pp. 1299–1317. DOI: 10.4310/CMS.250517001811.
Y.W. contributed to the conception, analysis, and writing of the project.

[8] Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. "Kan 2.0: Kolmogorov-arnold networks meet science". In: *Physical Review X* 15.4 (2025), p. 041051. DOI: 10.1103/4t7t-v19l.
Y.W. contributed to the conception and writing of the project.

[9] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. "KAN: Kolmogorov–Arnold Networks". In: *The Thirteenth International Conference on Learning Representations*. 2025. URL: https://openreview.net/forum?id=Ozo7qJ5vZi.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

[10] Yixuan Wang, Ziming Liu, Zongyi Li, Anima Anandkumar, and Thomas Y Hou. "High precision PINNs in unbounded domains: application to singularity formulation in PDEs". In: *arXiv preprint arXiv:2506.19243* (2025).
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

[11] Yixuan Wang, Jonathan W. Siegel, Ziming Liu, and Thomas Y. Hou. "On the expressiveness and spectral bias of KANs". In: *The Thirteenth International Conference on Learning Representations*. 2025. URL: https://openreview.net/forum?id=ydlDRUuGm9.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

[12] Yifan Chen, Thomas Y Hou, and Yixuan Wang. "Exponentially convergent multiscale finite element method". In: *Communications on Applied Mathematics and Computation* 6.2 (2024), pp. 862–878. DOI: 10.1007/s42967-023-00260-2.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the review paper.

[13] Thomas Y Hou and Yixuan Wang. "Blowup analysis for a quasi-exact 1D model of 3D Euler and Navier–Stokes". In: *Nonlinearity* 37.3 (2024), p. 035001. DOI: 10.1088/1361-6544/ad1c2f.
Y.W. contributed to the conception, analysis, computer-assisted verification, and writing of the project.

[14] Zongyi Li, Samuel Lanthaler, Catherine Deng, Yixuan Wang, Kamyar Azizzadenesheli, and Anima Anandkumar. "Scale-consistent learning with neural operators". In: *Neurips 2024 Workshop Foundation Models for Science: Progress, Opportunities, and Challenges*. 2024.
Y.W. contributed to the conception, analysis, and writing of the project.

[15] Yifan Chen, Thomas Y Hou, and Yixuan Wang. "Exponentially convergent multiscale methods for 2d high frequency heterogeneous Helmholtz equations". In: *Multiscale Modeling & Simulation* 21.3 (2023), pp. 849–883. DOI: 10.1137/22M1507802.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

[16] Haydn Maust, Zongyi Li, Yixuan Wang, Daniel Leibovici, Oscar Bruno, Thomas Hou, and Anima Anandkumar. “Fourier continuation for exact derivative computation in physics-informed neural operators”. In: *arXiv preprint arXiv:2211.15960* (2022).
Y.W. contributed to the conception, analysis, and writing of the project.

[17] Yifan Chen, Thomas Y Hou, and Yixuan Wang. “Exponential convergence for multiscale linear elliptic PDEs via adaptive edge basis functions”. In: *Multiscale Modeling & Simulation* 19.2 (2021), pp. 980–1010. DOI: 10.1137/20M1352922.
Y.W. contributed to the conception, analysis, numerical experiments, and writing of the project.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

Chapter 1

INTRODUCTION

Much of this thesis is motivated by the challenge of understanding singular behaviors in nonlinear partial differential equations, exemplified by the 3D incompressible Navier–Stokes equations (NSE):

$$\mathbf{u}_t + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla \mathbf{p} + \nu \Delta \mathbf{u}, \quad \nabla \cdot \mathbf{u} = 0\,. \tag{1.0.1}$$

The 3D NSE is a cornerstone of mathematical physics and fluid dynamics, yet foundational questions—such as whether smooth initial data can develop singularities in finite time—remain open. Singular behaviors like blowup and turbulence not only raise deep theoretical issues but also complicate numerical simulation and physical interpretation.

This thesis is guided by a central principle: that rigorous analysis and numerical insight are most powerful when developed in tandem. Our work illustrates how numerical methods not only provide guidance and experimental evidence for conjectures but also serve as a foundation for rigorous proofs and inspire new analytical frameworks. In turn, the theoretical insights yield more robust, interpretable, and generalizable numerical methods.

More generally, we consider finite-time singularities in evolution equations of the form

$$a_t = F(a), \quad \limsup_{t \to T^-} \|g(a(t))\|_{L^\infty} = \infty, \quad T < +\infty\,, \tag{1.0.2}$$

where $g(a)$ is the quantity of interest, such as a itself or its gradient ∇a.

A canonical mechanism for finite-time blowup is via self-similar solutions, which take the form

$$a(x,t) = (T-t)^\alpha \tilde{U}(y)\,, \quad y = x(T-t)^\beta\,. \tag{1.0.3}$$

This ansatz exploits the scaling symmetry of the equation, reducing the original PDE to a time-independent profile equation for $\tilde{U}$. Typically, $\tilde{U}$ is smooth and bounded, making both analysis and computation more tractable. Substituting into the original equation and equating powers of $T - t$, one obtains a steady-state equation of the form

$$\mathcal{F}(\tilde{\lambda}, \tilde{U}) = 0\,, \tag{1.0.4}$$

where the scalar $\tilde{\lambda}$ related to α and β governs the blowup rate, and the operator $\mathcal{F}$ consists of a scaling operator plus the original operator F. Specific forms of $\mathcal{F}$ will be illustrated in the examples that follow.

The study of blowup in the self-similar regime thus reduces to analyzing the existence and stability of solutions to (1.0.4). Since the profile is now a 'nice' function, one naturally asks: can numerical methods help identify such profiles, and more importantly, can they inform or even enable rigorous proofs?

A common strategy is to solve (1.0.4) directly by casting it as an optimization problem for $\tilde{U}$ and the associated scaling parameter $\tilde{\lambda}$, where neural networks become relevant due to their expressive power and seamless integration to modern computations at scale. However, one must build on techniques tailored to these specifically challenging problems. In particular, we build on **operator learning** and introduce Scale-Informed Neural Operator, which encodes scaling symmetries of PDEs, and Fourier Continuation for (Physics-Informed) Neural Operators for PDEs on the whole space. We also build on **high-precision training of Physics-Informed Neural Networks (PINNs)**, where exact enforcement of asymptotics and hard constraints becomes crucial for upgrading numerical candidates to rigorous proofs. Motivated by the goal of symbolic profile recovery, we propose the **Kolmogorov-Arnold Network (KAN)** architecture, which offers greater interpretability and compositional expressivity compared to conventional neural networks.

Alternatively, one can introduce artificial time marching and solve the equation forward in time with appropriate initial data, with the hope of converging to the steady profile. This dynamical approach, however, requires precise control over unstable directions, often arising from continuous symmetries such as scaling or translation. Ruling out these instabilities forms a key aspect of both our theoretical framework and our numerical design.

To upgrade from numerical candidates with residue to a rigorous proof, perturbative argument and stability analysis are required. One can either use a fixed-point argument to establish solutions to the nonlinear equation (1.0.4) of the perturbation around the approximate profile, or analyze the convergence to the steady state of its time-dependent counterpart.

Specifically, we consider a perturbative ansatz of the form $\tilde{\lambda} = \bar{\lambda} + \lambda, \tilde{U} = \bar{U} + U$, and substitute it into (1.0.4). This yields a perturbed equation of the form

$$0 = \mathcal{L}(\lambda, U) + \mathcal{N}(\lambda, U) + \mathcal{E}\,, \tag{1.0.5}$$

where $\mathcal{L}$ is the linearized operator associated with $\mathcal{F}$ at $(\bar{\lambda}, \bar{U})$, $\mathcal{N}$ is the nonlinear part, and $\mathcal{E}$ is the residue from the approximate profile, which can be made arbitrarily small via high-precision computation and verified with computer-assisted proofs.

A clear characterization of stability is therefore crucial in either approach. A common strategy is to introduce a carefully chosen energy norm E, typically induced by an inner product on a Hilbert space, under which the linear, nonlinear, and residual terms satisfy:

$$(\mathcal{L}, U)_E \leq -c_1 \|U\|_E^2 , \quad (\mathcal{N}, U)_E \leq c_2 \|U\|_E^3 , \quad \|\mathcal{E}\|_E \leq c_3 , \tag{1.0.6}$$

with positive constants. Here, c_3 is very small with $c_1^2 > 4c_2c_3$. Then we can construct an energy ball in either approach and demonstrate the existence and stability of a solution to (1.0.5). Of course, the construction of the norm is extremely challenging and requires a case-by-case analysis. Nonetheless, we provide a general recipe for a range of equations using **singular weights**, capturing localized behaviors of the solution.

In practice, one needs to deal with potentially unstable directions. Some of those are induced by scaling symmetries of the equation, and one can enforce **local orthogonality constraints** to those directions via the introduction of extra scaling parameters in the artificial time formulation, perturbing the symmetries, which constitutes one of our key theoretical and numerical contributions. Full stability and exact asymptotics can be inferred or precisely computed without prior knowledge, amenable to the case of implicit profiles and computer-assisted proofs. In other cases where only finite codimensional stability is expected, we couple the argument with a **topological argument** handling the unstable modes. For even more challenging scenarios where damping is hard to extract directly for the linearized operator $\mathcal{L}$, one can **decompose it into a coercive one plus a compact perturbation**. Applying finite-rank approximations to the compact part, one can formally regard it as a nonlinear component and only needs to verify boundedness inverting on the finite-dimensional space, where computer-assisted proofs can again be applied. We rigorously established the open problem of nonuniqueness of Leray solutions in the 3D incompressible Navier-Stokes equation [210]. Intuitively, the onset of nonuniqueness occurs after the singularity time, and building upon previous works, we reduce the problem to constructing a forward self-similar profile whose associated linearized operator admits a positive eigenvalue. We numerically compute the approximate profile and the eigenpair first, before rigorously establishing the existence of the exact solutions by a perturbative argument.

The remainder of this chapter is organized around the central themes introduced above. We highlight our contributions in developing both theoretical frameworks and numerical tools, including ideas that extend beyond self-similarity. Beyond singularity formation, we also point to broader connections with scientific computing, AI for science, and machine learning, where the methodologies developed here may provide useful perspectives and technical ingredients for related problems.

1.1 Numerics Inspire Proofs: Blowup via Local Modulations

We will elaborate on the general theoretical framework for stability analysis around an approximate profile, based on the perturbative formulation in (1.0.5) and the damping/nonlinearity/residue bounds in (1.0.6). As mentioned before, the linear damping estimate constitutes the most crucial part. We built upon the series of seminal works developed by Chen-Hou-Huang [76, 72, 70, 73] to introduce singular weights in L^2 estimates and extract damping. As an illustrative example, consider the following rescaled Riccati-type equation:

$$\hat{u}_\tau = -\hat{u} - \frac{1}{2} z \hat{u}_z + \hat{u}^2 \,, \tag{1.1.1}$$

where it can be interpreted as the Riccati ODE $a_t = a^2$ in the self-similar ansatz

$$a(x,t) = (T-t)^{-1} \hat{u}(x(T-t)^{-1/2}, -\log(T-t)) \,.$$

Equation (1.1.1) admits a family of explicit steady states $\bar{u} = (1+cz^2)^{-1}$. These profiles are known to be stable under even perturbations that vanish to fourth order at the origin.

Towards rigorously proving it, we can establish linear dynamical stability for even perturbations of $O(z^4)$ at the origin by a L^2 estimate with singular weight $|z|^{-\alpha}$ capturing the vanishing order at the origin. To be specific, we compute the damping estimate by an integration by parts near the origin as

$$(\mathcal{L}u, u\rho) \approx (-1 + \frac{(\rho z)_z}{4\rho} + 2\bar{u})(u, u\rho) \approx \frac{5-\alpha}{4}(u, u\rho) \,.$$

For $\alpha > 5$, which corresponds to the vanishing order of u greater than 2, we thus have damping near the origin. Notice that in this simple case, the initial vanishing order of the perturbation is preserved throughout the dynamics. To control the nonlinear terms, we complement the singularly-weighted L^2 estimate with a high-order H^k estimate for sufficiently large k.

Building on this idea, we established the finite time singularity of a modified Hou-Li model [219]. The Hou–Li model was introduced in [215] as a 1D reduction of the

3D axisymmetric Navier–Stokes equations along the symmetry axis. To be specific, the axisymmetric NSE (in cylindrical coordinates) takes the form

$$\begin{aligned} u_{1,t} + u^r u_{1,r} + u^z u_{1,z} &= 2u_1\psi_{1,z} + \nu\Delta u_1 \,, \\ \omega_{1,t} + u^r \omega_{1,r} + u^z \omega_{1,z} &= \left(u_1^2\right)_z + \nu\Delta\omega_1 \,, \\ -\left[\partial_r^2 + (3/r)\partial_r + \partial_z^2\right]\psi_1 &= \omega_1 \,, \end{aligned}$$

where $u_1 = u^\theta/r,\ \omega_1 = \omega^\theta,\ \psi_1 = \psi^\theta/r$, and u^θ, ω^θ, and ψ^θ are the angular velocity, angular vorticity, and angular stream function, respectively. The Hou-Li model, of the following form:

$$\begin{aligned} u_{1,t} + 2\psi_1 u_{1,z} &= 2\psi_{1,z}u_1 + \nu u_{1,zz} \,, \\ \omega_{1,t} + 2\psi_1\omega_{1,z} &= \left(u_1^2\right)_z + \nu\omega_{1,zz} \,, \\ -\psi_{1,zz} &= \omega_1 \,, \end{aligned}$$

is a reduction in the sense that if (ω_1, u_1, ψ_1) is an exact solution of the 1D model, we can obtain an exact solution of the 3D Navier-Stokes equations by using a constant extension in r. The model became particularly relevant in the investigation of potential interior singularities for NSE, and the authors in [215] established its well-posedness in C^m.

Since the discovery and proof of the Hou-Luo scenario of boundary singularity in the Euler equations [298, 299, 73], Hou discovered new numerical evidence that the 3D axisymmetric Euler and Navier-Stokes equations develop potential singular solutions at the origin [212, 213]. Hou's numerical simulations revealed that the axial velocity $u^z = 2\psi_1 + r\psi_{1,r}$ near the peak of u_1 is significantly weaker than $2\psi_1$. This weakening is attributed to the shift between the maxima of ψ_1 and u_1: ψ_1 peaks at $r = r_\psi$, while u_1 peaks at $r = r_u$ with $r_\psi < r_u$, implying $\psi_{1,r} < 0$ near the maximal of u_1. Thus the axial velocity u^z is actually weaker than its value at $r = 0$, highlighting that the original Hou–Li model—being confined to the symmetry axis—fails to capture this subtle, yet critical, three-dimensional effect. To better understand this phenomenon and its relation to singularity formation, we introduce the following 1D weak advection model.

$$\begin{aligned} u_t + 2a\psi u_z &= 2u\psi_z + \nu u_{zz} \,, \\ \omega_t + 2a\psi\omega_z &= \left(u^2\right)_z + \nu\omega_{zz} \,, \\ -\psi_{zz} &= \omega \,, \end{aligned} \tag{1.1.2}$$

where a is a parameter that measures the relative strength of advection in the Hou-Li model. We established the following theorems in [219]:

Theorem 1.1.1. *For the weak advection model* (1.1.2) *in the inviscid case* $\nu = 0$*, there exists a constant* $\delta > 0$ *such that for* $a \in (1-\delta, 1)$*, the weak advection model* (A.1.5) *develops a finite time singularity for some* C^∞ *initial data. Moreover, there exists a self-similar profile* $(\omega_\infty, \omega_\infty, \omega_\infty)$ *corresponding to a blowup that is neither expanding nor focusing. More precisely, the blowup solution to* (A.1.5) *has the form*

$$\omega(x,t) = \frac{1}{1+c_{u,\infty}t}\omega_\infty\,, u(x,t) = \frac{1}{1+c_{u,\infty}t}u_\infty\,, \psi(x,t) = \frac{1}{1+c_{u,\infty}t}\psi_\infty\,,$$

for some negative constant $c_{u,\infty}$ *with a blowup time given by* $T = \frac{-1}{c_{u,\infty}}$.

Theorem 1.1.2. *Consider the Hou-Li model* (A.1.4) *in the inviscid case* $\nu = 0$*. For any* $\alpha < 1$*,* (A.1.4) *develops a finite time singularity for some* C^α *initial data. Moreover, there exists a* C^α *self-similar profile corresponding to a blowup that is neither expanding nor focusing.*

Theorem 1.1.3. *Consider the weak advection model* (A.1.5) *with viscosity. There exists a constant* $\delta_1 > 0$ *such that for* $a \in (1-\delta_1, 1)$*, the weak advection model* (A.1.5) *develops a finite time singularity for some* C^∞ *initial data.*

We leave the details and the proofs concerning the Hou-Li model to Appendix A.

In the remainder of this section, we elaborate on our framework for proving finite-time blowup in PDEs beyond classical self-similar or fully stable scenarios, using local modulation and singularly-weighted energy estimates. In many cases—especially when viscous terms are present—the blowup deviates from exact self-similarity. To capture these behaviors, we introduce additional modulation parameters that enforce vanishing conditions locally, allowing us to infer both the blowup rate and its stability directly from the modulation dynamics. Those modulations, stemming from computational concerns of numerical stability, turn out to be keys to the theoretical stability. We illustrate this framework across a sequence of increasingly complex examples. In Subsection 1.1.1, we discuss the semilinear heat equation, which resembles the Riccati toy model (1.1.1). Subsection 1.1.2 follows as a much more technical resolution of the complex Ginzburg-Landau equation with its full stability without symmetry assumptions. We solve the open problem of the singularity formation of the 3D Keller-Segel equation with logistic damping in Subsection 1.1.3, where we combine modulation analysis with a topological argument to handle finite codimensional instability. We emphasize key conceptual ideas here for clarity, and refer to Chapter 2 for a complete treatise of the semilinear heat

equation based on [218], while leaving the technically challenging parts of complex Ginzburg-Landau [77] and 3D Keller-Segel equations [286] to Chapters 3 and 4.

1.1.1 Nonlinear heat equation

We consider the semilinear heat equation, one of the simplest and best-studied PDEs with singularities.

$$a_t = \Delta a + a^2 \,,$$

with the blowup profile

$$a(x,t) \sim \frac{1}{T-t}\bar{u}\left(\frac{x}{\sqrt{(T-t)|\log(T-t)|}}\right), \quad \bar{u}(\xi) = \frac{1}{1+|\xi|^2/8}.$$

One can see that its blowup scaling resembles the Riccati-type equation, except for a log-correction. Moreover, in its blowup scaling, the diffusion term becomes asymptotically small and to its leading order coincides with (1.1.1). This log-correction is of course subtle to compute and capture, requiring special attention.

Inspired by (1.1.1), we employ a similar stability argument using singularly-weighted L^2 norm and a high-order H^k norm. Extra modulations are required to ensure the vanishing conditions of the perturbation, which are inspired by numerical computations aimed at capturing the steady-state profile. We thus introduce general rescaling beyond self-similarity as

$$\hat{u}(z,\tau) = \hat{C}_u(\tau) a(\hat{C}_l(\tau) z, t(\tau)) \,,$$

where

$$\hat{C}_u = \hat{C}_u(0)\exp\Big(\int_0^\tau \hat{c}_u(s)ds\Big),\ \hat{C}_l = \exp\Big(-\int_0^\tau \hat{c}_l(s)ds\Big),\ t = \int_0^\tau \hat{C}_u(s)ds,$$

with $\hat{c}_u$ and $\hat{c}_l$ being determined. We introduced the extra degree of freedom $\hat{C}_u(0)$, which we will later choose to be small for the estimates of the viscous term. The renormalized equation for $\hat{u}$ reads as

$$\hat{u}_\tau = \hat{c}_u\hat{u} - \hat{c}_l z\hat{u}_z + \hat{u}^2 + \frac{\hat{C}_u}{\hat{C}_l^2}\hat{u}_{zz}.$$

We then consider the approximate profile $\bar{u} = (1+z^2/8)^{-1}$ that solves

$$\bar{c}_u\bar{u} - \bar{c}_l z\bar{u}_z + \bar{u}^2 = 0, \quad \text{for } \bar{c}_u = -1\,, \bar{c}_l = 1/2\,.$$

If we enforce that u is an even function satisfying $u(0,\tau) = u_{zz}(0,\tau) = 0$ for all time τ, we have by the dynamic rescaling equation,

$$\hat{c}_u + \bar{u}(0) + \frac{\hat{C}_u \bar{u}_{zz}(0)}{\hat{C}_l^2 \bar{u}(0)} = 0\,, \hat{c}_u - 2\hat{c}_l + 2\bar{u}(0) + \frac{\hat{C}_u(\bar{u}_{zzzz}(0) + u_{zzzz}(0))}{\hat{C}_l^2 \bar{u}_{zz}(0)} = 0\,.$$

Defining

$$\lambda(\tau) = \frac{\hat{C}_u(\tau)}{\hat{C}_l^2(\tau)} = \hat{C}_u(0) \exp\Big(\int_0^\tau (c_u(s) + 2c_l(s))ds \Big)\,,$$

we can simplify the normalization constraints into

$$c_u - \frac{1}{4}\lambda = 0\,, c_u - 2c_l - \big(\frac{3}{2} + 4u_{zzzz}(0)\big)\lambda = 0\,.$$

This gives

$$c_u = \frac{1}{4}\lambda\,, \quad c_l = -(\frac{5}{8} + 2u_{zzzz}(0))\lambda\,,$$

from which we derive the ODE for λ as

$$\lambda_\tau = \lambda(c_u + 2c_l) = -(1 + 4u_{zzzz}(0))\lambda^2\,.$$

We can formally infer the log-correction from the asymptotics $\lambda \approx 1/\tau \approx |\log(T-t)|^{-1}$. Those extra modulations ensure that we can work in the singularly weighted space, and we generalize to high dimensions via different spatial scalings in different coordinates. Computations also corroborate our theoretical findings beyond the small perturbation regime; for details see Chapter 2.

1.1.2 Complex Ginzburg-Landau equation

We consider the complex Ginzburg-Landau equation

$$\psi_t = (1 + \imath\beta)\Delta\psi + (1 + \imath\delta)|\psi|^{p-1}\psi - \gamma\psi,$$

which reduces to the semilinear heat equation we have considered when $\beta = \delta = \gamma = 0$, $p = 2$, while connecting with the nonlinear Schrödinger equation in the limit $\beta, |\delta| \to \infty$. It enjoys a similar blowup law with log corrections, and we aim to establish full stability beyond the assumption of even symmetries.

Upon introducing the phase-amplitude decomposition, we obtain

$$\begin{aligned}
\partial_t u &= \big[\Delta - |\nabla\theta|^2\big]u - \beta\big(2\nabla u \cdot \nabla\theta + u\Delta\theta\big) + u^p - \gamma u,\\
u\partial_t\theta &= \beta\big[\Delta - |\nabla\theta|^2\big]u + 2\nabla u \cdot \nabla\theta + u\Delta\theta + \delta u^p.
\end{aligned}$$

One sees that to the leading order in u, it resembles the nonlinear heat equation again, and we can adopt a similar stability argument, with the following extra difficulties, however.

- We need to deal with full stability, corresponding to modulations of terms all the way to second-order at the origin, with $1 + d + d(d+1)/2$ degree of freedom in d-dimension. We thus introduce a matrix rescaling in space as:

$$U(z,\tau) = H(\tau)u\big(\mathbf{R}(\tau)z + V(\tau), t(\tau)\big)\,,$$

$$\Theta(z,\tau) = \theta\big(\mathbf{R}(\tau)z + V(\tau), t(\tau)\big), \quad t(\tau) = \int_0^\tau H^{p-1}(s)ds,$$

  where $\mathbf{R}(\tau) \in \mathbb{R}^{d\times d}$ is upper triangular, $V(\tau) \in \mathbb{R}^d$ and $H(\tau) \in \mathbb{R}_+$. The modulation corresponds to the symmetries of the equation, with $d-1$ extra modulation parameters. Similarly to the semilinear heat equation, we now have a matrix $Q = H^{p-1}\mathbf{R}^{-1}\mathbf{R}^{-1,T}$ capturing the log-correction and making the viscous terms formally small.

- General nonlinearity and the phase equation necessitate a lower bound of U: we use the maximal principle and a weighted L^∞ estimate.

- Sharp decay estimates of $\nabla^i U$ require almost tight power for the weights and interpolation and embedding inequalities.

- Coupling of amplitude and phase needs a top-order energy estimate with special algebraic structure to cancel out top-order terms in diffusion.

$$\mathscr{D}_U = \Delta_Q U - 2\beta\langle\nabla U, \nabla\Theta\rangle_Q - U\langle\nabla\Theta,\nabla\Theta\rangle_Q - \beta U\Delta_Q\Theta\,,$$
$$\mathscr{D}_\Theta = \beta\frac{\Delta_Q U}{U} + 2\frac{\langle\nabla U,\nabla\Theta\rangle_Q}{U} - \beta\langle\nabla\Theta,\nabla\Theta\rangle_Q + \Delta_Q\Theta\,.$$

  We construct top-order energy as

$$(|\nabla^k W|^2, \rho_k) + (|\nabla^k\Phi|^2, U^2\rho_k)\,,$$

  where W, Φ are the perturbations in $U = \bar{U} + W, \Theta = \bar{\Theta} + \Phi$.

We refer to Chapter 3 for the technical details and estimates.

1.1.3 3D Keller-Segel equation with logistic damping

We now demonstrate the applicability of our method to singularity beyond full stability and without an explicit approximate profile. We focus on 3D Keller-Segel equation, a classical chemotaxis model, with a quadratic logistic damping term

$-\mu\rho^2$ modeling density-dependent mortality rate and show the existence of finite-time blowup solutions with nonnegative density and finite mass for a sharp range of $\mu \in \left[0, \frac{1}{3}\right)$. To be specific, consider the model

$$\begin{cases} \partial_t \rho = \Delta\rho - \nabla \cdot (\rho \nabla c) - \mu\rho^2, \\ \Delta c + \rho = 0. \end{cases}$$

The logistic damping makes the model inherently nonlocal, even in the radial symmetric assumption, unlike the original Keller-Segel equation, where a partial mass formulation could be introduced to make it local. Nonlocality makes the construction of approximate profiles challenging, and we deploy a phase-portrait method to establish the existence of a profile, treating the viscous term again as asymptotically small. To be specific, the radial profile satisfies

$$Q + \beta y \cdot \nabla_y Q = \nabla_y Q \cdot \nabla_y \Delta_y^{-1} Q + (1-\mu) Q^2,$$

with asymptotics $Q = \frac{1}{1-\mu} - |y|^{2j_0} + O(|y|^{2j_0+2})$ near the origin, where $\beta = \frac{1}{3(1-\mu)} + \frac{1}{2j_0}$. The left-hand side of the profile equation corresponds to the blowup scaling.

To establish the existence of the profile and make the viscous terms small, we require

$$\frac{1}{3(1-\mu)} < \beta < \frac{1}{2},$$

which coincides with the sharp range of μ.

We now conclude finite codimensional stability by introducing the rescaling

$$y = \frac{x}{\lambda^{2\beta}}, \quad \frac{d\tau}{dt} = \frac{1}{\lambda^2}, \quad \frac{\lambda_\tau}{\lambda} = -\frac{1}{2}, \quad \rho(t,x) = \frac{1}{\lambda^2} \Psi(\tau, y).$$

The equation in the rescaled variable is

$$\partial_\tau \Psi = \lambda^{2-4\beta} \Delta\Psi - \Psi - \beta y \cdot \nabla\Psi + \nabla\Psi \cdot \nabla\Delta^{-1}\Psi + (1-\mu)\Psi^2.$$

Consider the perturbative ansatz

$$\Psi = Q + \varepsilon^u(\tau, y) + \varepsilon^s(\tau, y), \text{ with } \varepsilon^u(\tau, y) = \sum_{j=0}^{K} c_j(\tau) \chi(y) |y|^{2j}.$$

For sufficiently large K, we have stability of the perturbation ε^s using singularly-weighted estimates, provided that ε^s vanishes to $O(|y|^{2K+2})$ near the origin, which will be enforced by local modulations of the coefficients c_j.

We can solve the dynamic equations for c_j as

$$\begin{cases} \dot{c}_j = \frac{j_0-j}{j_0} c_j + \lambda^{2-4\beta}[\Delta\Psi]_j + [N(\varepsilon^u)]_j, & 0 \le j < j_0, \\ \dot{c}_{j_0} = \sigma_{0,j_0} c_0 + \lambda^{2-4\beta}[\Delta\Psi]_{j_0} + [N(\varepsilon^u)]_j, & j = j_0, \\ \dot{c}_j = \frac{j_0-j}{j_0} c_j + \sum_{i=0}^{j-1} \sigma_{i,j} c_i + \lambda^{2-4\beta}[\Delta\Psi]_j + [N(\varepsilon^u)]_j, & j_0 < j \le K. \end{cases}$$

One can perform a standard bootstrap argument coupled with topological arguments to derive a finite codimensional stability with dimension $j_0 + 1$. Notice that here $j_0 < j \le K$ corresponds to fake unstable directions. We leave the technical details to Chapter 4.

1.2 Numerics Provide Basis for Proofs: High Precision NNs and KANs

In this section, we discuss the numerical tools we have developed for efficient computations of the profile; see the prototype equation (1.0.4). We will highlight how machine learning tools can contribute to the computation, especially when the profile is unstable, where numerical time marching might struggle with stability issues. The tools developed have a much broader impact beyond the interest of singularity formation, on general applications in scientific computing, AI for science, and general machine learning.

In Subsection 1.2.1, we discuss our contribution to operator learning, where we introduced Fourier continuation to deal with non-periodic problems, and also introduced Scale-Informed Neural Operator for scaling invariance of PDEs. We discuss the high-precision training of PINNs via exact enforcement of hard constraints and asymptotics in Subsection 1.2.2. KAN: Kolmogorov–Arnold Networks as a more interpretable and scaling-efficient neural network architecture, is introduced in Subsection 1.2.3. We sketch the high-level ideas here, and refer to our articles [313, 281] for details on developments on operator learning, Chapter 6 based on [445] for high precision PINN training, and Chapter 5 based on [293, 292, 446] for a complete overview of KAN.

1.2.1 Neural Operator with Fourier continuation and scaling invariance

Neural Operators (NOs) [277, 257] stand as an important idea to learn mappings directly between function spaces via parametrization of a nonlinear mapping in the spectral space by neural networks, thus offering the discretization-invariance property regardless of the physical mesh. In the context of PDE solving, NO aims to learn the solution operator directly, and can be coupled with the data loss and the physical loss, resulting in the Physics-Informed Neural Operator (PINO). Originally, PINO

leverages differentiation in the Fourier spectral space and is best suited for problems defined on a periodic domain. We extended the methodology by introducing Fourier continuation in FC-PINO [313] to deal with problems in the whole space, particularly suited for the aforementioned problems of singularity formation. FC-PINO enables efficient learning of blowup profiles, and more generally, families of profiles across varying scaling parameters—thanks to the operator learning formulation. Related to the scale-invariance in PDEs, we introduced the Scale-Informed Neural Operator [281]. Scale-consistency helps each model extrapolate to unseen scales, including the challenging Helmholtz and Navier-Stokes simulations. We refer to the details in our papers [313, 281]. Building on operator learning, which integrates supervised data with PDE constraints, we envision leveraging known profiles from similar equations to guide profile discovery in new regimes.

1.2.2 High precision training of PINNs

Motivated by the need for a numerical profile with high precision to be upgraded to a rigorous proof (1.0.6), we developed a high precision training procedure of PINNs on an unbounded domain in [445]. This approach emphasizes exact enforcement of asymptotic behavior, incorporation of hard constraints, and the use of advanced second-order optimization methods. We demonstrated an accuracy 4-digits better than previous state-of-the-art by the important work [449], in terms of a 2D Boussinesq equation related to the 3D Euler singularity on the boundary [73]. We fully open-sourced our codebase to facilitate future developments and reproducibility. Full technical details are provided in Chapter 6. Beyond architectures and physical inductive biases, we also work on developing optimizers for scalable and high-precision training, including our Self-scaled SOAP optimizer, and work on more challenging PDEs with singularities, including the weak convection model introduced in [289].

1.2.3 KAN

Partially motivated by an exact symbolic search for blowups, we proposed Kolmogorov–Arnold Networks (KANs) [293], as an interpretable alternative to modern machine learning architectures. The prevalent fully connected neural network, or Multi-Layer Perceptrons (MLPs), has learnable linear weights between layers and fixed nonlinear activation functions. It is the nonlinearity that renders them expressivity, albeit it often comes at the cost of a huge number of parameters to be trained. KANs leverage the Kolmogorov-Arnold representation theorem, which reduces the

representation of a multivariate function to a composition of univariate functions. This compositional construction aligns well with modern machine learning, and we generalize it to arbitrary depth. Better scaling laws are established under the compositionally smooth assumption empirically and theoretically, for a wide range of benchmarks of symbolic regression, PDE solving, and larger-scale problems of imaging. KANs demonstrate greater expressivity than MLPs and are particularly effective at capturing high-frequency components [446].

We further demonstrated the capacity of KANs to address a wide range of scientific problems [292], highlighting their synergy with AI for science. We also developed effective initialization schemes for KAN training [391] and integrated KAN into Operator Learning, providing a more interpretable and expressive alternative to existing Neural Operator frameworks [264]. We leave the detailed discussions to Chapter 5.

1.3 Numerics with Provable Guarantees: EKHMC and ExpMsFEM

Finally, in the last section, we talk about how insights from theory can guide the design of competitive solvers with theoretical guarantees.

We introduce two such methods: a preconditioned second-order sampler, Ensemble Kalman Hamiltonian Monte Carlo (EKHMC), in Subsection 1.3.1, and a state-of-the-art multiscale PDE solver, the Exponential Multiscale Finite Element Method (ExpMsFEM), in Subsection 1.3.2. We sketch the high-level ideas here and refer to Chapter 7 for a detailed exposition of ExpMsFEM, based on [81, 83, 82]. A full treatment of EKHMC appears in Appendix B, based on [291].

1.3.1 EKHMC

Consider the task of sampling from a target distribution

$$\pi(q) = \frac{1}{Z_q} \exp(-\Phi(q)) \,,$$

which arises naturally in Bayesian inverse problems, among other applications. The simplest approach is via simulation of the Langevin equation, which has the target density as its invariant distribution. Variants of Langevin dynamics have been proposed to boost convergence to the target distribution, among which are the Hamiltonian Monte Carlo with an auxiliary velocity variable introduced, and the preconditioning method to sample the target distribution with large condition numbers. Building upon the works of the Ensemble Kalman Sampler [158], where an interacting particle system was introduced to approximate the covariance as a

preconditioner, we proposed a second-order sampler [291] with empirically faster mixing rates.

$$\frac{\mathrm{d}q}{\mathrm{d}t} = p\,,$$
$$\frac{\mathrm{d}p}{\mathrm{d}t} = -C_q(\rho)D\Phi(q) - \gamma p + \sqrt{2\gamma C_q(\rho)}\frac{\mathrm{d}W}{\mathrm{d}t}\,.$$

We showed that EKHMC is affine-invariant and proposed a gradient-free version of the algorithm. In the context of Bayesian inverse problems with linear forward maps, we further demonstrated that the associated mean-field dynamics preserve Gaussianity and converge to the target distribution at a rate independent of the linear operator. We leave the detailed discussions to Appendix B.

1.3.2 ExpMsFEM

Consider the model problem in a bounded domain $\Omega \subset \mathbb{R}^d$ with a Lipschitz boundary Γ. Here, $d = 2$. For generality, the boundary can contain disjoint parts $\Gamma = \Gamma_1 \cup \Gamma_2$ where Γ_1 corresponds to the Dirichlet boundary conditions and Γ_2 corresponds to the Neumann and Robin type boundary conditions. The model equation is:

$$\begin{cases} -\nabla\cdot(A\nabla u) + Vu = f, \text{ in } \Omega \\ u = 0, \text{ on } \Gamma_1 \\ A\nabla u\cdot\nu = \beta u, \text{ on } \Gamma_2\,. \end{cases} \tag{1.3.1}$$

Here, A, V, β are functions in $L^\infty(\Omega)$ and can be rough, which makes the solution oscillating and difficult to solve. The vector ν is the outer normal to the boundary. In particular, when $V = 0$, the equation is the standard elliptic equation [81]. If $Vu = -k^2u$ and u is a complex-valued function, one obtains the Helmholtz equation [83] with wavenumber k.

Standard finite element methods often fail to resolve such problems accurately due to the limited regularity of the solution. The key challenge is to design numerical methods that adapt to the multiscale structure of the problem. The core idea is to construct local basis functions that adapt to the local coefficients of the equation. These bases allow efficient compression of the operator and can be reused to solve multiple problems with varying right-hand sides.

One key observation we proposed is to reduce the task of an accurate finite element solution to an accurate approximation. To be precise, we demonstrated via a posteriori estimates that once the bases can serve as good approximations to the solutions, the numerical solution automatically becomes a quasi-optimal approxi-

mation. Moreover, by a careful design of the local-to-global coupling, we showed that we only need to achieve a good approximation on each of the local patches.

Next, we discuss how to construct local approximations. This was achieved by a harmonic-bubble splitting. Roughly speaking, the harmonic part depends only on the information on the boundaries of the patch, and the bubble part depends only on the local right-hand side information. Even for challenging cases like the Helmholtz equation, we show that the bubble component remains small and that the harmonic part admits exponentially accurate approximation via oversampling. We rigorously proved exponential convergence and designed our algorithm to achieve such accuracy using only online basis functions, complemented by an offline modification of the right-hand side. We leave the detailed discussions and numerical experiments to Chapter 7.

Chapter 2

BLOWUPS VIA LOCAL MODULATIONS: NONLINEAR HEAT

In this chapter and the two subsequent chapters, we present our framework for inferring the precise law and stability of blowups beyond self-similarity via local modulations and singularly-weighted estimates, based on [218, 77, 286]. For problems with full stability, such as nonlinear heat (NLH) [218] or complex Ginzburg-Landau (CGL) [77] equations, we use a novel idea of enforcing stable normalizations for perturbations around the approximate profile and establish a weighted H^k stability, thereby avoiding the use of a topological argument and the analysis of a linearized spectrum. This result generalizes the L^2-based stability argument to blowups that are not exactly self-similar and can be adapted to higher dimensions. Full anisotropic scaling can be introduced to establish full stability beyond any symmetry assumption. The log correction for the blowup rate is automatically inferred via the local normalization conditions, captured by the energy estimates and refined estimates of the modulation parameters.

Crucially, this framework is applicable even when only numerical blowup profiles are available, providing a path toward rigorous analysis guided by computation. For cases involving finite-codimensional stability, our method can be combined with a topological argument. We illustrate those two points by solving the open problem of singularity formation in the 3D Keller–Segel (KS) equation with logistic damping [286]. Finally, numerical experiments confirm the effectiveness of our normalization strategy, even under large perturbations outside the strict theoretical regime.

For illustrative purposes, we present our framework in the setting of the NLH with even symmetry, covering both the fully stable and finite-codimensional cases in Section 2.5, in this chapter. Finally, in Section 2.6, we outline directions for extending this methodology to handle blowups involving multiple scales, and discuss prospects for turning the modulation approach into a robust computational algorithm beyond the search for singularities. The more technically involved cases, such as the CGL and 3D KS with logistic damping, are addressed in detail in the two following chapters respectively.

2.1 Introduction

We consider the semilinear heat equation

$$a_t = \Delta a + a^2, \tag{2.1.1}$$

where $a(t) : \mathbb{R}^n \to \mathbb{R}$, subject to the boundary condition $\lim_{|x|\to\infty} a(x,t) = 0$ and an initial data $a(0) = a_0$. Several blowup criteria were established in the past, dating back to the works of Kaplan [242], Fujita [156], Levine [268], Friedman-McLeod [153], etc. We refer to the book [380] for a comprehensive review on this subject. Given our interest in the singularity formation, in particular in characterizing blowup solutions to (2.1.1), we only mention works in this direction in the following, where a precise law of the blowup can be identified.

Singularity formation in nonlinear PDEs is generally connected to a group of symmetries associated with the problem under consideration. For the classical nonlinear heat equation (2.1.1), it is invariant under the scaling transformation

$$\forall \lambda > 0, \quad a_\lambda(x,t) = \frac{1}{\lambda^2} a\Big(\frac{x}{\lambda}, \frac{t}{\lambda^2}\Big), \tag{2.1.2}$$

in the sense that if a is a solution to (2.1.1), so is a_λ with the rescaled initial data $a_{0,\lambda} = \frac{1}{\lambda^2} a_0\big(\frac{x}{\lambda}\big)$. The earliest application of this scaling invariant property to equation (2.1.1) that we are aware of is the work by Berger-Kohn [27], where the authors developed a so-called rescaling algorithm to capture the blowup profile

$$a(x,t) \sim \frac{1}{T-t} \bar{u}\left(\frac{x}{\sqrt{(T-t)|\log(T-t)|}}\right), \quad \bar{u}(\xi) = \frac{1}{1+|\xi|^2/8}. \tag{2.1.3}$$

This blowup behavior is in agreement with classification results rigorously established in the works of Filippas-Kohn [150], Filippas-Liu [151], Herrero-Velazquez [197] and Velazquez [439] under the assumption of type-I blowup, namely that $\limsup_{t\to T} \|a(t)\|_{L^\infty}(T-t) < +\infty$, otherwise, the blowup is of type-II. In particular, Herrero-Velazquez [199] showed that the blowup behavior (2.1.3) is generic in dimension 1, and they claimed the same for higher dimensions in an unpublished work. A rigorous construction was later established by Bricmont-Kupiainen [45] to provide concrete examples of initial data leading to blowup behaviors classified in the works mentioned above. The method developed in [45] was generalized in the work of Merle-Zaag [324] through spectral analysis and a topological argument to establish the existence and stability of blowup solutions to (2.1.1) with the behavior (2.1.3).

It is worth mentioning that the scaling invariance (2.1.2) gives rise to the notion of energy-criticality in the sense that

$$\|a_\lambda\|_{\dot{H}^1} = \lambda^{n-6}\|a\|_{\dot{H}^1},$$

and the problem (2.1.1) is called energy-critical if $n = 6$, energy-subcritical if $n \le 5$ and energy-supercritical if $n \ge 7$. It is well known that the only type-I blowup occurs in the energy-subcritical case (i.e., $n \le 5$) from the work of Giga-Kohn [163, 164, 165], Giga-Matsui-Sasayama [166] (see also [167] for the case of convex domains and Quittner [379] for non-convex domains). The blowup in the energy-critical and supercritical cases is more subtle, where type-II blowup may also exist as predicted in [198] and [149] through formal matching asymptotic expansions. Concrete examples of initial data leading to type-II blowup for (2.1.1) with a general nonlinearity $|a|^{p-1}a$ were exhibited in several works [196], [331], [89], [90], [369], [405], [372], [370], [187], [371], [188] and references therein. In particular, the Type II blowup constructed in [188] for the energy-critical case in dimension $n = 6$ corresponding to the exponent $p = 2$ considered in this present work. Partial classification of type-II blowup was provided in [312, 311] and [91], even though a complete classification of all blowup patterns still remains open.

In this chapter, we are interested in adopting the idea of dynamic rescaling to provide rigorous proofs for the semilinear heat equation with a clear notion of stability. Specifically, just like in numerical algorithms, we introduce proper rescaling conditions to ensure the stability of the perturbation around the approximate steady state, whose proof constitutes the main goal of this chapter. We adopt a L^2-based stability analysis with properly chosen singular weights and normalization conditions, inspired by the line of work pioneered by [76, 72], and present our main result as Theorem 2.1.1. We introduce the weighted Sobolev space $\mathcal{E}_k$ for $k = 2n + 10$:

$$\mathcal{E}_k = \left\{u : \|u\|_{\mathcal{E}_k}^2 = \|u\|_\rho^2 + \mu\|\nabla^k u\|_{\rho_k}^2 < +\infty\right\}, \tag{2.1.4}$$

where $\|\cdot\|_\rho$ stands for the weighted L^2-norm with the weight functions ρ, ρ_k being defined in (2.3.1), (2.3.3), and the constant μ will be detailed later in (2.3.5).

Theorem 2.1.1. *Let $k = 2n + 10$, there exist positive constants C_0 and λ_0 such that for $0 < \lambda < \lambda_0$ and if the initial perturbation g is even*[1] *and satisfies $\|g\|_{\mathcal{E}_k} \le C_0\lambda$, the equation (2.1.1) with initial data*

$$a(x, 0) = \lambda^{-1}\big(\bar{u}(x) + g(x)\big),$$

[1] We call a multivariate function a even if it is even in each one of the coordinates, namely if $a(\xi_1 x_1, \xi_2 x_2, \cdots, \xi_n x_n) = a(x_1, x_2, \cdots, x_n)$ for any $\xi_i \in \{-1, 1\}$.

admits a solution $a(x,t)$ that blows up in some finite time T. Moreover, we have the following convergence in $\mathcal{E}_k$,

$$\lim_{t\to T}(T-t)a\left(((T-t)|\log(T-t)|)^{\frac{1}{2}}z,t\right)=\bar{u}(z).$$

Remark 2.1.2. *Using the scaling invariance of (2.1.1), we can introduce an initial rescaling in space (corresponding to introducing $\hat{C}_l(0)$ in the dynamic rescaling formulation (2.2.4) in Section 2.2) to obtain a result comparable to the theorems in [45, 324] that characterize the blowup time precisely. Here we highlight obtaining the correct rate and, for the sake of simplicity, do not rescale in space at $t=0$.*

Remark 2.1.3. *We work under the even assumption for illustration purposes of the dynamical rescaling technique developed in this chapter. We refer to our subsequent work [77] on the complex Ginzburg-Landau equation for a generalized dynamical rescaling technique to remove the even assumption, where translational and rotational modulations were introduced; see step 1 in Section 1.2 therein.*

Compared with most of the aforementioned works on semilinear equations that work in parabolic scaling, we work in variables that correspond to the true blowup scaling and obtain stability precisely with respect to the weighted H^k norms we construct, instead of resorting to a topological argument to identify the existence of the initial data for blowup.

2.1.1 Literature review and main contributions

The idea of dynamic rescaling formulation or the modulation technique to study blowup was originally introduced in the numerical study of self-similar blowup of the nonlinear Schrödinger equation [451, 418, 417, 315, 263]. Later on, the formulation has been generalized to various dispersive problems, both as numerical techniques and as an analysis tool; see for example nonlinear Schrödinger equation [246, 321], compressible Euler equations [49], and the nonlinear heat equation [27, 324]. Recently, this modulation technique has been adopted to establish self-similar singularity for incompressible Euler equations in [135, 72, 73].

When the equation admits an analytical approximate profile for blowups, analyzing the spectrum of the linearized operator has proven to be useful for establishing the blowup in many cases; see for example, the semilinear heat equation [45, 324] and the 2D Keller-Segel equation [385, 92]. While this methodology is powerful, it hinges on the fact that we are able to construct a simple and analytical approximate

steady state and can analyze the spectrum of the linearized operator explicitly (for semilinear heat equations) or at least asymptotically (for Keller-Segel equations). In this chapter, we provide a proof of blowup for the semilinear heat equation without analyzing the eigenvalues or the eigenfunction of the linearized operator at all, and we rule out the unstable directions via a clear characterization of a singularly weighted Sobolev space, instead of using Brouwer's fixed-point theorem and a topological argument. This framework can be adopted even if we only have a numerical or implicit profile and do not have explicit information on the spectrum of its linearized operator; see the follow-up work by the last author and collaborators [286] on 3D Keller-Segel equation with logistic damping.

On the other hand, a direct L^2 [76] or L^∞-based [73] stability argument with appropriate normalization conditions has been proven successful, even if no explicit approximate steady state can be identified. In fact, they are often combined with a numerical profile and rigorous computer-assisted proofs. See [76, 70, 219] for applications in various 1D models for the Euler equations, [72, 73] for 3D axissymmetric incompressible Euler equations, and [320] for the compressible Navier-Stokes equation (and the accompanying paper on the compressible Euler equation [319]). The nature of the blowup in [319, 320] and the current work is quite similar in the sense that the dominant behavior of the rescaled equations is driven by the inviscid part, although in each case one must use different scaling to get the precise asymptotic behavior. The methodology can be roughly summarized in the following two steps. Firstly, we link self-similar singularity with convergence to a steady state using the dynamic rescaling equation and obtain approximate steady states either analytically or numerically. Then, upon choosing appropriate normalization conditions, we can perform linear and nonlinear stability estimates to show that the perturbation around the approximate steady state will remain small. Therefore, we can obtain a self-similar blowup with rates prescribed by the normalizing constants.

This chapter adopts the L^2-based methodology to establish blowups beyond the self-similar setting; see also [93] on the 1D inviscid primitive equation and [24] on the harmonic map heat flow, where log corrections were also observed. We show that one can obtain the correct blowup rate by imposing proper vanishing conditions on the perturbation, without a priori knowledge of a formal blowup rate. Compared with a self-similar blowup, the crucial difference is that now we have an algebraic, instead of exponential, convergence of the normalizing constants in the rescaled time τ, inferred for example, by (2.2.2).

Another contribution is that we introduce different spatial rescalings in n different dimensions in Section 2.2.2, giving enough degrees of freedom for the normalization conditions. Those different rescaling constants in different dimensions will indeed converge to the same rescaling constant close to the blowup time. This approach may shed some light on the generalization of the dynamic rescaling framework to higher dimensions for other problems. We believe our method is robust for type-I singularities, especially for problems having non-self-adjoint linear operators, see for example our follow-up work [77] on the complex Ginzburg-Landau equation. Finally, we demonstrate our choice of normalization to be effective even beyond the regime of small perturbations, in Section 2.4 based on numerical experiments.

2.1.2 Notations

Throughout the chapter, we use $(\cdot,\cdot)$ to denote the inner product on $L^2(\mathbb{R}^n)$: $(f,g) = \int_{\mathbb{R}^n} fg$. We use C to denote absolute constants dependent only on the dimension n, which may vary from line to line. We use $A \lesssim B$ for positive B to denote that there exists a constant $C > 0$ such that $A \leq CB$. We adopt the notation of the Japanese bracket as $\langle z\rangle = \sqrt{1+|z|^2}$.

2.2 Dynamic Rescaling Formulation and Normalization Conditions

In this section, we discuss our dynamic rescaling formulation. Via enforcing local vanishing modulation conditions, we derive the law of the blowup formally and motivate the choices of singular weights for stability analysis.

2.2.1 1D case

We focus on the 1D case first and generalize to higher dimensions in Section 2.2.2. For the semilinear heat equation (2.1.1), we introduce the dynamic rescaling formulation

$$\hat{u}(z,\tau) = \hat{C}_u(\tau) a(\hat{C}_l(\tau) z, t(\tau)),$$

with

$$\hat{C}_u = \hat{C}_u(0)\exp\Big(\int_0^\tau \hat{c}_u d\tau\Big), \hat{C}_l = \exp\Big(\int_0^\tau -\hat{c}_l d\tau\Big), t = \int_0^\tau \hat{C}_u d\tau .$$

Here, we introduce an extra degree of freedom $\hat{C}_u(0)$ as in [70, 219], which we will later choose to be small for the estimates of the viscous term. We have

$$\hat{u}_\tau = \hat{c}_u\hat{u} - \hat{c}_l z\hat{u}_z + \hat{u}^2 + \frac{\hat{C}_u}{\hat{C}_l^2}\hat{u}_{zz} .$$

We know there exists an approximate profile $\bar{u} = (1 + z^2/8)^{-1}$ which solves

$$\bar{c}_u \bar{u} - \bar{c}_l z \bar{u}_z + \bar{u}^2 = 0 \, , \bar{c}_u = -1 \, , \bar{c}_l = 1/2 \, .$$

By using the dynamic rescaling formulation, we reduce the problem of establishing a blowup in the physical variables and quantifying its blowup rate to the problem of establishing stability in the dynamic rescaling formulation. We want to show that $\hat{u}$ converges to the steady state $\bar{u}$ of the dynamic rescaling equation and the normalization constants also converge. We put the ansatz

$$\hat{u} = \bar{u} + u \, , \hat{c}_u = \bar{c}_u + c_u \, , \hat{c}_l = \bar{c}_l + c_l \, .$$

We will elaborate on how to enforce normalization conditions c_u and c_l such that the perturbed solution u of the dynamic rescaling equation is stable for all time. Namely, we want to show that u remains small for all time, and thus $\hat{c}_u, \hat{c}_l$ will correspond to the correct blowup scaling.

If we enforce that the even perturbation satisfies $u(0)$ and $u_{zz}(0)$ vanish for all time, by the dynamic rescaling equation we have

$$\hat{c}_u + \bar{u}(0) + \frac{\hat{C}_u \bar{u}_{zz}(0)}{\hat{C}_l^2 \bar{u}(0)} = 0 \, , \hat{c}_u - 2\hat{c}_l + 2\bar{u}(0) + \frac{\hat{C}_u (\bar{u}_{zzzz}(0) + u_{zzzz}(0))}{\hat{C}_l^2 \bar{u}_{zz}(0)} = 0 \, . \quad (2.2.1)$$

Define

$$\lambda = \frac{\hat{C}_u}{\hat{C}_l^2} = \hat{C}_u(0) \exp \left(\int_0^\tau c_u + 2c_l d\tau \right) ,$$

we can simplify the normalization condition into

$$c_u - \frac{1}{4}\lambda = 0 \, , c_u - 2c_l - (\frac{3}{2} + 4u_{zzzz}(0))\lambda = 0 \, .$$

Therefore we solve

$$c_u = \frac{1}{4}\lambda \, , c_l = -(\frac{5}{8} + 2u_{zzzz}(0))\lambda \, . \quad (2.2.2)$$

And thus we can simplify the ODE for λ as

$$\lambda_\tau = \lambda(c_u + 2c_l) = -(1 + 4u_{zzzz}(0))\lambda^2 \, . \quad (2.2.3)$$

Remark 2.2.1. *Notice that formally, when the perturbation is small, we can further solve this ODE to obtain*

$$\lambda \approx 1/\tau = \frac{1}{|\log (T - t)|} \, .$$

Therefore, the effect of the viscosity terms can be treated perturbatively. We will make this heuristic argument rigorous later on by choosing $\hat{C}_u(0)$ small.

Remark 2.2.2. *To motivate our choice of normalization conditions, we can plug in an ansatz $\rho = z^{-\alpha}$ for the singular weight we use in the L^2 estimate, and calculate linear damping for the evolution of u. Via an integration by parts, we know that up to the linear part near the origin, we have*

$$(u_\tau, u\rho) \approx (-1 + \frac{1}{4}\frac{(\rho z)_z}{\rho} + 2)(u, u\rho) = (1 - \frac{\alpha - 1}{4})(u, u\rho)\,.$$

We calculate that we need $\alpha > 5$ to extract linear damping, and therefore we need to enforce the perturbation u to vanish to higher orders.

Of course, we need to take care of nonlinear estimates. Thus, the singular weights cannot be as simple as $\rho = z^{-6}$, but this serves as the starting point of our stability analysis.

2.2.2 nD case

A crucial idea in the n-dimensional case is that we introduce n different scaling parameters in different directions. This gives us more freedom to enforce the normalization conditions and obtain a perturbation with the same vanishing orders. Consider

$$\hat{u}(z, \tau) = \hat{C}_u(\tau) a(\hat{C}_l^1(\tau) z_1, \hat{C}_l^2(\tau) z_2, \cdots, \hat{C}_l^n(\tau) z_n, t(\tau))\,,$$

with the same $\hat{C}_u$ and $t(\tau)$ defined as before, and

$$\hat{C}_l^i = \exp\,(\int_0^\tau -\hat{c}_l^i d\tau)\,. \tag{2.2.4}$$

The equation for $\hat{u}$ is

$$\hat{u}_\tau = \hat{c}_u \hat{u} - \sum_i \hat{c}_l^i z_i \hat{u}_i + \hat{u}^2 + \sum_i \lambda_i \hat{u}_{ii}\,, \tag{2.2.5}$$

where we use the short-hand notation for partial derivatives: we denote $f_i = \partial_{z_i} f$ and $f_{ij} = \partial_{z_j}\partial_{z_i} f$. Using the same radial approximate steady state $\bar{u}, \bar{c}_l, \bar{c}_u$ and a similar ansatz

$$\hat{u} = \bar{u} + u\,, \hat{c}_u = \bar{c}_u + c_u\,, \hat{c}_l^i = \bar{c}_l + c_l^i\,, \tag{2.2.6}$$

we can enforce the same normalization condition that u is of $O(|z|^4)$. Notice that if we choose u to be an even perturbation, we only need to enforce $u(0) = 0$ and $u_{ii}(0) = 0$. Those $n + 1$ constraints can be solved exactly to obtain

$$c_u = \frac{1}{4}\sum_i \lambda_i\,, c_l^i = -\sum_j \lambda_j (\frac{1 + 4\delta_{ij}}{8} + 2u_{iijj}(0))\,, \tag{2.2.7}$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. Here we define

$$\lambda_i = \frac{\hat{C}_u}{(\hat{C}_l^i)^2} = \hat{C}_u(0) \exp\left(\int_0^\tau c_u + 2c_l^i d\tau\right),$$

and we obtain the ODE for λ as follows:

$$\partial_\tau \lambda_i = \lambda_i(c_u + 2c_l^i) = -(\sum_j 4u_{iijj}(0)\lambda_j + \lambda_i)\lambda_i \,. \tag{2.2.8}$$

Notice that (2.2.7), (2.2.8) are consistent with (2.2.2), (2.2.3) in the 1D case.

2.3 Stability of Perturbation and Finite Time Blowup

Building upon the general strategy of a weighted L^2-based stability argument as in [76, 72], we will prove Theorem 2.1.1 in this section. We denote $\lambda = \max_i \lambda_i$.

2.3.1 L^2 stability analysis

Plugging in the ansatz (2.2.6) into the dynamic rescaling equation (2.2.5) and using the fact that $\bar{u}$ is an approximate steady state, we write down the evolution equation for u as follows:

$$u_\tau = L(u) + N(u) + \sum_i F_i(z,\tau) + \sum_i \lambda_i V_i(u) \,,$$

where we reorganize the different terms into the linear, nonlinear, error, and viscous terms respectively as

$$L = (-1 + c_u)u - \sum_i (\frac{1}{2} + c_l^i) z_i u_i + 2\bar{u}u \,, \quad N = u^2 \,,$$

$$F_i = \frac{1}{4}\lambda_i \bar{u} - c_l^i z_i \bar{u}_i + \lambda_i(\bar{u}_{ii} + \sum_j \frac{1}{2} u_{iijj}(0) z_j^2 \chi(|z|)) \,,$$

$$V_i = u_{ii} - \sum_j \frac{1}{2} u_{iijj}(0) z_j^2 \chi(|z|) \,.$$

Here $\chi(z)$ is a 1D even smooth cutoff function such that $\chi(z) = 0$ for $|z| \geq 2$ and $\chi(z) = 1$ for $|z| \leq 1$. We introduce such a cutoff function to make each one of the four terms integrable in the weighted L^2 space. We name and group the terms in such a way that is convenient for our analysis. The "linearized operator L" is obtained by treating the scaling parameters c_u and c_l as known parameters, although they actually depend on u. As a result, the "linearized operator L" actually contains nonlinear terms in the original physical variables.

To show that the dynamic rescaling equation is stable for even perturbations and converges to a steady state, we will perform a weighted L^2 estimate with a singular weight ρ and a weighted L^2 norm

$$\rho = |z|^{-5-n} + 10^{-3}|z|^{1-n}, \quad \|f\|_\rho = (f^2, \rho)^{1/2}. \tag{2.3.1}$$

We choose such a weight to extract damping near the origin as in Remark 2.2.2, while also having good control of growth at infinity, to make L^∞ and thus the nonlinear estimates easier. Via an integration by parts[2], we have a standard L^2 estimate for the linear part:

$$(L, u\rho) = ([(-1 + c_u) + \frac{1}{2}\sum_i (\frac{1}{2} + c_l^i)\frac{(z_i\rho)_i}{\rho} + 2\bar{u}]u, u\rho).$$

We plug in the singular weight (2.3.1), using $O(\lambda)$ notations due to the form (2.2.7), and simplify as

$$\begin{aligned}&(-1 + c_u) + \frac{1}{2}\sum_i (\frac{1}{2} + c_l^i)\frac{(z_i\rho)_i}{\rho} + 2\bar{u}\\ =&O((1 + |\nabla^4 u(0)|)\lambda) - \frac{1}{4} + \frac{0.006}{4(10^{-3} + z^{-6})} - \frac{2z^2}{8 + z^2}.\end{aligned}$$

By a straightforward computation and the AM-GM inequality we have

$$0.006(8 + z^2) - 4(10^{-3} + z^{-6})2z^2 = 0.048 - 0.002z^2 - 8z^{-4} \le 0.$$

Therefore, we have the simple linear stability

$$(L, u\rho) \le (-\frac{1}{4} + O((1 + |\nabla^4 u(0)|)\lambda))(u, u\rho) \le (-\frac{1}{4} + C(1 + |\nabla^4 u(0)|)\lambda)\|u\|_\rho^2.$$

The estimate of the nonlinear term is straightforward:

$$(N, u\rho) \le \|u\|_\infty \|u\|_\rho^2.$$

[2]We can justify the integration by parts here, and similarly the subsequent ones, by a density argument. Notice that both terms are indeed integrable. For compactly supported smooth functions, we can, of course, do integration by parts since the boundary integral vanishes as the radius goes to infinity. For general cases, we can take a sequence of compactly supported smooth functions that approximates the functions in the weighted spaces and then take limits. By the Cauchy-Schwarz inequality, the integrals also converge, and we validate the integration by parts. Such approximate functions exist since we can first truncate the function in an annulus $\epsilon \le |z| \le 1/\epsilon$ such that the norm outside of the annulus is small; then the weighted norm is equivalent to a regular L^2-norm and we can approximate by compactly supported smooth functions inside the annulus [430] and zero extend to the whole space.

Now we compute the error terms. We compute $\bar{u}_{ii} = -\frac{\bar{u}^2}{4} + \frac{z_i^2\bar{u}^3}{8}$. As a consequence, we use Fubini's principle to arrive at

$$\sum_i F_i = \frac{\sum_i \lambda_i}{4}(\bar{u} + \frac{1}{2}z\cdot\nabla\bar{u} - \bar{u}^2) + \sum_i \frac{\lambda_i}{2}(z_i\bar{u}_i + \frac{z_i^2\bar{u}^3}{4}) + \sum_j\sum_i \frac{\lambda_i u_{iijj}(0)}{2}(4z_j\bar{u}_j + z_j^2\chi(|z|)) .$$

Using the fact that $\bar{u}$ is an approximate solution to the dynamic rescaling equation, we know that the first term vanishes. We can compute to simplify

$$\sum_i F_i = -\sum_i \lambda_i z_i^2|z|^2\frac{\bar{u}^3}{64} + \sum_j\sum_i \frac{\lambda_i u_{iijj}(0)z_j^2}{2}(\chi(|z|) - \bar{u}^2) .$$

We know that the error term is $O(|z|^4)$ at $z = 0$ and $O(|z|^{-2})$ at ∞; thus lies in the weighted space. We conclude that

$$(\sum_i F_i, u\rho) \le C(1 + |\nabla^4 u(0)|)\lambda\|u\|_\rho .$$

The viscous part is more subtle since we need to deal with the singularity carefully. Notice that

$$|\nabla^2\rho| \lesssim |\rho/|z|^2| .$$

We do integration by parts twice to derive

$$\begin{aligned}(V_i, u\rho) &= (-\frac{1}{2}u_{iijj}(0)z_i^2\chi(z)|z|, \frac{u}{|z|}\rho) - (u_i, u_i\rho) - (u_i, u\rho_i)\\ &\le -\|u_i\|_\rho^2 + C(|\nabla^4 u(0)|^2 + \|\frac{u}{|z|}\|_\rho^2) .\end{aligned}$$

Finally, we decompose the whole space into the near field $I = [-1, 1]^d$ and its complement I^c, notice that $u = O(z^4)$ at $z = 0$, we have the estimate

$$\|\frac{u}{|z|}\|_\rho^2 \lesssim \int_I \frac{u^2}{|z|^{7+n}} + \int_{I^c} u^2\rho \lesssim (\sup_I \frac{u}{|z|^4})^2 + \|u\|_\rho^2 \lesssim \|\nabla^4 u\|_\infty^2 + \|u\|_\rho^2 .$$

Denote $E_0^2 = (u, u\rho)$. We collect the L^2 estimate as

$$\begin{aligned}\frac{1}{2}\partial_\tau E_0^2 = (L + N + \sum_i (F_i + \lambda_i V_i), u\rho) &\le (-\frac{1}{4} + C(1 + \|\nabla^4 u\|_\infty)\lambda + \|u\|_\infty)E_0^2\\ &+ C\lambda(1 + \|\nabla^4 u\|_\infty)E_0 + C\lambda\|\nabla^4 u\|_\infty^2 .\end{aligned} \tag{2.3.2}$$

To close the L^2 estimate, we need higher-order estimates to control L^∞ norms.

2.3.2 Higher order stability analysis

Consider the weighted H^k norm for $k = 2n + 10$:

$$E_k^2(u) = (\nabla^k u, \nabla^k u \rho_k)\,, \quad \rho_k = 1 + 10^{-3k}|z|^{2k+1-n}\,, \tag{2.3.3}$$

and we will estimate

$$\frac{1}{2}\partial_\tau E_k^2 = (\nabla^k L + \nabla^k N + \sum_i (\nabla^k F_i + \lambda_i \nabla^k V_i), \nabla^k u \rho_k)\,.$$

Before we start, we will state the following lemma concerning interpolation inequalities of lower order terms and L^∞ estimates of a Morrey-type. We define the weighted auxiliary norms $D_j = \|\nabla^j u \langle z\rangle^{j+(1-n)/2}\|_2$. By (2.3.1) and (2.3.3), we know that $D_0 \lesssim E_0$, $D_k \lesssim E_k$.

Lemma 2.3.1. *For any $\nu > 0$, there exists a $C(\nu)$ such that the inequalities hold:*

$$D_j \le \nu D_k + C(\nu) D_0\,, \quad 0 \le j < k\,,$$
$$\|\nabla^j u \langle z\rangle^{i+1/2}\|_\infty \lesssim \|\nabla^{j+n} u \langle z\rangle^{i+(n+1)/2}\|_2, \quad 0 \le j, i\,.$$

Proof. We use an integration by parts to compute for $k > j > 0$:

$$D_j^2 = -\sum_i \int (\partial_i^2 \nabla^{j-1} u \cdot \nabla^{j-1} u \langle z\rangle^{2j+1-n} + \partial_i \nabla^{j-1} u \cdot \nabla^{j-1} u \partial_i \langle z\rangle^{2j+1-n})\,.$$

Noticing that $(\partial_i \langle z\rangle^{2j+1-n})^2 \lesssim \langle z\rangle^{2j+1-n}\langle z\rangle^{2j-1-n}$, by Cauchy-Schwarz inequalities, we have $D_j^2 \lesssim D_{j-1}(D_j + D_{j+1})$. By a weighted AM-GM inequality, we compute for any $\nu > 0$, $D_j^2 \le \nu D_{j+1}^2 + C(\nu) D_{j-1}^2$.

Now we prove the first interpolation inequality. Since ν is arbitrary, we only need to prove for $j = k - 1$, which we can use induction on k and the obtained inequality $D_j \le \nu D_{j+1} + C(\nu) D_{j-1}$ to conclude.

For the second inequality, we borrow the idea of proof for the embedding (3.13) of Proposition 1 in our follow-up paper [77]. We can assume $u \in C_c^\infty$ by a density argument and consider WLOG $z \in \mathbb{R}^n$ with $z_i \ge 0$ for all components. In the region $\Omega(z) = \{y \in \mathbb{R}^n, y_i \ge z_i\}$ we have $|y| \ge |z|$ for any $y \in \Omega(z)$. By Lebniz's rule and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |\nabla^j u(z)| &\lesssim \int_{\Omega(z)} |\partial_1 \partial_2 \cdots \partial_n \nabla^j u(y)| dy \\ &\lesssim \|\nabla^{j+n} u \langle y\rangle^{i+(n+1)/2}\|_2 \Big(\int_{|y|\ge|z|} \langle y\rangle^{-2i-n-1} dy\Big)^{1/2}\,. \end{aligned}$$

We thus collect the pointwise bound and conclude the proof of the inequality. □

With the lemma in mind, we denote the terms as lower order terms (l.o.t. for short) if their ρ_k-weighted L^2-norms are bounded by $\nu E_k + C(\nu)E_0$ for any $\nu > 0$. Notice that for $0 < j \le k$, $\nabla^j \bar{u} \nabla^{k-j} u$ are l.o.t. since we can estimate

$$|\nabla^j \bar{u}| \rho_k^{1/2} \lesssim \langle z \rangle^{-2-j} \langle z \rangle^{k+(1-n)/2} \lesssim \langle z \rangle^{k-j+(1-n)/2} ,$$

and thus their ρ_k-weighted L^2-norms are bounded by D_{k-j}. Therefore we collect the linear estimate via an integration by parts and $O(\lambda)$ notations as

$$\begin{aligned}(\nabla^k L, \nabla^k u \rho_k) \le ([-1 - \frac{k}{2} + \frac{1}{4}\frac{(z\rho_k)_z}{\rho_k} + 2\bar{u}]\nabla^k u, \nabla^k u \rho_k) \\ + C(1 + \|\nabla^4 u\|_\infty)\lambda E_k^2 + \nu E_k^2 + C(\nu) E_0^2 .\end{aligned}$$

We can compute the damping as

$$-1 - \frac{k}{2} + \frac{n}{4} + \frac{1}{4}\frac{(2k+1-n)10^{-3k}|z|^{2k+1-n}}{1 + 10^{-3k}|z|^{2k+1-n}} + \frac{2}{1+|z|^2/8} \le -\frac{1}{2} ,$$

where the last inequality is equivalent to

$$(1 + |z|^2/8)(2k + 2 - n + 10^{-3k}|z|^{2k+1-n}) - 8(1 + 10^{-3k}|z|^{2k+1-n}) \ge 0 ,$$

which can be implied by an AM-GM inequality via

$$\begin{aligned}&3n + \frac{10^{-3k}}{8}|z|^{2k+3-n} \ge (2k+3-n)(\frac{3n}{2})^{\frac{2}{2k+3-n}}(\frac{10^{-3k}}{8(2k+1-n)})^{1-\frac{2}{2k+3-n}}|z|^{2k+1-n} \\ &> 10^{-3k}/8(10^{6k/(2k+3-n)})|z|^{2k+1-n} > 8 \times 10^{-3k}|z|^{2k+1-n} .\end{aligned}$$

We collect the linear estimate by choosing a small enough ν to get

$$(\nabla^k L, \nabla^k u \rho_k) \le (-\frac{1}{4} + C(1 + \|\nabla^4 u\|_\infty)\lambda) E_k^2 + C E_0^2 .$$

For the nonlinear term $\nabla^k N$, by Leibniz's rule, we know that it will be a linear combination of $\nabla^{k-j} u \nabla^j u$. For a typical term, assume WLOG that $j \le k/2$. By the interpolation lemma, we have

$$\|\nabla^{k-j} u \nabla^j u\|_{\rho_k} \le D_{k-j} \|\nabla^j u \langle z \rangle^j\|_\infty \lesssim D_{k-j} D_{j+n} \lesssim (D_k + D_0)^2 .$$

Therefore, we can collect the nonlinear estimate

$$(\nabla^k N, \nabla^k u \rho_k) \le C(E_k + E_0)^2 E_k .$$

For the error term, notice that $\nabla^k F_i$ is $O(z^{-2-k})$ at ∞. Therefore, it is square integrable with the weight ρ_k and we can estimate

$$(\nabla^k F_i, \nabla^k u \rho_k) \le C\lambda(1 + \|\nabla^4 u\|_\infty) E_k .$$

We estimate the viscous term using integration by parts twice

$$(\nabla^k V_i, \nabla^k u\rho_k) = -(\nabla^k u_i, \nabla^k u_i\rho_k) - (\nabla^k u_i, \nabla^k u(\rho_k)_i) \le \frac{1}{2}(\nabla^k u, \nabla^k u(\rho_k)_{ii}) \le CE_k^2 .$$

Finally, we gather our H^k estimate as

$$\begin{aligned} \frac{1}{2}\partial_\tau E_k^2 &\le (-\frac{1}{4} + C(1 + \|\nabla^4 u\|_\infty)\lambda)E_k^2 + CE_0^2 + C(E_k + E_0)^2 E_k \\ &+ C\lambda(1 + \|\nabla^4 u\|_\infty)E_k + C\lambda E_k^2 . \end{aligned} \tag{2.3.4}$$

Using again the interpolation lemma, we know that $\|\nabla^4 u\|_\infty + \|u\|_\infty \lesssim E_k + E_0$. Combined with (2.3.2), we know that there exists a constant μ_0, such that for $0 < \mu < \mu_0$, if we consider the energy

$$E^2 = E_0^2 + \mu E_k^2 , \tag{2.3.5}$$

we have the estimate

$$\frac{1}{2}\partial_\tau E^2 \le (-\frac{1}{10} + C(1 + E)\lambda)E^2 + CE^3 + C\lambda(1 + E)E + C\lambda E^2 .$$

Namely that

$$\partial_\tau E \le (-\frac{1}{10} + CE\lambda + CE)E + C\lambda + C\lambda E . \tag{2.3.6}$$

Notice that here $C \ge 1$ is an absolute constant.

2.3.3 Finite time blowup

Recall the ODE (2.2.8) for λ_i, and noticing that $\lambda = \max \lambda_i$, we define $\gamma = 1/\lambda$. $\gamma(0) = 1/\hat{C}_u(0)$ will be the constant we choose now. The ODE for γ is

$$\partial_\tau \gamma = -\frac{\partial_\tau \lambda_i}{\lambda_i^2} = 1 + 4\sum_j u_{iijj}(0)\frac{\lambda_j}{\lambda_i} , \quad i = \text{argmax}\lambda_i .$$

a priori estimate yields $|\nabla^4 u(0)| \le CE$. We can assume WLOG that $C \ge 1$. Defining $G = E\gamma$, we have

$$|\partial_\tau \gamma - 1| \le 4nC\frac{G}{\gamma} . \tag{2.3.7}$$

We will show that E decays as $1/\tau$. We calculate an ODE for G:

$$\begin{aligned} \partial_\tau G &\le (-\frac{1}{10} + CE\lambda + CE)G + C + CE + E(1 + 4nC\frac{G}{\gamma}) \\ &\le (-\frac{1}{10} + 2C\frac{1}{\gamma})G + C + 8nCG^2(\frac{1}{\gamma} + \frac{1}{\gamma^2}) . \end{aligned} \tag{2.3.8}$$

We choose $\hat{C}_u(0) = 1/\gamma(0) \leq 1/(10000nC^2)$ small enough such that if we start from $G(0) < 100C$, we will have the bootstrap estimate

$$G < 100C, \quad \gamma \geq \gamma(0) \geq 10000nC^2 \tag{2.3.9}$$

for all time via a standard bootstrap argument.

Proof of the bootstrap bound (2.3.9). In fact, we know by (2.3.7) that $\partial_\tau \gamma(0) > 0$. Assume that the bootstrap estimate is false, then by continuity, there exists a rescaled time $\tau_0 > 0$ such that (2.3.9) holds for $0 < \tau < \tau_0$ and either $G(\tau_0) \geq 100C$ or $\gamma(\tau_0) \leq \gamma(0)$. We compute by (2.3.7) that in $(0, \tau_0)$, $\partial_\tau \gamma \geq 1 - \frac{400nC^2}{10000nC^2} > 0$ which rules out the latter case. As a consequence, we estimate the ODE for G (2.3.8) in $(0, \tau_0)$ as

$$\partial_\tau G \leq -\frac{1}{20}G + C + G\frac{1}{50}\,.$$

By continuity, we estimate $\partial_\tau G(\tau_0) < 0$ and we conclude that the former case cannot hold either. We reach a contradiction and conclude the bootstrap. □

Based on the bootstrap estimates, we have the following estimate for γ:

$$|\partial_\tau \gamma - 1| \leq \frac{400nC^2}{\gamma}\,.$$

Thus $\gamma/\tau \to 1$ as $\tau \to \infty$. Thus we have

$$|(\hat{C}_u)_t + 1| = |\frac{(\hat{C}_u)_t t_\tau}{\hat{C}_u} + 1| = |\frac{(\hat{C}_u)_\tau}{\hat{C}_u} + 1| = |\hat{c}_u + 1| = |c_u| \leq \frac{n}{4\gamma}\,, \quad \tau_t = 1/\hat{C}_u\,.$$

We can finally show that there exists a blowup time $T > 0$, such that

$$\lim_{t\to T} \frac{\hat{C}_u}{T-t} = 1\,, \quad \lim_{t\to T} \frac{\tau}{|\log(T-t)|} = 1\,.$$

Moreover, defining $\kappa = \sum_i \frac{1}{\lambda_i}$, we compute the following ODE

$$\partial_\tau \kappa = n + 4\sum_i \sum_j u_{iijj}(0)\frac{\lambda_j}{\lambda_i} \leq n + 4nC\frac{G}{\gamma^2}\kappa \leq n + \frac{400nC^2}{\gamma^2}\kappa\,.$$

Therefore for sufficiently large τ, $\partial_\tau \kappa \leq n + 800nC^2\frac{\kappa}{\tau^2}$. We integrate and get for sufficiently large τ, $\kappa \leq n\tau + 1600n^2C^2\log\tau$. Therefore since $\lambda = \max\lambda_i = 1/\gamma \to 1/\tau$, we have $n \leq \liminf \sum_i \frac{\lambda}{\lambda_i} \leq \limsup \sum_i \frac{\lambda}{\lambda_i} \leq n$. Namely we get $\lambda_i\tau \to \frac{\lambda_i}{\lambda} \to 1$, and thus we arrive at the law

$$\lim_{t\to T} \frac{\hat{C}^i_l}{\sqrt{(T-t)|\log(T-t)|}} = \lim_{t\to T}\sqrt{\frac{\hat{C}_u}{(T-t)\lambda_i\tau}} = 1\,.$$

Therefore for $C_0 = 100C, \lambda_0 = 1/(10000nC^2)$ and for initial data satisfying the assumption of Theorem 2.1.1, we know that $\gamma(0) = 1/\lambda$ and $u = g$ defined in the rescaled formulation satisfy the bootstrap assumption. We conclude Theorem 2.1.1 based on the asymptotics of $\hat{C}_u, \hat{C}_l^i$.

2.4 Numerical Experiments

In this section, we conduct numerical experiments to corroborate our analysis that our choice of normalization in Section 2.2 indeed preserves a stable blowup and therefore we are able to capture the log-correction numerically, both in the 1D case and in the case of higher dimensions with nonradial perturbations. We remark that our proofs in the work are derived independently of the numerical results in this section.

Data availability statement: The data and the code will be available upon request.

In practice, we hope to compute the profile even if we do not have prior knowledge. Therefore in our numerical experiment, we solve (2.2.5) with initial data as a large perturbation to the approximate steady state. We will compute $\hat{u}$ dynamically and recall that our choice of normalization $\hat{c}_l$, $\hat{c}_u$ in (2.2.1) ensures that $\hat{u}(0)$, $\hat{u}_{zz}(0)$ remain constants in time.

2.4.1 1D case

In our numerical study, we choose the initialization that is more general than the assumption of our theorem as

$$\hat{u}(0,z) = (1 + z^2/8 + z^4/10)^{-1}, \quad \hat{C}_u(0) = 1, \quad \lambda = 1.$$

At each time step τ_m, we first determine the normalization constants as

$$\hat{c}_u = -\hat{u}(0) - \frac{\lambda \hat{u}_{zz}(0)}{\hat{u}(0)}, \hat{c}_l = \frac{\hat{c}_u}{2} + \hat{u}(0) + \frac{\lambda \hat{u}_{zzzz}(0)}{2\hat{u}_{zz}(0)}.$$

Next, we can determine the time step k via the standard numerical stability conditions for a convection-diffusion equation, and then we use the 4-th order Runge-Kutta scheme for the discretization in time and a cubic spline for the discretization in space to evolve the equation

$$\hat{u}_\tau = \hat{c}_u \hat{u} - \hat{c}_l z \hat{u}_z + \hat{u}^2 + \lambda \hat{u}_{zz}.$$

Finally, we update our λ for time $\tau_{m+1} = \tau_m + k$ by a 4-th order Runge-Kutta discretization scheme of the ODE

$$(\log \lambda)_\tau = (2\hat{c}_l + \hat{c}_u).$$

We use a fixed nonuniform mesh in space with even symmetry considered, and our computational domain is $[0, 10^5]$ with 2000 gridpoints in space. We report that after 10^9 iterations in time, the rescaled time $\tau \approx 3.9887 \times 10^5$ and $\log(\hat{C}_u) \approx -3.9886 \times 10^5$. This means that the amplitude of the solution in the physical space grows $\exp(3.9886 \times 10^5)$ times, which is impossible to compute if we do not use a dynamic rescaling formulation. We remark that the computation is very stable and we stopped after 10^9 iterations only due to concerns of computational time. In theory, we can compute for an arbitrarily long time and witness an arbitrary growth of the amplitude in the physical space.

To see that the profile $\hat{u}$ converges indeed to the steady state $\bar{u}$, we plot the profile after $m = 5 \times 10^4, 5 \times 10^5, 5 \times 10^6, 10^7, 1.5 \times 10^7, 2 \times 10^7$ iterations and compare it with the steady state. We see that the profile converges fast; see Figure 2.1. Furthermore, we investigate the convergence rate of the profile. Define $\gamma(\tau) = \sup_z\{\hat{u}(\tau) - \bar{u}\}$. We plot $\gamma\tau$ after 2×10^7 until 5×10^7 iterations, corresponding to $\tau \in [218, 11638]$. We see that the residue is approximately of order $1/\tau$; see Figure 2.2. However, we are only using a finite domain and as time becomes larger, the effect of the finite domain size becomes more obvious, and $\gamma\tau$ will increase slightly.

To see that we can recover the correct convergence rate, we plot $(1/2 - \hat{c}_l)\tau$ and $(\hat{c}_u + 1)\tau$ in time to see that they indeed converge to the correct constant $5/8$ and $1/4$ respectively and therefore will give the correct log-scaling; see for example indicated by (2.2.2). Again for visualization purposes, we only plot for the first 5×10^7 iterations and we can see that they converge to the desired constants very fast; see Figure 2.3.

2.4.2 2D case

For the 2D example, we choose a nonradial initialization as

$$\hat{u}(0, x, y) = (1 + (x^2 + y^2)/8 + x^4/100)^{-1}, \quad \hat{C}_u(0) = 1, \quad \lambda_1 = \lambda_2 = 1.$$

At each time step τ_m, we first determine the normalization constants as

$$\hat{c}_u = -\hat{u}(0,0) - \frac{\lambda_1 \hat{u}_{xx}(0,0) + \lambda_2 \hat{u}_{yy}(0,0)}{\hat{u}(0,0)},$$

$$\hat{c}_l^1 = \frac{\hat{c}_u}{2} + \hat{u}(0,0) + \frac{\lambda_1 \hat{u}_{xxxx}(0,0) + \lambda_2 \hat{u}_{xxyy}(0,0)}{2\hat{u}_{xx}(0,0)},$$

$$\hat{c}_l^2 = \frac{\hat{c}_u}{2} + \hat{u}(0,0) + \frac{\lambda_1 \hat{u}_{xxyy}(0,0) + \lambda_2 \hat{u}_{yyyy}(0,0)}{2\hat{u}_{yy}(0,0)}.$$

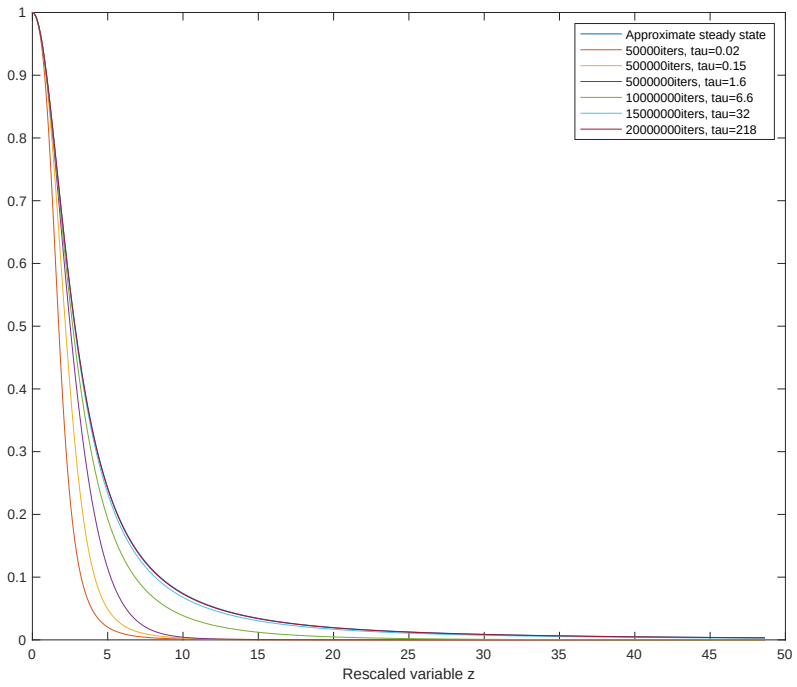


Figure 2.1: Comparison of the profile to the approximate steady state

Next, we can determine the time step k via the standard numerical stability conditions for a convection-diffusion equation, and then we use the 4-th order Runge-Kutta scheme for the discretization in time and a cubic spline for the discretization in space to evolve the equation

$$\hat{u}_\tau = \hat{c}_u\hat{u} - \hat{c}_l^1 x\hat{u}_x - \hat{c}_l^2 y\hat{u}_y + \hat{u}^2 + \lambda_1\hat{u}_{xx} + \lambda_2\hat{u}_{yy}\,.$$

Finally, we update our λ_1, λ_2 for time $\tau_{m+1} = \tau_m + k$ by a 4-th order Runge-Kutta discretization scheme of the ODE

$$(\log\lambda_i)_\tau = (2\hat{c}_l^i + \hat{c}_u)\,, \quad i = 1, 2\,.$$

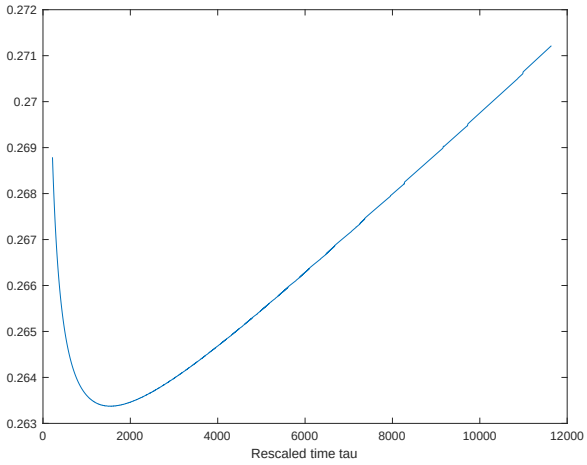


Figure 2.2: Plot of the residue multiplied by the rescaled time

We use a fixed nonuniform mesh in space with even symmetry considered, and our computational domain is $[0, 4000]$ with 200 gridpoints in space in each direction. To see that we can recover the correct convergence rate, we plot $R_i := (1/2 - \hat{c}_l^i)\tau$ and $R_u := (\hat{c}_u + 1)\tau$ as a function of τ after 10^7 iterations to see that they indeed converge to the correct constant $3/4$ and $1/2$ respectively and therefore will give the correct log-scaling; see for example indicated by (2.2.7). We can see that they converge to the desired constants very fast; see Figure 2.4.

2.5 Finite Codimensional stability

In this section, we will briefly sketch the high-level idea to study high-order vanishing type-I blowup for the 1D semilinear heat equation (2.1.1) under *radial (even symmetric)* setting.

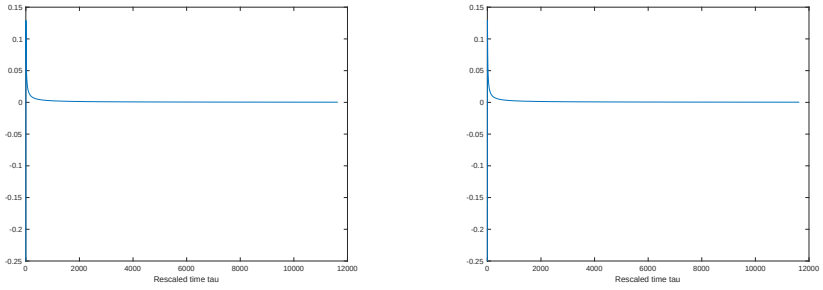


Figure 2.3: Fitting the law of the normalization constants. Left: $(1/2 - \hat{c}_l)\tau - 5/8$ versus τ; right: $(\hat{c}_u + 1)\tau - 1/4$ versus τ.

2.5.1 Self-similar renormalization and the approximate solution

For any fixed $m \in \mathbb{Z}_{>1}$, we introduce the self-similar coordinate

$$y = \frac{x}{\lambda^{\frac{1}{m}}}, \quad \frac{d\tau}{dt} = \frac{1}{\lambda^2}, \quad \tau\Big|_{t=0} = 0, \quad \frac{\lambda_\tau}{\lambda} = -\frac{1}{2}, \tag{2.5.1}$$

and corresponding renormalization

$$a(t,x) = \frac{1}{\lambda^2} U(\tau, y), \tag{2.5.2}$$

then $\lambda(\tau) = \lambda_0 e^{-\frac{1}{2}\tau}$ and U solves the equation

$$\partial_\tau U = \lambda^{2-\frac{2}{m}} U_{yy} - U - \frac{1}{2m} y \cdot \nabla U + U^2, \tag{2.5.3}$$

where the diffusion term can be regarded as a perturbation since $2 - \frac{2}{m} > 0$. This in turn motivates us to find an approximate solution solving the equation

$$-U - \frac{1}{2m} y \cdot \nabla U + U^2 = 0. \tag{2.5.4}$$

In particular, this equation can be explicitly solved by

$$U_*(y) = (1 + c y^{2m})^{-1}. \tag{2.5.5}$$

Here $c > 0$ is a constant, and $m > 1$ describes the vanishing order of the next order expansion of U_* near the origin.

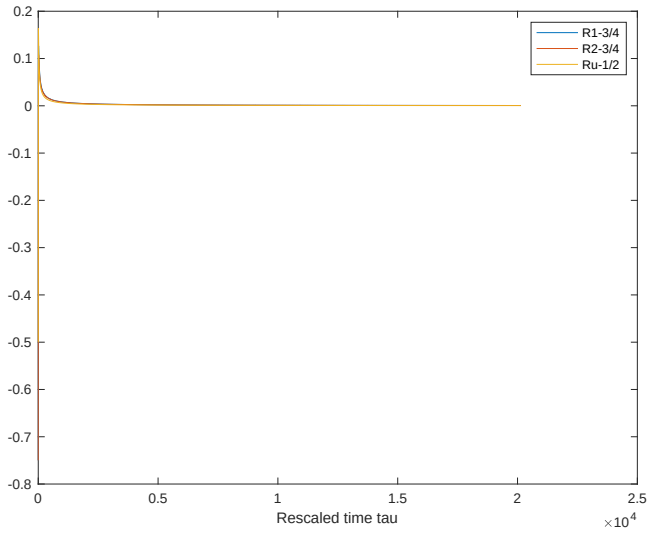


Figure 2.4: Fitting the law of the normalization constants for 2D.

2.5.2 Linear stability

We fix $m > 1$ and $c = 1$ in (2.5.5), and plug the ansatz $U = U_* + \epsilon$ into (2.5.3), it then follows that ε solves

$$\partial_\tau \varepsilon = \lambda^{2-\frac{2}{m}} U_{yy} + \mathcal{L}\varepsilon + \varepsilon^2,$$

where the linearized operator reads

$$\mathcal{L}\varepsilon = -\varepsilon - \frac{1}{2m} y \partial_y \varepsilon + 2U_* \varepsilon. \tag{2.5.6}$$

Next, we introduce a weighted L^2_Θ space with singular weight $\Theta(y) = y^{-4m-4}$ near the origin to extract damping, which is the essential step to close the nonlinear

stability. Via the integration by parts, we have the coercivity near the origin

$$(\mathcal{L}\varepsilon,\varepsilon)_{L^2_\Theta}=\left(\left(-1+2U_*+\frac{(\Theta y)_y}{4m\Theta}\right)\varepsilon,\varepsilon\right)_{L^2_\Theta}\approx-\frac{3}{4m}(\varepsilon,\varepsilon)_{L^2_\Theta}. \tag{2.5.7}$$

In particular, with careful analysis, there is a small constant $0<\kappa=\kappa(m,U_*)\ll 1$ such that (2.5.7) can be extended to

$$(\mathcal{L}\varepsilon,\varepsilon)_{L^2_{\Theta+\kappa}}\le-\frac{1}{4m}(\varepsilon,\varepsilon)_{L^2_{\Theta+\kappa}}. \tag{2.5.8}$$

Additionally, we need to introduce the higher Sobolev norm $\dot{H}^{\bar{K}}$ to close the bootstrap argument. Precisely,

$$\begin{aligned}(\mathcal{L}\varepsilon,\varepsilon)_{\dot{H}^{\bar{K}}}&=\left(\left(-1-\frac{\bar{K}}{2m}+2U_*+\frac{1}{4m}\right)\partial_y^{\bar{K}}\varepsilon,\partial_y^{\bar{K}}\varepsilon\right)_{L^2}+O(\|\varepsilon\|_{H^{\bar{K}-1}}\|\varepsilon\|_{\dot{H}^{\bar{K}}})\\&\le-\frac{2\bar{K}-4m-1}{4m}\|\varepsilon\|^2_{\dot{H}^{\bar{K}}}+O(\|\varepsilon\|_{H^{\bar{K}-1}}\|\varepsilon\|_{\dot{H}^{\bar{K}}}),\end{aligned} \tag{2.5.9}$$

where the leading order enjoys damping once we choose $\bar{K}=\bar{K}(m)\gg 1$.

2.5.3 Modulation ODEs and nonlinear stability

With the singular weight $\Theta(y)=|y|^{-4m-4}$ given previously, introducing a cutoff function χ, we can further *radially* decompose ε into

$$\varepsilon(\tau,y)=\varepsilon_u(\tau,y)+\varepsilon_s(\tau,y),\qquad\text{with}\qquad\varepsilon_u=\sum_{j=0}^{m}c_j(\tau)\chi(y)y^{2j}, \tag{2.5.10}$$

such that $\varepsilon_s(\tau,y)=O(y^{2m+2})$ near the origin, which yields an ODE system for modulation parameters $\{c_j\}_{j=0}^m$:

$$\begin{cases}\dot{c}_j=\left(1-\frac{j}{m}\right)c_j+[\varepsilon_u^2]_j+\lambda^{2-\frac{2}{m}}[U_{yy}]_j, & 0\le j<m-1,\\ \dot{c}_m=[\varepsilon_u^2]_m-2c_0+\lambda^{2-\frac{2}{m}}[U_{yy}]_m, & j=m.\end{cases} \tag{2.5.11}$$

Here $[h]_j$ is the $2j$-th order coefficient of Taylor expansion of $h(r)$ at the origin. Additionally, $\varepsilon_s=O(y^{2m+2})$ solves the equation

$$\partial_\tau\varepsilon_s=\mathcal{L}\varepsilon_s+2\varepsilon_u\varepsilon_s+\varepsilon_s+G[\lambda,U,\varepsilon_u],$$

with the modulation term $G[\lambda,U,\varepsilon_u]=O(y^{2m+2})$ given by

$$G[\lambda,U,\varepsilon_u]=\left(\lambda^{2-\frac{2}{m}}U_{yy}+\mathcal{L}\varepsilon_u+\varepsilon_u^2\right)-\sum_{j=0}^{K}\left[\lambda^{2-\frac{2}{m}}U_{yy}+\mathcal{L}\varepsilon_u+\varepsilon_u^2\right]_j\chi y^{2j}.$$

Finally, we can use the standard topological argument together with (2.5.7) and (2.5.9) to derive the nonlinear stability with finite codimension $m+1$.

Remark 2.5.1. *We expect that this nonlinear stability result can be improved to finite-codimension $m-1$, which is two dimensions lower than our previous findings. The key underlying reason is the presence of two degrees of freedom, namely, the choice of the blowup time $T>0$ and the shrinking rate. These degrees of freedom can be utilized through a matching argument to recover the corresponding unstable directions as in [273, 275].*

Alternatively, one may employ a method of dynamical rescaling to establish stability with finite codimension $m-1$. Specifically, we modify the coordinate (2.5.1) as

$$y=\frac{x}{\mu^{\frac{1}{m}}},\quad \frac{d\tau}{dt}=\frac{1}{\lambda^2},\quad \tau\big|_{t=0}=0,\quad \frac{\lambda_\tau}{\lambda}=-\frac{1}{2}+\frac{1}{2}c_a,\quad \frac{\mu_\tau}{\mu}=-\frac{1}{2}-mc_s.$$

We then define the corresponding renormalization

$$u(t,x)=\frac{1}{\lambda^2}U(\tau,y).$$

Under this coordinate transformation, U satisfies

$$\partial_\tau U=\lambda^2\mu^{-\frac{2}{m}}U_{yy}+(-1+c_a)\,U-\left(\frac{1}{2m}+c_s\right)y\partial_y U+U^2,$$

where the parameters (c_a,c_s) are determined by the modulation conditions

$$U(\tau,0)=U_*(0)\qquad \text{and}\qquad [U(\tau,0)]_m=[U_*]_m,$$

and we eliminate neutral modes by fixing $c_0=c_m=0$. By applying a similar argument of modulation ODEs, we obtain the nonlinear stability with finite codimension $m-1$. Notably, introducing extra scaling parameters to perturb the scaling symmetry is crucial for extending the argument to the nonradial setting; see previous works by the second author and collaborators [77, 218].

Remark 2.5.2. *Compared with the semilinear heat equation, analyzing the Keller-Segel equation with logistic damping involves several additional challenges. For example, the profile U_* introduced in (2.5.5) is an explicit solution to the first-order and separable local equation (2.5.4). In contrast, for Keller-Segel equation with logistic damping, the associated profile equatiois inherently nonlocal and cannot be trivially solved. This nonlocality requires a more delicate analysis.*

Moreover, since there is no explicit nontrivial solution to the profile equation, additional effort is required to derive quantitative properties of the profile Q. Combined with the nonlocal nature, these complexities make the establishment of linear coercivity of Keller-Segel equation more intricate than in the case of the semilinear heat equation. Detailed strategies to handle these obstacles will be presented in Section 4 of [286].

2.6 Future works

Having established our robust framework of blowups via local modulations and singularly weighted estimates, we outline some potential directions for generalizing our methods for more challenging singularities with multiple scales, and towards investigation as promising numerical methods.

An intriguing blowup phenomenon happens with a traveling-wave type singularity, say of the following

$$\bar{w}(x,t) = (T-t)^{-3/2}\bar{\Omega}(\frac{x - 1/2(T-t)^{1/2}}{(T-t)}) .$$

One goal is to generalize the idea of local modulations to traveling-wave type singularities, where there are two scales at play: the inner scale that determines the blowup profile at $x = 1/2(T-t)^{1/2} + O(T-t)$, and the outer scale that determines the blowup location at $x = O((T-t)^{1/2})$. We have demonstrated that via arguments of local modulations, the outer scale is shown to be stable, leveraging the precise vanishing order of the profile and making the perturbation vanish at a higher order.

By studying this simple model, one can gain insights to establish singularity for more complicated models like Keller-Segel or potentially Navier-Stokes equations, which involve multiple blowup scales. Again, our framework of local modulations comes at the benefit of using only limited non-explicit information of the approximate profile, seamlessly amenable to computations of profiles and computer-assisted proofs.

On the other hand, numerical investigations showcased the efficiency of our method in the computation of the semilinear heat equation [218]. As indicated by our analysis, we enforced $\hat{u}(0,t)$ and $\nabla^2\hat{u}(0,t)$ to be constant for data with even symmetries. In practice, one can come up with more stable enforcement of those vanishing conditions, say with a splitting method to enforce those constant quantities, potentially improving the numerical stability of our method. Implementation of the numerical methods for general classes of blowups, including those with multiple scales, is of immediate future interest.

For general nonlinear steady-state equations, one approach is to introduce artificial time and perform time marching for convergence. Since we demonstrated that a clever nonlinear rescaling can rule out the potentially unstable directions and enhance convergence, we have reasons to believe that our methodology can be adopted broadly as numerical solvers beyond singularity formation.

Chapter 3

BLOWUPS VIA LOCAL MODULATIONS: COMPLEX GINZBURG-LANDAU

Building upon the idea in [218], we establish the stability of the type-I[1] blowup with log correction for the complex Ginzburg-Landau equation. In the amplitude-phase representation, a generalized dynamic rescaling formulation is introduced, with modulation parameters capturing the spatial translation and rotation symmetries of the equation and novel anisotropic modulation parameters perturbing the scaling symmetry. This new formulation provides enough degrees of freedom to impose normalization conditions on the rescaled solution, completely eliminating the unstable and neutrally stable modes of the linearized operator around the blowup profile. It enables us to establish the full stability of the blowup by enforcing vanishing conditions via the choice of normalization and using weighted energy estimates, for a non-variational problem. No topological argument or spectrum analysis is needed, opening up the possibility to tackle a wide range of type-I singularities. The log correction for the blowup rate is automatically inferred via the local normalization conditions, captured by the energy estimates and refined estimates of the modulation parameters.

3.1 Introduction

We consider the complex Ginzburg-Landau equation

$$\psi_t = (1+\imath\beta)\Delta\psi + (1+\imath\delta)|\psi|^{p-1}\psi - \gamma\psi, \tag{CGL}$$

where $\psi(t) : \mathbb{R}^d \to \mathbb{C}$, β, δ, γ are real constants and $p > 1$. The model equation (CGL) was first derived by Stewardson and Stuart in [424] (see also [109], [120]) to examine afresh the problem of plane Poiseuille flow in a wave system. The equation is also used to describe various phenomena in many fields, among which are nonlinear optics with dissipation [329], turbulent behavior [41], Rayleigh-Bénard convection or Taylor-Couette flow in hydrodynamics [122], [344], [343], reaction-diffusion systems [186], [229], [399], [404], the theory of superconductivity [30], [66], [125], [171], etc. For further details on the physical background and derivation

[1] A blowup solution is of Type I if it satisfies the bound $\lim_{t\to T}(T-t)^{-\frac{1}{p-1}}\|u(t)\|_\infty < \infty$, otherwise, blowup is of Type II.

of the complex Ginzburg-Landau equation, we refer to the surveys [9], [328], and the references therein.

The local Cauchy problem has been well established through a semigroup approach in the works [169, 168, 170]. A solution to (CGL) blows up in finite time if $\lim_{t\to T} \|\psi(t)\|_{L^\infty(\mathbb{R}^d)} = +\infty$ for some $T < +\infty$. Singularity formation has been intensively studied for the two limiting models of (CGL): the classical nonlinear heat equation in the limit $\beta, \delta, \gamma \to 0$,

$$\partial_t \psi = \Delta\psi + |\psi|^{p-1}\psi, \quad \psi(t) : x \in \mathbb{R}^d \to \mathbb{R}, \tag{NLH}$$

and the nonlinear Schrödinger equation in the limit $\beta, |\delta| \to \infty$,

$$\imath\partial_t \psi + \Delta\psi + \mu|\psi|^{p-1}\psi = 0, \quad \mu = \pm 1. \tag{NLS}$$

We refer to [380] and [147] for intensive lists of references from the early 1960s concerning blowup results of these two equations. However, singularities in (CGL) (collapse, chaotic, or blowup) are much less understood in comparison with what have been established for (NLH) and (NLS). The study of singularity in (CGL) is a challenging problem due to the lack of variational structure, no maximum principle, non-self-adjoint linearized operator, etc. Nevertheless, singularity in (CGL) was experimentally reported in [253], [254] where the authors described an extensive series of experiments on traveling-wave convection in an ethanol/water mixture and collapse solutions were observed. We have a sharp sufficient criterion for collapse in (CGL) for the case of subcritical bifurcation described in [434]. In [148], the authors used the modulation theory and numerical observations to show that the collapse dynamic is governed in the (CGL) limit of the L^2-critical cubic (NLS). For the existence of blowup, there are the results of [63] and [64] in which the authors studied (CGL) for the case $\beta = \delta$. In [50, 395], [373], the authors gave some evidence for the existence of a radial solution that blows up in a self-similar way, their arguments were based on the combination of rigorous analysis and numerical computations. In [470] and [310], the authors rigorously constructed particular examples of initial data for which the solutions of (CGL) blow up in finite time for (β, δ) in the *subcritical* range

$$\mathfrak{b}_* := p - \delta^2 - \beta\delta(p+1), \quad \mathfrak{b}_* > 0 \quad \text{(subcritical range)}. \tag{3.1.1}$$

The constructed blowup solution in the subcritical case admits the asymptotic be-

havior

$$\psi(x,t) \sim |\log(T-t)|^{i\mu}\Big[(T-t)\big(p-1+c_p|Z|^2\big)\Big]^{-\frac{1+i\delta}{p-1}}, \quad Z = \frac{x}{\sqrt{(T-t)|\log(T-t)|}}, \tag{3.1.2}$$

where the constants c_p and μ are given by

$$c_p = \frac{(p-1)^2}{4\mathfrak{b}_*} > 0\,, \quad \mu = -\frac{\beta(1+\delta^2)}{2\mathfrak{b}_*}\,. \tag{3.1.3}$$

The spectral analysis for a non-self-adjoint operator developed in [310] can be implemented for other problems where an energy-type method is not applicable, see for example [161]. The blowup for the critical range, $\mathfrak{b}_* = 0$, was solved in [348], [130] following the approach of [310]. The blowup for (CGL) in the supercritical range, $\mathfrak{b}_* < 0$, has recently been solved in [131] for the special choice $\beta = 0$. We remark that in the mentioned works ([470], [310], [348], [130], [131]), the authors focused on the case of dimension $d = 1$, and briefly described the stability properties of constructed blowup solutions through a spectral approach in a restricted (well-prepared) class of initial data. One can go further in the spectral analysis to establish type II singularities, for example, the energy supercritical case for (NLH) in [196] and for (NLS) in [322].

In this chapter, we aim to develop a new approach based on the dynamical rescaling formulation and the energy method to study blowup solutions to (CGL). Compared to the aforementioned works that employ radial scalings to the spatial variable corresponding to the scaling symmetry, our approach has the following key novelties. We introduce novel *nonradial, anisotropic* scalings to the spatial variable that breaks the scaling symmetry and incorporates it, along with modulation parameters capturing the symmetries of the equations, to develop a novel generalized *anisotropic* rescaling in Step 1 in Subsection 3.1.2. This new approach allows us to use a purely energy-based method to establish asymptotically self-similar blowup in (CGL) and a clear notion of stability capturing the logarithm correction (3.1.2) in the subcritical case for a large class of initial data in all dimensions $d \geq 1$. Throughout this chapter, we use the amplitude-phase representation,

$$\psi(x,t) = u(x,t)e^{i\theta(x,t)}, \tag{3.1.4}$$

where u and θ are real-valued functions of time and space solving the coupled system

$$\partial_t u = \big[\Delta - |\nabla\theta|^2\big]u - \beta\big(2\nabla u\cdot\nabla\theta + u\Delta\theta\big) + u^p - \gamma u, \tag{3.1.5a}$$

$$u\partial_t\theta = \beta\big[\Delta - |\nabla\theta|^2\big]u + 2\nabla u\cdot\nabla\theta + u\Delta\theta + \delta u^p. \tag{3.1.5b}$$

The case $\beta = 0$ is related to a class of reaction-diffusion equations appearing in the study of pattern formation, see for example [186] and references therein.

3.1.1 Main result

For any $k \geq 1$, we introduce the functional spaces $\mathfrak{E}_k$ and $\mathfrak{F}_k$

$$\mathfrak{E}_k = \Big\{ w : \ \|w\|_{\mathfrak{E}_k} = \sum_{j=0}^{k} \|\nabla^j w\|_{\rho_j} < +\infty \Big\}, \quad \mathfrak{F}_k = \Big\{ \phi : \ \|\phi\|_{\mathfrak{F}_k} = \sum_{j=1}^{k} \|\nabla^j \phi\|_{\mathring{\rho}_j} < +\infty \Big\}, \tag{3.1.6}$$

where $\|\cdot\|_{\rho_k}$ and $\|\cdot\|_{\mathring{\rho}_k}$ stand for the standard weighted L^2-norm with ρ_k and $\mathring{\rho}_k$ being defined as in (3.1.28). Let $\bar{U}$ be the universal profile

$$\bar{U}(z) = \Big(p - 1 + \frac{(p-1)^2}{4\mathfrak{b}_*} |z|^2 \Big)^{-\frac{1}{p-1}}, \quad \forall\, z \in \mathbb{R}^d, \tag{3.1.7}$$

and V_0 be a non-degenerate global maximizer of u_0 defined by

$$V_0 = \arg\max u_0(z), \quad u_0(V_0) > 0, \quad -\nabla^2 u_0(V_0) \succ 0, \tag{3.1.8}$$

where $A \succ 0$ means that A is a positive definite matrix. The main result of this chapter is the following theorem.

Theorem 3.1.1 (Existence and stability of blowup solutions to (CGL))**.** *Consider β, δ in the sub-critical range (3.1.1), i.e. $\mathfrak{b}_* > 0$, $p > 1$ and $d \geq 1$. Let $K = K(d,p) \in \mathbb{N}$ be defined as in (3.1.29). There exists an open set $\mathcal{O} \subset \mathfrak{E}_K \times \mathfrak{F}_K$ of initial data $\psi_0 = u_0 e^{\iota\theta_0}$ with the property (3.1.8) such that the corresponding solution $\psi = u e^{\iota\theta}$ to (CGL) blows up in finite time T and the following asymptotic behaviors hold.*

(i) (The amplitude-phase decomposition)

$$\Big\| H(t) u\big(\mathbf{R}(t) z + V(t), t\big) - \bar{U}(z) \Big\|_{\mathfrak{E}_K} + \Big\| \theta\big(\mathbf{R}(t) z + V(t), t\big) - \mu(t) - \delta \log \bar{U}(z) \Big\|_{\mathfrak{F}_K} \leq \frac{C}{1 + |\log(T-t)|}, \tag{3.1.9}$$

where $H(t)$ and $\mu(t)$ are scalar functions, $\mathbf{R}(t)$ is an upper triangular matrix and $V(t)$ is a vector in $\mathbb{R}^d$,

$$\lim_{t\to T} \frac{H(t)^{p-1}}{T-t} = 1, \quad \lim_{t\to T} \frac{\mathbf{R}(t)}{\sqrt{(T-t)|\log(T-t)|}} = \mathrm{I}_d, \quad \lim_{t\to T} V(t) = V_T, \tag{3.1.10}$$

for some $V_T \in \mathbb{R}^d$, and $\mu(t)$ admits the expansion[2]

$$\mu(t) = -\frac{\delta}{p-1}\log(T-t) - \frac{d\beta(1+\delta^2)}{2\mathfrak{b}_*}\log|\log(T-t)| + \hat{\mu}(t), \quad \lim_{t\to T}\hat{\mu}(t) = \hat{\mu}_T, \tag{3.1.11}$$

for some scalar function $\hat{\mu}(t)$ and $\hat{\mu}_T \in \mathbb{R}$.

(ii) (L^∞ asymptotic behavior)

$$\left\| |\log(T-t)|^{\iota\frac{d\beta(1+\delta^2)}{2\mathfrak{b}_*}} (T-t)^{\frac{1+\iota\delta}{p-1}} e^{-\iota\hat{\mu}(t)}\, \psi(\mathbf{R}(t)z+V(t),t) - \bar{U}^{1+\iota\delta} \right\|_{L^\infty} \le \frac{C}{1+|\log(T-t)|^{\sigma'}}, \tag{3.1.12}$$

where $\sigma' = \min\left\{1, \frac{4}{p-1}\right\}$ and $C = C(u_0,\theta_0) > 0$.

Remark 3.1.2 (Description of the set $\mathcal{O}$ of initial data)**.** *For initial data u_0 satisfying the property (3.1.8), we can define an upper triangular matrix $\mathcal{M}_0$ with $\mathcal{M}_{0,ii} > 0$*[3] *and the rescaled variables (U_0, Θ_0)*

$$\begin{aligned} &H_0 = \frac{\kappa_0}{u_0(V_0)}, \quad \mathcal{M}_0^T\mathcal{M}_0 = -\frac{\kappa_0\nabla^2 u_0(V_0)}{\kappa_2 u_0(V_0)} = H_0\frac{\nabla^2 u_0(V_0)}{\kappa_2}, \quad \kappa_0 = \bar{U}(0), \quad \kappa_2 = \partial_1^2\bar{U}(0), \\ &U_0(z) = H_0 u_0(\mathcal{M}_0^{-1}z + V_0), \quad \Theta_0(z) = \theta_0(\mathcal{M}_0^{-1}z + V_0), \end{aligned} \tag{3.1.13}$$

where $\bar{U}$ is defined in (3.1.7). Since $\nabla u(V_0, 0) = 0$, by definition, (3.1.13) implies the following normalization

$$U_0(0) = \kappa_0 = \bar{U}(0)\,, \quad \nabla U_0(0) = 0\,, \quad \nabla^2 U_0(0) = \nabla^2\bar{U}(0) = \kappa_2 I_d\,. \tag{3.1.14}$$

Let $\nu > 0$ be small, ϵ_2 and C_b be defined in (3.1.30) and (3.3.7), the set of initial data in Theorem 3.1.1 consists of initial data (u_0, θ_0) satisfying (3.1.8) and its rescaled variable (U_0, Θ_0) satisfies

$$U_0\bar{U}^{-1-\epsilon_2} > 2C_b, \quad H_0^{p-1} < \nu, \quad u_0(V_0)^{-p}\mathrm{tr}(\nabla^2 u_0(V_0)) < \nu, \tag{3.1.15}$$

and

$$\|W_0\|_{\mathfrak{E}_K} = \|U_0 - \bar{U}\|_{\mathfrak{E}_K} < \nu, \quad \|\Phi_0\|_{\tilde{\mathfrak{F}}_K} := \|\Theta_0 - \delta\log\bar{U}\|_{\mathfrak{F}_{K-1}} + \|\langle z\rangle^{K-\frac{d}{2}}\nabla^K(\Theta_0 - \delta\log\bar{U})\|_{L^2} < \nu. \tag{3.1.16}$$

[2] While the $\mathfrak{F}_K$ norm in (3.1.9) only involves $\nabla^i\phi, i \ge 1$ and $\mu(t)$ does not play a role in (3.1.9), we keep $\mu(t)$ in (3.1.9) to indicate that it captures the phase of ψ. See (3.1.11), (3.1.12).

[3] Simple linear algebra shows that $\mathcal{M}$ is uniquely determined.

The last quantity in (3.1.15) *is invariant under the parabolic rescaling:* $u_{0,l}(z) = l^{1/(p-1)}u_0(l^{1/2}z)$. *We will use its smallness to show that the viscous terms are small compared to the nonlinear terms. The lower bound* $U_0\bar{U}^{-1-\epsilon} > 2C_b$ *in* (3.1.15) *ensures that* $U_0(z) \neq 0$ *for any* z, *without which can lead to low regularity of rescaled velocity* $|U|^p$ *of* u^p (3.1.5a).

Remark 3.1.3 (Positive definiteness of the Hessian of the initial data)**.** *While the limiting blowup profile* $\bar{U}$ (3.1.7) *is isotropic near* $z = 0$, *we do not need to assume that the initial data* u_0 *is isotropic near* V_0, *i.e.,* $\nabla^2 u_0(V_0)$ *is close to* cI_d *for some* $c \neq 0$. *By introducing the upper triangular matrix* $\mathbf{R}(t)$ *in the rescaling (see* (3.2.1)*), we can handle a much larger class of initial data with non-degenerate global maximizer* (3.1.8).

Remark 3.1.4. *The asymptotics of the blowup solution* (3.1.12) *recovers the constructed result of Masmoudi-Zaag [310] for the case* $d = 1$. *The set of initial data leading to the blowup solution described in Theorem 3.1.1 contains different elements (see Remark 3.1.2) from the one described in [310] which is a subset of* $L^\infty(R)$ *via spectral analysis of a non-selfadjoint linear operator. We note that there is a free phase-shift* $\hat{\mu}$ *in* (3.1.12) *corresponding to the phase invariant of* (CGL) *that was fixed to be* $\hat{\mu}(t) = 0$ *in [310] by a specific choice of initial data through a topological argument. We remark that the relaxing asymptotics* (3.1.10) *and* (3.1.11) *are natural for rigorous stability analysis in all dimensions* $d \geq 1$ *treated in this chapter.*

Note that the asymptotic behavior (3.1.9), (3.1.12) involves the 4 parameter functions $H, \mathbf{R}, V$ and μ (or $\hat{\mu}$) which are responsible for all the symmetries of (CGL).[4] Theorem 3.1.1 is stated in terms of the rescaled profiles, with a singular weight at the origin. We note that the condition (3.1.8) and parameters $\mathcal{M}_0, H_0, V_0$ in (3.1.13) are C^2-stable if the global maximizer is unique. We can therefore simplify the assumptions in Theorem 3.1.1 to obtain the following stability results with a more explicit description of the open set of initial data.

Theorem 3.1.5 (Stability of blowup solutions to (CGL))**.** *Let* $K = K(d, p) \in \mathbb{N}$ *be defined as in* (3.1.29), *and* $\mathcal{H}^K, \mathfrak{F}_{K-1}$ *be the norms defined in* (3.3.9), (3.1.6). *Suppose that* (u_0, θ_0) *satisfies the assumptions* (3.1.8), (3.1.15), (3.1.16) *and* V_0

[4] Although $H(t)$ is absent in (3.1.12), we can replace the factor $(T-t)^{1/(p-1)}$ by $H(t)$ using (3.1.10).

is the unique global maximizer: $u_0(V_0) > u_0(z)$ *for all* $z \neq V_0$. *There exists* $\epsilon_0 = \epsilon_0(u_0) > 0$ *such that if*

$$\|\tilde{u}_0 - u_0\|_{\mathcal{H}^K} + \|(\tilde{u}_0 - u_0)\bar{U}^{-1-\epsilon_2}\|_{L^\infty} + \|\tilde{\theta}_0 - \theta_0\|_{\mathfrak{F}_{K-1}} + \|\langle z\rangle^{K-\frac{d}{2}}\nabla^K(\tilde{\theta}_0 - \theta_0)\|_{L^2} < \epsilon_0, \tag{3.1.17}$$

the solution $\tilde{\psi} = \tilde{u}e^{\iota\tilde{\theta}}$ *to* (CGL) *with the initial data* $\tilde{\psi}_0 = \tilde{u}_0 e^{\iota\tilde{\theta}_0}$ *blows up in finite time* $\tilde{T}$. *Moreover, there exists* $H(t), \mathbf{R}(t), V(t), \mu(t)$ *satisfying* (3.1.10) *and* (3.1.11) *such that* (3.1.9) *and* (3.1.12) *hold for* $(\tilde{u}(t), \tilde{\theta}(t))$ *with* T *being replaced by* $\tilde{T}$.

From the proof of Theorem 3.1.5, it can be shown that ϵ_0 depends on u_0 through its certain norms. We do not state the dependence explicitly for simplicity.

Remark 3.1.6. *The assumptions in Theorems 3.1.1,3.1.5 are satisfied, e.g. for* $u_0 = C\bar{U}$, $\theta_0 = \bar{\Theta}_0$ *with* C *sufficiently large. The weighted norms* $\|\cdot\|_{\mathcal{H}^K}$ (3.3.9) *and* $\|\cdot\|_{\mathfrak{F}_K}$ (3.1.6) *are well-defined for sufficiently smooth functions with fast decay. We do not require that* $u_0 - \tilde{u}_0$ *agrees up to* $O(|z - V_0|^k)$, $k > 0$ *near the maximizer* V_0 *of* u_0.

3.1.2 Dynamic rescaling formulation with extra modulation parameters

The dynamic rescaling formulation or the modulation technique was developed to study singularity formation in the nonlinear Schrödinger equation [315], [263] numerically and various nonlinear PDEs; see the comprehensive references on (NLH), (NLS), and related models in [380], [147]. Recently, researchers also generalized this technique for fluid mechanics [72, 73, 74], [135]. We can establish singularity in two steps. Firstly, one constructs an approximate steady state of the dynamic rescaling equation (analytically or numerically). Secondly, one performs linear and nonlinear stability analysis for perturbation around the approximate steady state with appropriate normalization conditions. The law of blowup will then be prescribed by the normalizing constants. One can establish stability using a L^2-based [76, 75] or L^∞-based [73] argument. In these arguments, one of the crucial steps is to design appropriate singular weights depending on the profile, and then use the weights to derive damping terms for the energy estimates. The approach does not require an explicit profile and is robust to small perturbations, which makes it possible to combine weighted energy estimates for stability analysis, a numerical *implicit* profile, and computer-assisted proofs to construct blowup solutions. See for example, [73, 74] for applications in 3D incompressible Euler equations with smooth data and [76, 75], [70], [219] for related 1D models.

In [218], the authors generalized the L^2-based methodology to establish a type-I blowup for the semilinear heat equation beyond the self-similar setting where there is a logarithm correction in the self-similar scaling for the spatial variable. Compared with previous works [45], [197], [324], [115], [310] where the authors heavily relied on a spectral analysis with detailed properties of the associated linearized operator to establish the existence and stability, we simply suppress unstable directions and neutral modes via a clear characterization of weighted Sobolev spaces, without using Brouwer's fixed-point theorem or a topological argument. The correct Type I blowup rate is automatically inferred by enforcing proper vanishing conditions of the perturbation. We remark that for Type I blowup, there is indeed a link between the vanishing conditions at the origin and the vanishing coefficients projected onto the unstable and neutral spectral eigenfunctions. It is worth mentioning that the method used in [218] shares a similar approach to that of [115], where the authors employed self-similar variables and derived the blowup law by imposing orthogonality conditions on the perturbation with respect to the eigenfunctions (Hermite polynomials) of the associated linear operator in an exponentially weighted L^2 space. In particular, these orthogonality conditions are analogous to the vanishing conditions at the origin in the correct blowup variables considered in [218]. A further explanation of this connection is discussed in Step 2 in Section 3.1.3.

A generalized dynamical rescaling formulation with extra modulation parameters[5]. In this article, we further generalize the above framework and the ideas in [218]. In particular, it consists of three main steps:

Step 1 (Renormalization with extra modulation parameters): The renormalization is an essential step in the study of nonlinear PDEs with symmetries including incompressible/compressible fluids equations. For the equation (CGL), there are trivial symmetries (see Section 3.2.1) from which we introduce the following renormalization in terms of the amplitude-phase representation $\psi(x,t) = u(x,t)e^{\iota\theta(x,t)}$,

$$U(z,\tau) = H(\tau)u\big(\mathbf{R}(\tau)z+V(\tau),t(\tau)\big), \quad \Theta(z,\tau) = \theta\big(\mathbf{R}(\tau)z+V(\tau),t(\tau)\big), \quad t(\tau) = \int_0^\tau H^{p-1}(s)ds, \tag{3.1.18}$$

where $\mathbf{R}(\tau) \in \mathbb{R}^{d\times d}$ is a upper triangular matrix, $V(\tau) \in \mathbb{R}^d$ and $H(\tau) \in \mathbb{R}_+$. Here, H is responsible for time, V for spatial translation.

[5]The modulation parameters are also known as normalization constants in the dynamic rescaling formulation.

The key novelty is that in addition to modulation parameters corresponding to the symmetries, we introduce *extra anisotropic modulation parameters*. Instead of applying the same rescaling to z_i, we rescale z_i with *different but similar* scalings following the ideas [218]. See also a recent work [211] on the generalized Navier-Stokes equations, in which the author developed a generalized dynamic rescaling formulation by using different rescalings for the r and z directions respectively and a self-similar blowup was observed numerically. In the case of (CGL), we rescale z_i with scaling slightly perturbed from the parabolic scaling. We remark that the choices of different scalings for z_i violate the scaling symmetries. Yet, in the case of (CGL), we will show that the violation is asymptotically small and $\mathbf{R}(\tau)$ converges to $c(\tau)I_d$ asymptotically for some scalar function $c(\tau)$. Therefore, the above renormalization (3.1.18) asymptotically agrees with the classical dynamical rescaling formulation [73, 74], [76, 75]. These extra parameters provide us extra $d-1$ degrees of freedom, and we have crucial $1+d+\frac{d(d+1)}{2}$ degrees of freedom in total in choosing the dynamic variables $H(\tau), V(\tau), \mathbf{R}(\tau)$. A natural idea to represent the scaling in z_i and capture the rotation symmetries is to choose $\mathbf{R}(\tau) = D(\tau)Q(\tau)$ with a diagonal matrix D and an orthogonal matrix Q. Yet, it is challenging to parametrize a time-dependent orthogonal matrix in $\mathbb{R}^{d\times d}$. We overcome this difficulty by using the upper triangular matrix $\mathbf{R}(\tau)$ with $\frac{d(d+1)}{2}$ parameters.

To determine these modulation parameters, we impose normalization conditions on $\nabla^i U(0), i = 0, 1, 2$, so that the perturbation of U vanishes $O(|z|^3)$ near $z = 0$ and we can perform weighted energy estimates mentioned above. Note that the number of (different) equations and that of the degrees of freedom are *exactly* the same. For (CGL), these conditions allow us to completely eliminate the unstable and neutrally stable modes of the linearized operator. See Step 2 in Section 3.1.3 for more details.

Step 2 (Equations of the profiles and modulation parameters): We derive the equations of $F = (U, \Theta)$ and matrix (or vectors) $\mathcal{Q}(\tau)$ governing the modulation parameters $\mathbf{R}(\tau), V(\tau), H(\tau)$,

$$\partial_\tau F = \mathcal{N}_F(F, \mathcal{Q}), \quad \frac{d}{d\tau}\mathcal{Q} = \mathcal{N}_{\mathcal{Q}}(F, \mathcal{Q}), \tag{3.1.19}$$

where $\mathcal{N}_F$ is a nonlinear function and $\mathcal{N}_{\mathcal{Q}}$ is a matrix. Then the rescaling system is completely determined, and we further construct the approximate steady state $(\bar{F}, \bar{\mathcal{Q}})$ analytically or numerically.

Step 3 (Stability analysis and the log correction): In general, we *do not* know *a-priori* that the approximate steady state $(\bar{F}, \bar{\mathcal{Q}})$ is stable in some suitable topology.

Nevertheless, if we can establish stability of $(\bar{F}, \bar{Q})$ following the strategy mentioned above and $H(\tau)^{p-1}$ is integrable, then we can obtain finite time blowup using (3.1.18) and the law of blowup will then be prescribed by the normalizing constants.

For (CGL), we will use the energy method with an energy E for the perturbation $F - \bar{F}$ to establish

$$\frac{d}{d\tau}E \le -c_1 \cdot E + C\mathrm{tr}(Q) + l.o.t., \qquad \frac{d}{d\tau}\mathrm{tr}(Q) \le -c_2 \cdot \mathrm{tr}(Q)^2 + l.o.t.,$$

for some $c_1, c_2, C > 0$, where Q is a positive definite matrix and $l.o.t.$ denotes some terms that are very small. The second ODE of $\mathrm{tr}(Q)$ further implies that $|\mathrm{tr}(Q)| \lesssim (1+\tau)^{-1}$. A further refinement of this algebraic decay in the self-similar time implies a log correction $\log(T-t)$ in the blowup rate. We will elaborate more in Step 2 in Section 3.1.3.

One can thus hope to combine the above method for a log correction and the framework [73, 74] to problems with numerical steady states, while spectral analysis heavily hinges on a simple and analytical approximate steady state with explicit (nonlinear heat [324]) or at least asymptotical spectral information of the linearized operator (Keller-Segel [92]). For example, constructing a smooth (approximate) steady state in the dynamic rescaling equations analytically for 3D incompressible Euler or Navier-Stokes equations (NSE) is challenging and remains an open problem. On the other hand, constructing a numerical approximate steady state with computer-assistance is much more feasible. See [73, 74], [450], [445] for the construction in 3D Euler equations. For NSE, self-similar blowup with a perfect self-similar scaling has been ruled out [433], [341]. Yet, one can construct a blowup violating these non-blowup results by adding a log correction in the spatial variable. See numerical evidence on the singular behavior of NSE with a potential logarithm correction in the potential blowup by the second author [213].

3.1.3 Ideas of the blowup analysis

We first discuss some of difficulties in the study of singularity formation in (CGL). Then we follow the generalized dynamic rescaling framework to establish the existence and stability of asymptotically self-similar blowup solutions to (CGL) by briefly discussing the strategy and main ideas of our analysis.

Difficulties: Compared with (NLH) or (NLS), the analysis for the complex Ginzburg-Landau equation (CGL) has the following additional challenges.

1. The complex Ginzburg-Landau equation (CGL) is not of a gradient form, rendering energy estimates hard. To overcome this challenge, we use the amplitude-phase representation (3.1.4), (3.1.5) to analyze (CGL).

2. We remove the even symmetry assumption of the perturbation required in [218] to recover full stability. Without the even symmetry assumption, we have more potentially unstable modes for the linearized operator. We control these unstable modes using the generalized dynamic rescaling formulation in Step 1 in Section 3.1.2.

3. We consider the whole range of the nonlinearity $p > 1$. For the analysis of the phase equation (3.1.5b) and the nonlinearity u^p (3.1.5a), (3.1.5b), we need to bound the rescaled amplitude U from below, which we establish using the maximal principle and a weighted L^∞ estimate. Due to the non-integer power p to control $\nabla^K(U^p)$ in the H^K estimate, which leads to terms like $U^{p-K}(\nabla U)^K$, we need to obtain sharp decay estimates for $\nabla^i U$. This is done by choosing an almost tight power in the far field of the weight for the weighted H^k energy estimates and using interpolation and embedding inequalities following [78]. An additional difficulty comes from the coupling between u, θ in the viscous terms in (3.1.5a), (3.1.5b). We design the top order energy with a special algebraic structure to cancel out the top order terms and show that the viscous terms have a good sign. See Step 3(b) in Section 3.1.3.

Ideas and strategy: We briefly discuss the strategy and main ideas of our analysis.

Step 1 (Dynamical rescaling formulation): We follow Step 1 in Section 3.1.2 to perform the rescaling (3.1.18). Then, we introduce the following factors governing the evolution of these parameters

$$\frac{H_\tau}{H} = c_U, \quad c_U = -\frac{1}{p-1} + c_W, \quad \mathcal{M} = e^{-\frac{\tau}{2}}\mathbf{R}^{-1}, \quad \mathcal{V} = -\mathbf{R}^{-1}V_\tau, \quad \mathcal{P} = \mathcal{M}_\tau \mathcal{M}^{-1}. \tag{3.1.20}$$

Step 2 (Normalization and vanishing conditions): Let $\bar{U}$ be the profile defined in (3.1.7). To determine the law for the parameter functions $H(\tau), V(\tau), \mathbf{R}(\tau)$, we enforce the following normalization conditions on the amplitude U:

$$k = 0, 1, 2, \quad \nabla^k U(0, \tau) = \nabla^k \bar{U}(0). \tag{3.1.21}$$

Since $\nabla^2 U \in \mathbb{R}^{d\times d}$ is symmetric, we have $1+d+\frac{d(d+1)}{2}$ different equations, which match the degrees of freedom of the dynamic variables exactly. The above conditions determine the initial modulation parameters $H(0), \mathbf{R}(0), V(0)$ and the leading order system of $c_W, \mathcal{P}, \mathcal{V}$ (3.1.20)

$$c_W = \frac{2(1-\beta\delta)}{4\mathfrak{b}_*}\mathrm{tr}(Q) + O(\mathcal{E}_0), \quad \mathcal{V} = O(\mathcal{E}_0), \quad \mathcal{P} = O(|Q| + \mathcal{E}_0), \quad \text{where} \quad Q = H^{p-1}e^{\tau}\mathcal{M}\mathcal{M}^T, \tag{3.1.22}$$

where $\mathcal{E}_0$ tracks some lower order terms depending on the perturbation (W, Φ) (3.1.24a), (3.1.24b) and Q. The main unknown $Q \in \mathbb{R}^{d\times d}$ (a positive definite matrix) solves the following ODE

$$\frac{d}{d\tau}\mathrm{tr}(Q) = -\mathrm{tr}(Q^2) + O(\mathcal{E}_0|Q|) \le -\frac{1}{d}(\mathrm{tr}(Q))^2 + O(\mathcal{E}_0|Q|). \tag{3.1.23}$$

From (3.1.23), (3.1.22), (3.1.20), we can control all the modulation parameters. A refined estimate using $\mathrm{tr}(Q)$ and $\mathrm{tr}(Q^{-1})$ yields $Q = \frac{1}{\tau}\mathrm{I}_d + O(\tau^{-3/2+})$, together with an asymptotic refinement of the phase yield the asymptotics in Theorem 3.1.1. See Section 3.2.3 for deriving (3.1.22), Section 3.4 for the estimates of Q and Proposition 3.4.2 for the refinement of the phase.

Roughly speaking, imposing (3.1.21) for U is equivalent to imposing local orthogonality conditions for the perturbation $W = U - \bar{U}$ to $1, z_i, z_i z_j, 1 \le i, j \le d$. These functions are all the neutrally stable and unstable modes of $L = \mathrm{Id} - \frac{1}{2}z\cdot\nabla$, which behaves similarly to the main linearized operator $\mathscr{L}_{\bar{U}}$ in (3.1.24a) for $|z|$ small. We then get a damping in the weighted L^2 energy estimate.

Step 3 (Stability analysis): We linearize (U, Θ) around the approximate steady state $(\bar{U}, \bar{\Theta})$ defined in (3.1.7) and (3.1.27) and obtain the equations for the perturbation $W = U - \bar{U}, \Phi = \Theta - \bar{\Theta}$,

$$W_\tau = \mathscr{L}_{\bar{U}}W + \mathscr{F}_U + \mathscr{N}_U + \mathscr{D}_U, \quad \mathscr{L}_{\bar{U}}W = \big(-\frac{1}{p-1} + p\bar{U}^{p-1} - \frac{1}{2}z\cdot\nabla\big)W, \tag{3.1.24a}$$

$$\Phi_\tau = -\frac{1}{2}z\cdot\nabla\Phi + \mathscr{F}_\Theta + \mathscr{N}_\Theta + \mathscr{D}_\Theta, \tag{3.1.24b}$$

where $\mathscr{F}_U, \mathscr{N}_U, \mathscr{D}_U, \mathscr{F}_\Theta, \mathscr{N}_\Theta, \mathscr{D}_\Theta$ are small residue, nonlinear, and viscous terms (3.3.3), and will be treated perturbatively. Below, we outline the stability estimates and focus on the linear part and the viscous terms.

(a) Estimates of the (W, Φ) : $\mathscr{L}_{\bar{U}}$ to the leading order is the linearized equation for the Riccati equation or the semilinear heat equation. With the vanishing conditions

$\nabla^k W = 0, k = 0, 1, 2$ (3.1.21), we obtain its stability using weighted H^k estimates with singular weights. See [218] in the case of $p = 2$ and [77], [76, 75], [70].

The right-hand side of (3.1.24) only involves $\nabla\Phi$, which enjoys better stability. We estimate $\nabla\Phi$ by performing a similarly weighted H^k estimate (starting from $k = 1$) on (3.1.24b) and exploiting the term $-\frac{1}{2}z \cdot \nabla\Phi$.

In the nonlinear estimates and the estimates of the phase, we need to control $1/U$. We use the maximal principle and a weighted L^∞ estimate to obtain a lower bound of U.

(b) Estimates of the viscous terms: For the viscous terms $\mathscr{D}_U, \mathscr{D}_\Theta$ (3.2.9), the main difficulty is the coupling between U and Θ. At the top H^K estimate, the highest order derivative terms read

$$\nabla^K \mathscr{D}_U = \Delta_Q \nabla^K U - \beta U \Delta_Q \nabla^K \Theta + l.o.t. := I_1 + I_2 + l.o.t.,$$
$$\nabla^K \mathscr{D}_\Theta = \beta \frac{\Delta_Q \nabla^K U}{U} + \Delta_Q \nabla^K \Theta + l.o.t. := I_3 + I_4 + l.o.t.,$$

where $\Delta_Q F$ is a weighted elliptic operator defined in (3.2.6). The terms I_1, I_4 lead to damping terms of $\nabla^{K+1} U$ and $\nabla^{K+1}\Theta$ via integration by parts. To control I_2, I_3, we exploit their cancellation using the energy J_1 below with some weight ρ_K independent of U, Θ and couple their estimates in J_2

$$J_1 = \int (|\nabla^K W|^2 + U^2 |\nabla^K \Phi|^2)\rho_K, \quad J_2 = \int ((-\beta U \Delta_Q \nabla^K \Theta)\cdot \nabla^K U + (\beta \frac{\Delta_Q \nabla^K U}{U})\cdot U^2 \nabla^K \Phi)\rho_K. \tag{3.1.25}$$

Applying integration by parts, J_2 reduces to some lower order terms and we can close the viscous estimates. For estimates of intermediate-order terms, we use interpolation inequalities following [78].

(c) Choosing the weights: To extract damping in the energy estimates, we need to design various suitable weights. The weights (3.1.28) for W are very similar to those of the semilinear heat equation [218]. They are singular near $z = 0$ for the lower order energy estimates and regular for the top order energy estimates so that the viscous terms will have a good sign. In addition, we choose an almost optimal rate at the infinity for these weights to obtain a sharp decay estimate for (W, Φ) using interpolation and embedding following [78].

Organization of the chapter: The rest of the chapter is organized as follows. In Section 3.2, we introduce the generalized dynamic rescaling formulation using the

symmetries of (CGL) and derive the ODEs governing the modulation parameters. Section 3.3 is devoted to the stability analysis of the profile. In Section 3.4, we establish the asymptotics of the blowup rate. In Section 3.4.3, we prove Theorem 3.1.1 and Theorem 3.1.5.

Notations: We use ι to denote the imaginary number, $\bar{f}$ to denote approximate profiles for the variable f, e.g.,$\bar{U}$, rather than conjugates, and $(\cdot,\cdot)$ to denote the inner product on $\mathbb{R}^d$: $(f,g)=\int_{\mathbb{R}^d} fg$. For a weight ρ, we denote $\|f\|_\rho=(|f|^2,\rho)^{1/2}$. For matrix notations, we use $\mathrm{tr}(R)$ to denote the trace of a matrix R, $\mathbf{T}^u(R)$ to denote the upper triangular part of R; namely $(\mathbf{T}^u(R))_{ij}=R_{ij}\mathbf{1}_{i\le j}$. We use $\delta_{ij}=\mathbf{1}_{i=j}$ to denote the Kronecker delta function, and $|\mathbf{T}|:=(\sum_i \mathbf{T}_i^2)^{1/2}$ with summation over all entries $\mathbf{T}_i$ to denote the tensor norm of a tensor $\mathbf{T}$, e.g., higher-order derivatives $\nabla^k f$. We use C to denote an absolute constant only dependent on the constants p,β,δ,γ and the dimension d, which may vary from line to line. $C(\mu)$ denotes a constant depending on μ. We denote $A=O(B)$ or $A\lesssim B$ if there exists an absolute constant $C>0$, such that $|A|\le CB$, and denote $A\approx B$ if $A\lesssim B$ and $B\lesssim A$. Furthermore, we denote

$$\Lambda=z\cdot\nabla,\quad \langle z\rangle=\sqrt{1+|z|^2}\,. \tag{3.1.26}$$

Parameters and special functions: We introduce

$$\bar{\Theta}=\frac{\delta}{p-1}\tau+\delta\log\bar{U}\,,\quad c_p=\frac{(p-1)^2}{4\mathfrak{b}_*},\quad \sigma=-\frac{2}{p-1}. \tag{3.1.27}$$

We choose the weights for the H^k estimates as follows:

$$\begin{aligned}
&\rho_k=|z|^{-6+\epsilon-d+2k}+c_0|z|^{-2\sigma-\epsilon-d+2k},\ 0\le k\le\frac{d+5}{2}\,,\quad \rho_k=1+c_1|z|^{-2\sigma-\epsilon-d+2k},\ \frac{d+5}{2}<k\,,\\
&\mathring{\rho}_k=|z|^{2k-1-d}\,,\ 0<k\le\frac{d}{2}\,,\qquad \mathring{\rho}_k=1+|z|^{2k-1-d}\,,\ \frac{d}{2}<k<K\,,\qquad \mathring{\rho}_K=U^2\rho_K\,,
\end{aligned} \tag{3.1.28}$$

where we determine the constants in the following order:

$$K=2d+4+2\left\lceil\frac{p+1}{\min\{p-1,c_p\}}\right\rceil, \tag{3.1.29}$$

$$\epsilon=\min\{\frac{p-1}{5(p+3)(K+p)},\frac{4}{5(p+3)(K+p)}\},\quad \epsilon_2=\frac{(p-1)\epsilon}{4}. \tag{3.1.30}$$

For c_0,c_1 used in (3.1.28), we determine c_0 via (3.3.26) and c_1 via (3.3.47). Note that c_0,c_1 only depends on K,ϵ,p,δ. Hence, they are considered as fixed constants throughout the chapter.

3.2 Generalized dynamical rescaling formulation

In this section, we introduce a generalized dynamic rescaling formulation and decompose the complex Ginzburg-Landau equation into the equation of the phase and the amplitude. We will consider a linearization around the approximate profiles and choose the modulation parameters based on the vanishing conditions. Finally, we will estimate the ODE of the modulation parameters.

3.2.1 Symmetries and renormalization

We exploit the following symmetries of equation (CGL) to study stability for general perturbation, which will motivate our choice of rescaling. If $\psi(x,t)$ solves (CGL), all of the following also solves (CGL):

1. Phase shift: $\psi_a(x,t) \coloneqq e^{ia}\psi(x,t)$, for $a \in \mathbb{R}$.
2. Parabolic scaling for $\gamma = 0$: $\psi^l(x,t) \coloneqq l^{1/(p-1)}\psi(l^{1/2}x, lt)$, for $l \in \mathbb{R}$.
3. Translation: $\psi_b(x,t) \coloneqq \psi(x-b,t)$, for $b \in \mathbb{R}^d$.
4. Rotation: $\psi^R(x,t) \coloneqq \psi(Rx,t)$, for orthogonal matrix $RR^T = I_d$.

We use the symmetry groups of the parabolic scaling via a rescale in amplitude, of the translation via a shift in space, and of the rotation via a rotation and rescaling in the spatial variable parametrized by an upper triangular matrix. In sum, we can exploit modulation with $1+d+\frac{(d+1)d}{2}$ degree of freedom. The phase shift invariance corresponding to a constant addition in Θ is taken care of in (3.1.11) and (3.1.12). It is irrelevant to the dynamical modulation of stability since the right-hand sides of (3.1.5a) and (3.1.5b) only involve the derivatives of Θ. We remark that the modulation of symmetries due to Galilean transformations, including the rotation symmetry, has been used successfully to obtain shock formation in compressible Euler equations with fine characterization [48]. Below, we will use a general upper triangle matrix, which simplifies the parametrization of the time-dependent orthogonal matrix.

For solution ψ to (CGL), we consider the amplitude-phase form $\psi(x,t) = u(x,t)e^{i\theta(x,t)}$, where $u(t) : x \in \mathbb{R}^d \to \mathbb{R}_+$ and $\theta(t) : x \in \mathbb{R}^d \to \mathbb{R}$. For the amplitude u and phase ψ, we introduce the generalized dynamic rescaling formulation

$$U(z,\tau) = H(\tau)u\big(\mathbf{R}(\tau)z+V(\tau),t(\tau)\big)\,, \quad \Theta(z,\tau) = \theta\big(\mathbf{R}(\tau)z+V(\tau),t(\tau)\big)\,, \tag{3.2.1}$$

where the main unknown parameter functions are $\mathbf{R} \in C^1\big([\tau_0, +\infty), \mathbb{R}^{d\times d}\big)$ a non degenerate upper triangular matrix, $V \in C^1\big([\tau_0, +\infty), \mathbb{R}^d\big)$ and H is given by

$$H = H(0)\exp\big(\int_0^\tau c_U(s)ds\big)\,, \quad t(\tau) = \int_0^\tau H^{p-1}(s)ds\,. \tag{3.2.2}$$

In a compact form, we have

$$U(z,\tau)e^{\iota\Theta(z,\tau)} = H(\tau)(ue^{\iota\theta})\big(\mathbf{R}(\tau)z + V(\tau), t(\tau)\big), \quad \psi = ue^{\iota\theta}. \tag{3.2.3}$$

We decompose the solution into the approximate steady states (3.1.27), with perturbations W, Φ:

$$U = \bar{U} + W\,, \quad \Theta = \bar{\Theta} + \Phi\,, \quad c_U = -\frac{1}{p-1} + c_W\,, \quad H = e^{-\frac{\tau}{p-1}}C_W. \tag{3.2.4}$$

If $\mathbf{R}(\tau)$ is a scalar factor and $V = 0$, (3.2.3) reduces to the standard dynamic rescaling formulation, see e.g., [75, 76, 72]. If $\mathbf{R}(\tau)$ is a diagonal matrix and $V = 0$, it reduces to a formulation similar to [218]. We will show that $\mathbf{R}$ is close to some identity matrix and V is some lower order term.

We first compute the spatial derivative

$$\nabla U = He^{-\iota\Theta}\nabla\psi\mathbf{R} - \iota H\psi e^{-\iota\Theta}\nabla\Theta\,,$$

$$\nabla^2 U = He^{-\iota\Theta}\mathbf{R}^T\nabla^2\psi\mathbf{R} - \iota He^{-\iota\Theta}(\nabla\psi\mathbf{R}\nabla\Theta^T + \nabla\Theta\mathbf{R}^T\nabla\psi^T) - H\psi e^{-\iota\Theta}\nabla\Theta\nabla\Theta^T - \iota H\psi e^{-\iota\Theta}\nabla^2\Theta\,.$$

We then write from (CGL) the equation for U,

$$\begin{aligned} U_\tau = {} & -\iota\Theta_\tau U + c_U U - (\frac{1}{2}z + \mathcal{P}z + \mathcal{V})\cdot\nabla U - \iota U(\frac{1}{2}z + \mathcal{P}z + \mathcal{V})\cdot\nabla\Theta + (1+\iota\delta)U^p - C_U^{p-1}\gamma U \\ & + (1+\iota\beta)(\Delta_Q U + 2\iota\langle\nabla U, \nabla\Theta\rangle_Q - U\langle\nabla\Theta, \nabla\Theta\rangle_Q + \iota U\Delta_Q\Theta)\,, \end{aligned}$$

where $\mathcal{P}, \mathcal{V}$ are related to the matrix $\mathbf{R}$ as

$$\mathcal{M} = e^{-\tau/2}\mathbf{R}^{-1}, \quad \mathcal{V} = -\mathbf{R}^{-1}V_\tau, \quad \mathcal{P} = \mathcal{M}_\tau\mathcal{M}^{-1}, \quad Q := C_W^{p-1}\mathcal{M}\mathcal{M}^T\,, \tag{3.2.5}$$

and we use the notation

$$\Delta_Q f := \mathrm{tr}(Q\nabla^2 f)\,, \quad \langle x, y\rangle_Q := x^T Q y\,, \quad \forall x, y \in \mathbb{R}^d. \tag{3.2.6}$$

Taking the real and imaginary parts we arrive at the following equations for U and Θ:

$$U_\tau = c_U U - (\frac{1}{2}z + \mathcal{P}z + \mathcal{V})\cdot\nabla U + U^p - C_U^{p-1}\gamma U + \mathscr{D}_U\,, \tag{3.2.7}$$

$$\Theta_\tau = -(\frac{1}{2}z + \mathcal{P}z + \mathcal{V})\cdot\nabla\Theta + \delta U^{p-1} + \mathscr{D}_\Theta\,, \tag{3.2.8}$$

where $\mathscr{D}_U$ and $\mathscr{D}_\Theta$ consist of the viscous terms and the nonlinear quadratic term we define the viscous terms as follows:

$$\mathscr{D}_U = \Delta_Q U - 2\beta\langle\nabla U, \nabla\Theta\rangle_Q - U\langle\nabla\Theta, \nabla\Theta\rangle_Q - \beta U\Delta_Q\Theta\,, \tag{3.2.9a}$$

$$\mathscr{D}_\Theta = \beta\frac{\Delta_Q U}{U} + 2\frac{\langle\nabla U, \nabla\Theta\rangle_Q}{U} - \beta\langle\nabla\Theta, \nabla\Theta\rangle_Q + \Delta_Q\Theta\,. \tag{3.2.9b}$$

We will show that the diffusion and $Q, c_W, \mathcal{V}, \mathcal{P}, H^{p-1}$ are lower order terms. See Remark 3.2.2. Dropping these terms and setting $\partial_\tau U = 0$, we obtain the leading order parts of (3.2.7) and (3.2.8):

$$-\frac{1}{p-1}\bar{U} - \frac{1}{2}\Lambda\bar{U} + \bar{U}^p = 0\,, \quad \bar{\Theta}_\tau = -\frac{1}{2}\Lambda\bar{\Theta} + \delta\bar{U}^{p-1}\,,$$

whose solution are given by the approximate profiles $(\bar{U}, \bar{\Theta})$ defined in (3.1.7), (3.1.27).

3.2.2 Initial rescaling and normalization conditions

We will choose some proper initial modulation parameters $H(0), V(0), \mathcal{M}(0)$ and the dynamic variables $c_W, \mathcal{V}, \mathcal{P}$ such that the perturbation W vanishes to the third order. We denote the following constants

$$\kappa_0 = \bar{U}(0) = (p-1)^{-\frac{1}{p-1}}\,, \quad \kappa_2 = \partial_1^2\bar{U}(0) = -\frac{2c_p\kappa_0}{(p-1)^2}\,, \quad \kappa_4 = \partial_1^4\bar{U}(0) = \frac{12pc_p^2\kappa_0}{(p-1)^4}\,. \tag{3.2.10}$$

Given initial data (u, θ) (3.2.3) satisfying (3.1.8), we first define $\mathcal{M}_0, V_0, H(0), \Theta_0, U_0$ using (3.1.13). Then we determine other initial rescalings and initial data using

$$V(0) = V_0, \quad \mathcal{M}(0) = \mathcal{M}_0, \quad \mathbf{R}(0) = \mathcal{M}_0^{-1}, \quad \Theta(z,0) = \Theta_0(z), \quad U(z,0) = U_0(z).$$

We impose the following normalization conditions in time as

$$U(0,\tau) = \bar{U}(0) = \kappa_0\,, \quad \nabla U(0,\tau) = \nabla\bar{U}(0) = 0\,, \quad \nabla^2 U(0,\tau) = \nabla^2\bar{U}(0) = \kappa_2 I_d\,.$$

From (3.1.14), the above holds for $\tau = 0$. By the ansatz (3.2.3), it reduces a dynamical condition in time

$$\partial_\tau\nabla^k U(0,\tau) = 0\,, \quad k = 0,1,2\,,$$

which we can use (3.2.7) to simplify as

$$\begin{aligned}
&c_U + \kappa_0^{p-1} - H^{p-1}\gamma + \frac{\mathscr{D}_U(0)}{\kappa_0} = 0,\\
&\kappa_2\mathcal{V} = \nabla\mathscr{D}_U(0),\\
&(c_U - 1 + p\kappa_0^{p-1} - H^{p-1}\gamma)\delta_{ij} - (\mathcal{P}_{ij} + \mathcal{P}_{ji}) + \frac{\partial_{ij}\mathscr{D}_U(0) - \mathcal{V}\cdot\nabla\partial_{ij}U(0)}{\kappa_2} = 0\,,
\end{aligned} \tag{3.2.11}$$

for any indices i, j. Notice that the inverse of an upper-triangular matrix is still upper-triangular, and as a consequence $\mathcal{P} = \mathcal{M}_\tau \mathcal{M}^{-1}$ is upper-triangular. We can further simplify the equations for c_U and $\mathcal{P}$ by the ansatz (3.2.4) as follows:

$$c_W = -\frac{\mathscr{D}_U(0)}{\kappa_0} + H^{p-1}\gamma\,, \quad (1+\delta_{ij})\mathcal{P}_{ij} = -\frac{\mathscr{D}_U(0)}{\kappa_0}\delta_{ij} + \frac{\partial_{ij}\mathscr{D}_U(0) - \mathcal{V}\cdot\nabla\partial_{ij}W(0)}{\kappa_2}\,, \tag{3.2.12}$$

for any $i \le j$. We will estimate equations (3.2.12) and (3.2.11) in the next subsection.

3.2.3 ODE for the modulation parameters

In this subsection, we simplify equations (3.2.12) and (3.2.11) to derive a leading order ODE. We will treat the perturbations W, Φ as low-order terms and estimate them in Section 3.3. Denote

$$\Gamma = \max_{0\le k\le 5, 1\le l\le 5} (\|\nabla^k W\|_\infty, \|\nabla^l \Phi\|_\infty)\,, \quad \mathcal{E}_0 = |Q|(\Gamma + \Gamma^4) + H^{p-1}. \tag{3.2.13}$$

Clearly, we have $\Gamma^i |Q| \lesssim \mathcal{E}_0$, $1 \le i \le 4$. We use $\mathcal{E}_0$ to track some lower order terms.

Lemma 3.2.1. *We have the following estimates for the modulation parameters:*

$$c_W = \frac{2(1-\beta\delta)}{(p-1)^2} c_p \mathrm{tr}(Q) + O(\mathcal{E}_0)\,, \quad \mathcal{V} = O(\mathcal{E}_0)\,, \tag{3.2.14}$$

and

$$\mathcal{P} = O(|Q|(1+\Gamma^4) + H^{p-1})\,, \quad Q_\tau = -(Q_u + \frac{1}{2}Q_d)Q - Q(Q_u^T + \frac{1}{2}Q_d) + O(\mathcal{E}_0|Q|), \tag{3.2.15}$$

where Q_u, Q_d are the strictly upper part and diagonal part of Q.

Remark 3.2.2. *Formally to the leading order when we take the trace, we have*

$$\mathrm{tr}(Q)_\tau \approx -\mathrm{tr}(Q^2)\,, \quad \mathrm{tr}(Q^{-1})_\tau \approx -d\,.$$

Recall that $Q = C_W^{p-1}\mathcal{M}\mathcal{M}^T$ is positive. We can estimate $\mathrm{tr}(Q), \mathrm{tr}(Q^{-1})$ and obtain $Q \approx \tau^{-1}\mathrm{I}_d$. Therefore $Q, c_W, \mathcal{V}, \mathcal{P}$ are indeed small and the viscous terms can be treated perturbatively. We will make this heuristic rigorous by choosing $H(0) = C_W(0)$ small; see Section 3.3.6. Although we allow anisotropicity in the initial data, the profile will converge to a isotropic one, i.e. Q will converge to a diagonal matrix.

Proof. Notice that by (3.2.9a), we have

$$\mathscr{D}_U = \mathrm{tr}(QS_1)\,, \quad S_1 = \nabla^2 U - 2\beta\nabla U\nabla\Theta^T - U\nabla\Theta\nabla\Theta^T - \beta U\nabla^2\Theta\,. \tag{3.2.16}$$

We can simplify (3.2.12) by an asymptotic expansion of S_1 near the origin up to the second order.

$$\bar{U} = \kappa_0 + \frac{\kappa_2}{2}|z|^2 + \frac{\kappa_4}{24}|z|^4 + O(|z|^6)\,, \quad \nabla\bar{\Theta} = \frac{\delta}{\bar{U}}\nabla\bar{U} = \delta\frac{\kappa_2}{\kappa_0}z + O(|z|^3)\,,$$
$$\nabla^2\bar{U} = \kappa_2 I_d + \frac{\kappa_4}{6}|z|^2 I_d + \frac{\kappa_4}{3}zz^T + O(|z|^4)\,,$$
$$\nabla^2\bar{\Theta} = \frac{\delta}{\bar{U}}\nabla^2\bar{U} - \frac{\delta}{\bar{U}^2}\nabla\bar{U}\nabla\bar{U}^T = \delta(\frac{\kappa_2}{\kappa_0}I_d + (\frac{\kappa_4}{6\kappa_0} - \frac{\kappa_2^2}{2\kappa_0^2})(|z|^2 I_d + 2zz^T)) + O(|z|^4)\,.$$

Decomposing $U = \bar{U} + W, \Theta = \bar{\Theta} + \Phi$ (3.2.4) and using Γ (3.2.13) to control the perturbation W, Φ, we expand

$$\begin{aligned} S_1 = {} & \kappa_2(1-\beta\delta)I_d + (\frac{\kappa_4}{3} + (-2\beta-\delta)\delta\frac{\kappa_2^2}{\kappa_0} - \beta\delta(\frac{\kappa_4}{3} - \frac{\kappa_2^2}{\kappa_0}))zz^T \\ & + (\frac{\kappa_4}{6} - \beta\delta(\frac{\kappa_2^2}{2\kappa_0} + \frac{\kappa_4}{6} - \frac{\kappa_2^2}{2\kappa_0}))|z|^2 I_d + O(|z|^4 + \Gamma + \Gamma^3)\,. \end{aligned}$$

The estimates for $\nabla^i S_1$ are similar. We have chosen Γ (3.2.13) to control $\nabla^k W, \nabla^{k+1}\Phi$ with high enough order k. In particular, for an error term I in S_1 bounded by $|z|^4 + \Gamma + \Gamma^3$, e.g.,$(U-\bar{U})\nabla^2(\Theta - \bar{\Theta})$, we have

$$|\nabla^i I| \lesssim |z|^{4-i} + \Gamma + \Gamma^3,\ i = 1, 2.$$

Since we only need the expression at $z = 0$, the error term $O(|z|^j), j \geq 1$ vanishes in the following derivations. For this reason, we do not track the constant associated with $|z|^j$.

As a consequence, we have the expressions for derivatives of $\mathscr{D}_U$ as

$$\mathscr{D}_U(0) = \kappa_2(1-\beta\delta)\mathrm{tr}(Q) + O(\mathcal{E}_0)\,, \quad \nabla\mathscr{D}_U(0) = O(\mathcal{E}_0)\,,$$
$$\partial_{ij}\mathscr{D}_U(0) = \delta_{ij}(1-\beta\delta)\frac{\kappa_4}{3}\mathrm{tr}(Q) + 2((1-\beta\delta)\frac{\kappa_4}{3} - (\beta+\delta)\delta\frac{\kappa_2^2}{\kappa_0})Q_{ij} + O(\mathcal{E}_0)\,.$$

We plug the estimates into (3.2.12) and (3.2.11) and get

$$\begin{aligned} c_W = {} & -\frac{\kappa_2}{\kappa_0}(1-\beta\delta)\mathrm{tr}(Q) + O(\mathcal{E}_0)\,, \quad v = O(\mathcal{E}_0)\,, \\ \mathcal{P} = {} & \mathbf{T}^u\Big[\frac{\kappa_4}{6\kappa_2}(1-\beta\delta)\mathrm{tr}(Q)I_d + ((1-\beta\delta)\frac{\kappa_4}{3\kappa_2} - \big(\beta+\delta\big)\delta\frac{\kappa_2}{\kappa_0}\big)\big(2Q - \mathrm{diag}(Q)\big)\Big] \\ & - \mathbf{T}^u(\frac{\kappa_2}{2\kappa_0}(1-\beta\delta)\mathrm{tr}(Q)I_d) + O(\mathcal{E}_0)\,, \end{aligned}$$

where we recall that $\mathbf{T}^u$ is the upper-triangular part of the matrix. Notice that by (3.2.10), we have the relationship

$$\frac{\kappa_4}{6\kappa_2} = p\frac{\kappa_2}{2\kappa_0}\,, \quad (1-\beta\delta)\frac{\kappa_4}{3\kappa_2} - (\beta+\delta)\delta\frac{\kappa_2}{\kappa_0} = (p - \delta^2 - \beta\delta(1+p))\frac{\kappa_2}{\kappa_0} = -\frac{1}{2}\,.$$

Therefore, we collect

$$\mathcal{P} + \frac{p-1}{2} c_W I_d = -\mathbf{T}^u (Q - \frac{1}{2}\mathrm{diag}(Q)) + O(\mathcal{E}_0)\,. \tag{3.2.17}$$

Recall that $\mathcal{M}_\tau = \mathcal{P}\mathcal{M}$ by definition, and we can compute

$$Q_\tau = (p-1)c_W Q + C_W^{p-1}(\mathcal{P}\mathcal{M}\mathcal{M}^T + \mathcal{M}\mathcal{M}^T\mathcal{P}^T) = \big(\mathcal{P} + \frac{p-1}{2} c_W I_d\big)Q + Q\big(\mathcal{P} + \frac{p-1}{2} c_W I_d\big)^T\,.$$

If we decompose Q into the strictly upper, lower, and diagonal parts as $Q = Q_u + Q_u^T + Q_d$, then we can simplify

$$Q_\tau = -(Q_u + \frac{1}{2}Q_d)Q - Q(Q_u^T + \frac{1}{2}Q_d) + \mathcal{E}_Q Q + Q\mathcal{E}_Q^T, \quad \mathcal{E}_Q = O(\mathcal{E}_0).$$

and we conclude the proof of the lemma. □

Remark 3.2.3. *In the above ODE of Q, we get -1 for the coefficient of Q^2 since we have normalized the profile (3.1.27). The factor $p - \delta^2 - \beta\delta(1+p) > 0$ (corresponding to the subcritical case) appears in the constant c_p. For the critical case*

$$p - \delta^2 - \beta\delta(1+p) = 0\,,$$

without such a normalization, we can see that if we do something similar, the coefficient of Q^2 will be zero. We could keep track of the next order terms of size Γ to derive a system of size $|Q|^3$. This can potentially help us establish a result similar to [130] but we do not pursue it here.

3.3 Stability analysis and finite time blowup

In this section, we perform stability analysis and establish nonlinear stability of the perturbation around the approximate steady state following the ideas and strategy outlined in Section 3.1.3.

We linearize (3.2.7) and (3.2.8) around the approximate profile as in ansatz (3.2.4) and obtain the equations of the perturbations as follows:

$$W_\tau = \mathscr{L}_{\bar{U}} W + \mathscr{F}_U + \mathscr{N}_U + \mathscr{D}_U\,, \tag{3.3.1}$$

$$\Phi_\tau = -\frac{1}{2}\Lambda\Phi + \mathscr{F}_\Theta + \mathscr{N}_\Theta + \mathscr{D}_\Theta\,, \tag{3.3.2}$$

where we recall from (3.2.9a) and (3.2.9b) the definition of $\mathscr{D}_U$ and $\mathscr{D}_\Theta$, and define the linear, residue, and nonlinear parts respectively as

$$\begin{aligned}
&\mathscr{L}_{\bar U} W = c_{\bar U} W - \frac{1}{2}\Lambda W + p\bar U^{p-1} W\,, \quad \Lambda = z\cdot\nabla,\\
&\mathscr{F}_U = c_W U - (\mathcal{P} z + \mathcal{V})\cdot\nabla U - C_U^{p-1}\gamma U\,, \quad \mathscr{N}_U = (\bar U + W)^p - \bar U^p - p\bar U^{p-1} W\,,\\
&\mathscr{F}_\Theta = -(\mathcal{P} z + \mathcal{V})\cdot\nabla\Theta\,, \quad \mathscr{N}_\Theta = \delta((\bar U + W)^{p-1} - \bar U^{p-1})\,.
\end{aligned} \tag{3.3.3}$$

We will group the terms by integrability: $\mathscr{L}_U$ and $\mathscr{N}_U$ vanish to the third order at the origin. Recall that by Lemma 3.2.1 and Remark 3.2.2, we know that $Q, c_W, \mathcal{V}, \mathcal{P}$ are small. The viscous terms $\mathscr{D}_U$, $\mathscr{D}_\Theta$ are small of order $|Q|$ with a typical size of $1/\tau$. Also obviously $H = C_W e^{-\tau/(p-1)}$ is small.

We define the weighted H^k energy as follows

$$E_k^2 = (|\nabla^k W|^2, \rho_k)\,, \quad 0\le k\le K\,, \quad F_k^2 = (|\nabla^k\Phi|^2, \mathring\rho_k)\,, \quad 0 < k\le K\,. \tag{3.3.4}$$

Our goal is to prove the following nonlinear stability results.

Theorem 3.3.1. *Denote $E_Q = \mathrm{tr}(Q)$. There exists $0 < E_* < 1$ sufficiently small and $\mu_3 > 0$, such that for any initial perturbation satisfying*

$$U\rho > 2C_b, \quad E(0) < E_*, \quad E_Q < E_*, \quad H^{p-1}(0) < E_*, \tag{3.3.5}$$

we have the following estimates for all $\tau > 0$

$$\begin{aligned}
&E_Q(\tau) \le \min(2E_Q(0), 4d/\tau), \quad U(\tau) > C_b\bar U^{1+\epsilon_2}, \quad H^{p-1}(\tau) < H^{p-1}(0)e^{-\tau/2},\\
&E(\tau) \le e^{-\lambda\tau/2}(E(0) + \mu_3 H^{p-1}(0) + \mu_3 E_Q(0)) + \mu_3\min(E_Q(0), 1/\tau).
\end{aligned} \tag{3.3.6}$$

Here $C_b, H(\tau), E(\tau), \epsilon_2$ are defined in (3.3.7), (3.2.2), (3.3.50), and (3.1.30). Moreover, the bootstrap assumptions 3.3.2, 3.3.9, 3.3.10 (introduced below) hold for all $\tau > 0$.

Note that the parameters μ_1, μ_2 will be introduced in the proof in Section 3.3.6.

We impose the following weak bootstrap assumptions for nonlinear estimates. To control $1/U$, which will appears in the estimate of $\mathscr{N}_U, \mathscr{N}_\Theta$, we impose a lower bound on U, which is almost comparable with $\bar U$, up to a small power. To simplify the notations in the nonlinear estimates and to ensure that $\mathbf{R}, \mathcal{M}$ are invertible (3.2.5), (3.2.4), we impose the following weak assumptions on the energy E_i, F_j and $\det(Q)$.

Assumption 3.3.2. *Let ϵ_2 be defined in (3.1.30). We impose the following bootstrap assumptions*

$$U \geq C_b \bar{U}^{1+\epsilon_2}, \quad C_b = \min_{|z|\leq 1} \bar{U}^{-\epsilon_2}(z)/4 > 0, \tag{3.3.7}$$

$$\max(E_K, E_0, F_K, F_1) \leq 1, \qquad \det(Q) > 0. \tag{3.3.8}$$

Below, we will first establish some functional inequalities in Section 3.3.1. We will start with the L^2 analysis of perturbations W and $\nabla\Phi$ in Sections 3.3.2, 3.3.3, and then build higher-order estimates in Section 3.3.4. We obtain a lower bound of the amplitude U via the maximal principle in Section 3.3.5, inspired by [75]. Then we close the nonlinear estimates and prove Theorem 3.3.1 via a bootstrap argument in Section 3.3.6.

3.3.1 Functional inequalities

In this section, we establish a few functional inequalities, which will be used to estimate the decay of the solution and close the nonlinear estimates. We introduce the following norms

$$||f||_{\dot{\mathcal{H}}^k} := ||\nabla^k f g_k^{1/2}||_{L^2}, \quad ||f||_{\mathcal{H}^k} := ||f||_{\dot{\mathcal{H}}^k} + ||f||_{\dot{\mathcal{H}}^0}, \quad g_k = \langle z\rangle^{-2\sigma-\epsilon-d+2k}. \tag{3.3.9}$$

By definition of ρ_k (3.1.28) and (3.3.4) for E_k, we have

$$g_k \lesssim \rho_k, \quad ||f||_{\dot{\mathcal{H}}^k} \lesssim E_k, \quad ||f||_{\mathcal{H}^k} \lesssim E_0 + E_k. \tag{3.3.10}$$

We define the low-order terms $\mathcal{E}_1, \mathcal{E}_2$ that we later show to be small:

$$\begin{aligned} \mathcal{E}_1 &:= |Q| + H^{p-1} + \sum_{i\leq K-1} E_i + \sum_{1\leq j\leq K-1} F_j + (F_K + F_1 + E_K + E_0)^2, \\ \mathcal{E}_2 &:= |Q| + H^{p-1}. \end{aligned} \tag{3.3.11}$$

We treat $\mathcal{E}_1$ as a lower order term since it either contains nonlinear terms or energy with order lower than E_K, F_K. Note that $\mathcal{E}_2$ is the low-order term of order P in Lemma 3.2.1 by asuming (3.3.8), which implies $\Gamma \lesssim 1$. See (3.3.15).

Following Lemma C.4 in [78], we have the following weighted interpolation and embedding inequalities.

Proposition 3.3.3 (Interpolation)**.** *Let $\sigma = -\frac{2}{p-1}$ and ϵ be the constants defined in (3.2.10), (3.1.30). For any $\mu > 0$, there exists a constant $C(\mu)$, such that the*

following interpolation inequalities hold:

$$E_k \le \mu E_l + C(\mu)E_0\,, \quad \forall 0 \le k < l \le K\,, \tag{3.3.12a}$$

$$F_k \le \mu F_l + C(\mu)F_1\,, \quad \forall 1 \le k < l \le K\,, \tag{3.3.12b}$$

$$||f||_{\dot{\mathcal{H}}^k} \le \mu||f||_{\dot{\mathcal{H}}^l} + C(\mu)||f||_{\mathcal{H}^0}, \quad \forall 0 \le k < l \le K. \tag{3.3.12c}$$

Moreover, we have the following embedding

$$|\nabla^l W| \lesssim \langle z\rangle^{-l+\sigma+\epsilon/2} \min(E_{l+d}, ||W||_{\mathcal{H}^{l+d}}),\ \forall 0 \le l \le K-d\,, \tag{3.3.13a}$$

$$|\nabla^l \Phi| \lesssim \langle z\rangle^{-l+1/2} F_{l+d}, \quad \forall 1 \le l \le K-d-1\,. \tag{3.3.13b}$$

As a result, for $k < l$, we have the following product rule for $i \le n, j \le m$ with $i + d \le n$, or $j + d \le m$,

$$||\nabla^i F \nabla^j G g_{i+j}^{1/2}||_{L^2} \lesssim ||F||_{\mathcal{H}^n}||G||_{\mathcal{H}^m}. \tag{3.3.14}$$

Finally, assuming (3.3.8)*, for $0 \le i \le K, 1 \le j \le K$ and Γ defined in* (3.2.13)*, we have*

$$E_i \lesssim 1, \quad F_j \lesssim 1, \quad \Gamma \lesssim 1. \tag{3.3.15}$$

Since K (3.1.29) is absolute, $l, k \le K$, and the parameters c_0, c_1 in the weights (3.1.28) and norms depend on absolute constants K, ϵ (3.1.30), we only need to track the constants related to μ.

Proof. **(a) Interpolation inequalities.** To prove (3.3.12), we use integration by parts. For (3.3.12a), we compute for $K > k > 0$ that

$$E_k^2 = -\sum_i \int (\partial_i^2 \nabla^{k-1} W \cdot \nabla^{k-1} W \rho_k + \partial_i \nabla^{k-1} W \cdot \nabla^{k-1} W \partial_i \rho_k)\,.$$

Notice that the weights (3.1.28) satisfy

$$\rho_k^2 \lesssim \rho_{k+1}\rho_{k-1}\,, \quad (\partial_i \rho_k)^2 \lesssim \rho_k \rho_{k-1}\,.$$

Combined with a Cauchy-Schwarz inequality, we obtain

$$E_k^2 \lesssim E_{k-1}(E_k + E_{k+1})\,.$$

Since ϵ only depends on K (3.1.30), by a weighted AM-GM inequality ($ab \le \nu a^2 + 1/(4\nu)b^2$), for any $\mu > 0$, we have

$$E_k^2 \le C(\mu)E_{k-1}^2 + \mu E_{k+1}^2\,.$$

From here, to conclude the first inequality, since $\mu > 0$ is arbitrary, we only need to that show it holds for $k = l - 1$, which we can combine the above estimates for $k = 1, 2, .., l - 1$ to establish.

The proof of (3.3.12c) follows from the same argument.

For the second inequality (3.3.12b), we can repeat the same procedure to conclude, provided that the weights $\mathring{\rho}_k$ satisfy the same inequalities for $K > k > 1$ (3.1.28):

$$\mathring{\rho}_k^2 \lesssim \mathring{\rho}_{k+1}\mathring{\rho}_{k-1}\,, \quad (\partial_i\mathring{\rho}_k)^2 \lesssim \mathring{\rho}_k\mathring{\rho}_{k-1}\,.$$

When $k + 1 < K$, this is obvious. For $k = K - 1$, we only need to show

$$\langle z\rangle^{2K-1-d} \lesssim \mathring{\rho}_K \approx U^2\langle z\rangle^{-2\sigma+2K-\epsilon-d}\,,$$

which is true, since by the choice of ϵ, ϵ_2 (3.1.30) and the bootstrap Assumption 3.3.2 we have

$$U \geq C_b\bar{U}^{1+\epsilon_2} \approx \langle z\rangle^{\sigma-\epsilon/2}, \quad 2\epsilon < 1.$$

(b) Embedding (3.3.13)**.** To prove the L^∞ estimates (3.3.13), one can proceed as in [218] and invoke the weighted Morrey-type inequality. Below, we present a simpler proof. By a density argument, we can assume that $W \in C_c^\infty$. Without loss of generality, we fix $z \in \mathbb{R}^d$ with $z_i \geq 0$ and estimate $\nabla^l W(z)$. Consider the region $\Omega(z) = \{y \in \mathbb{R}^d, y_i \geq z_i\}$. We have $|y| \geq |z|$ for any $y \in \Omega(z)$. Denote $\delta = -2\sigma - \epsilon - d > -d$ (3.1.30). We have

$$|\nabla^l W(Z)| \lesssim_l \int_{\Omega(z)} |\partial_1\partial_2..\partial_d\nabla^l W(y)|dy \lesssim_l ||\langle y\rangle^{l+d+\delta/2}\nabla^{l+d}W||_{L^2}\Big(\int_{|y|\geq|z|}\langle y\rangle^{-2l-2d-\delta}dy\Big)^{1/2}.$$

For $0 \leq l \leq K - d - 1$, using $\rho_{l+d} \gtrsim \langle y\rangle^{2(l+d)+\delta}$ (3.1.28), $-2l - d - \delta - 1 < -1$, and the first inequality (3.3.12a) with $k = l + d < K$, for any $\mu > 0$, we further obtain

$$|\nabla^l W(y)| \lesssim_K E_{l+d}(\int_{R\geq|y|}\langle R\rangle^{-2l-2d-\delta}R^{d-1}dR)^{1/2} \lesssim_K E_{l+d}\langle y\rangle^{-l-(d+\delta)/2}.$$

Rearranging the power on both sides, we prove (3.3.13a).

The proof of (3.3.13b) with $1 \leq l \leq K - d - 1$ is similar by replacing W by $\nabla\Phi$ in the above argument and using $\mathring{\rho}_{l+d} \gtrsim \langle y\rangle^{2(l+d)+\delta_2}$ for $\delta_2 = -d - 1$ (3.1.28) and $-2l - d - \delta_2 - 1 = -2l < -1$ for $l \geq 1$.

(c) Other inequalities. For (3.3.14), without loss of generality, we assume that $i + d \leq n$. Using $g_{i+j}^{1/2} \lesssim \langle z\rangle^i g_j^{1/2}$, $\sigma + \frac{\epsilon}{2} < 0$ (3.3.12c), we prove

$$||\nabla^i F\nabla^j G g_{i+j}^{1/2}||_{L^2} \lesssim ||F||_{\mathcal{H}^{i+d}}||\nabla^j G g_j^{1/2}||_{L^2} \lesssim ||F||_{\mathcal{H}^n}||G||_{\mathcal{H}^m}.$$

The inequalities (3.3.15) follow from (3.3.12a), (3.3.12b) and the assumption (3.3.8). □

Combining Assumption 3.3.2 and the decay estimates (3.3.13a), (3.3.13b), we have the following estimates.

Corollary 3.3.4. *Denote $W = \bar{U} - U$. Under the Assumption 3.3.2, for any $\alpha \in [0,1]$ and any $q \in \mathbb{R}$ we have*

$$\begin{aligned} \langle z\rangle^{\sigma-\epsilon/2} \lesssim U \lesssim \langle z\rangle^{\sigma+\epsilon/2}, \\ |(\bar{U}+\alpha W)(z)|^q \lesssim C(|q|)\langle z\rangle^{\sigma q+|\epsilon q|/2}, \\ ||\bar{U}+\alpha W||_{\mathcal{H}^K}, ||U||_{\mathcal{H}^K} \lesssim 1, \end{aligned} \tag{3.3.16}$$

where σ, ϵ are defined in (3.2.10). As a result, for $i + j \le K$, we have the following estimates for the weights

$$\begin{aligned} &\mathring{\rho}_i \lesssim \rho_i \langle z\rangle^{\sigma+\epsilon/2-1/2}, i \le K-1, \quad \mathring{\rho}_K \lesssim \rho_K \langle z\rangle^{\sigma+\epsilon/2}, \\ &\rho_{i+j} \lesssim \rho_i \langle z\rangle^{2j}, \quad \mathring{\rho}_{i+j} \lesssim \mathring{\rho}_i \langle z\rangle^{2j}. \end{aligned} \tag{3.3.17}$$

Proof. **(a) Estimate of U.** By definition of ϵ_2 (3.1.30), we get $-\frac{2}{p-1}(1+\epsilon_2) = -\frac{2}{p-1} - \frac{\epsilon}{2}$. Thus, under Assumption 3.3.2, we yield

$$U \gtrsim \bar{U}^{1+\epsilon_2} \gtrsim \langle z\rangle^{-\frac{2}{p-1}-\frac{\epsilon}{2}}.$$

Since $\bar{U} + \alpha W = \bar{U} + \alpha(U - \bar{U}) = \alpha U + (1-\alpha)\bar{U}$, which is between $U, \bar{U}$, and $\bar{U}, U > 0$, using (3.3.13a), $K > d$ (3.1.29) and the above estimate, we prove

$$|(\bar{U}+\alpha W)| \lesssim \bar{U} + U \lesssim (1 + E_K + E_0)\langle z\rangle^{\sigma+\epsilon/2}, \quad |(\bar{U}+\alpha W)|^{-1} \lesssim \min(\bar{U}, U)^{-1} \lesssim \langle z\rangle^{-\sigma+\epsilon/2}.$$

The first estimate with $\alpha = 1$ implies the upper bound for U in (3.3.16). Raising the above estimates to $|q|$-th power proves the second estimate in (3.3.16).

For the last estimate in (3.3.16), using triangle inequality, $|\nabla^i \bar{U}| \lesssim \langle z\rangle^{\sigma-i}$, and (3.3.10), we prove

$$||\bar{U}+\alpha W||_{\mathcal{H}^K} + ||\bar{U}+W||_{\mathcal{H}^K} \lesssim ||\bar{U}||_{\mathcal{H}^K} + ||W||_{\mathcal{H}^K} \lesssim 1 + E_0 + E_K \lesssim 1.$$

(b) Estimate of weights. We consider (3.3.17). Using $U \lesssim \langle z\rangle^{\sigma+\epsilon/2}$, clearly, we have

$$\begin{aligned} &|z| \le 1: \ \mathring{\rho}_i \lesssim \rho_i \lesssim \rho_i \langle z\rangle^{2\sigma+\epsilon-1}, \\ &|z| \ge 1: \ \mathring{\rho}_i \lesssim |z|^{2k-1-d} \lesssim \rho_i \langle z\rangle^{2\sigma+\epsilon-1}, i \le K-1, \quad \mathring{\rho}_K \lesssim U^2 \rho_K \lesssim \rho_K \langle z\rangle^{2\epsilon+\epsilon}. \end{aligned}$$

For $(f,\tau)=(\rho,-2\sigma-\epsilon)$ or $(\mathring{\rho},-1)$ and $i+j<K$, from the definition of f_i (3.1.28), we have

$$f_{i+j}\lesssim f_i\lesssim f_i\langle z\rangle^{2j},\quad |z|\le 1, f_{i+j}\approx |z|^{2i+2j+\tau-d}\lesssim |z|^{2i+\tau-d}\langle z\rangle^{2j}\lesssim f_i\langle z\rangle^{2j}, |z|\ge 1.$$

For $i+j=K$, the above estimate still holds for $f=\rho$. For $\mathring{\rho}_i$ and $\mathring{\rho}_K$, using $U\lesssim\langle z\rangle^{\sigma+\epsilon/2}$, we obtain

$$\mathring{\rho}_K\lesssim\langle z\rangle^{2\sigma+\epsilon}\rho_K\lesssim\langle z\rangle^{2K-1-d}=\langle z\rangle^{2i-1-d}\langle z\rangle^{2j}\lesssim\mathring{\rho}_i\langle z\rangle^{2j}.$$

We complete the proof of (3.3.17). □

Proposition 3.3.5. *Suppose that* (3.3.8) *holds true. For* $0\le k\le K$, $j_1+j_2\le k$ *and any* $\alpha\in[0,1]$, *denote*

$$V=\bar{U}+\alpha W,\quad I_{(j_1,j_2)}=\nabla^{j_1}W_1\cdot\nabla^{j_2}W_2\nabla^{k-j_1-j_2}V^{p-2}.$$

We have the following product estimates

$$||I_{(j_1,j_2)}||_{g_k}\lesssim||W_1||_{\mathcal{H}^{\max(j_1,K-1)}}||W_2||_{\mathcal{H}^{\max(j_2,K-1)}},\tag{3.3.18}$$

and

$$||W_1W_2V^{p-2}||_{\mathcal{H}^K}\lesssim(||W_1||_{\mathcal{H}^{K-1}}||W_2||_{\mathcal{H}^K}+||W_1||_{\mathcal{H}^K}||W_2||_{\mathcal{H}^{K-1}}),\tag{3.3.19a}$$

$$||W_1W_2V^{p-2}||_{\rho_0}\lesssim||W_2||_{\rho_0}||W_1||_{\mathcal{H}^d}.\tag{3.3.19b}$$

Moreover, we have

$$||\langle z\rangle^{\sigma+\epsilon/2}\nabla^{l+1}U\nabla^m(U^{-1})||_{g_k}\lesssim||U||_{\mathcal{H}^{\max(l+1,m,k)}},\quad l+m=k\le K,\tag{3.3.20}$$

$$|\nabla^{l+1}U\cdot\nabla^m(U^{-1})|\lesssim\langle z\rangle^{-l-m},\quad l,m\le K-1-d,\tag{3.3.21}$$

Proof. A direct computation yields

$$I_{j_1,j_2}\le\sum_{2\le q\le k+2,\,\sum_{l=1}^q j_l=k,j_l\ge 0}\sum I_{\vec{j},q},\quad I_{\vec{j},q}=|V|^{p-q}\cdot|\nabla^{j_1}W_1||\nabla^{j_2}W_2|\prod_{l=3}^{q}|\nabla^{j_l}V|.\tag{3.3.22}$$

Here $\vec{j}$ stands for the tuple $(j_1,j_2,\cdots,j_l)$. For a fixed (j,q), we denote

$$J_1=W_1,\quad J_2=W_2,\quad J_l=V,\quad l\ge 3,\quad i=\arg\max_{l\le q}j_l.$$

If there are more than one index a with $j_a=\arg\max_{l\le q}j_l$, we just pick one of them. Clearly, we have $j_l\le k/2, l\ne i$ (3.1.29). By Proposition 3.3.3 (3.3.13a), we have

$$|\nabla^{j_l}J_l|\lesssim\langle z\rangle^{-j_l+\sigma+\epsilon/2}||J_l||_{\mathcal{H}^{j_l+d}}$$

Applying the above L^∞ estimates to $J_l, l \neq i$, and Corollary 3.3.4 for $V = \bar{U} + \alpha W$, we obtain

$$I_{\vec{j},q} \lesssim |\nabla^{j_i} J_i| \prod_{l\neq i} \langle z\rangle^{-j_l+\sigma+\epsilon/2} ||J_l||_{\mathcal{H}^{j_l+d}} \langle z\rangle^{(p-q)\sigma+|p-q|\epsilon/2}. \tag{3.3.23}$$

Combining the exponents of the $\langle z\rangle^{\cdot}$ terms, and using the definitions of $\sigma = -\frac{2}{p-q}$ (3.2.10) and ϵ (3.1.30), we yield

$$\begin{aligned}\xi &= \sum_{1\leq l\neq i\leq q} (-j_l + \sigma + \epsilon/2) + (p-q)\sigma + |p-q|\epsilon/2 \\ &= -(k-j_i) + (p-q+q-1)\sigma + (p+q+q-1)\epsilon/2 = -(k-j_i) - 2 + (p+K)\epsilon < -(k-j_i).\end{aligned} \tag{3.3.24}$$

Since $\rho_k^{1/2}\langle z\rangle^{-(k-j_i)} \lesssim \rho_{k-j_i}$ (3.1.28), applying weighted L^2 bound to $\nabla^{j_i} J_i$, we further obtain

$$I_{\vec{j},q} \lesssim ||J_i||_{\mathcal{H}^{j_i}} \prod_{l\neq i} ||J_l||_{\mathcal{H}^{j_l+d}}.$$

Since $j_l + d \leq k/2 + d \leq K-1$ for $l \neq i, k \leq K$, $j_i \leq k$, and $||V||_{\mathcal{H}^K} \lesssim 1 + E_0 + E_K$, we obtain $j_l + d \leq \max(j_l, K-1)$ and thus

$$I_{\vec{j},q} \lesssim ||W_1||_{\mathcal{H}^{\max(j_1,K-1)}} ||W_2||_{\mathcal{H}^{\max(j_2,K-1)}}.$$

Using $\max(p-q,0)+q \leq p+K+q$ and summing the estimates of $I_{\vec{j},q}$, we conclude the proof of (3.3.18).

The estimate (3.3.19a) follows from summing the estimates of $I_{(j_1,j_2)}$ (3.3.18) over (j_1, j_2) with $j_1 + j_2 \leq K$, and using the fact that we have $j_1 \leq K-1$ or $j_2 \leq K-1$. The estimate (3.3.19b) follows from applying $L^2(\rho_0)$ estimate to W_2 and L^∞ estimate to W_1, V similar to the above.

For the last estimate (3.3.20), we note that $U = \bar{U} + W$. Applying the Leibniz rule, we obtain

$$|\nabla^{l+1} U \nabla^m (U^{-1})| \lesssim \sum_{1\leq q\leq k+1,\, \sum_{l=1}^q j_l = k+1, j_1\geq 1, j_l\geq 0} T_{\vec{j},q}, \quad T_{\vec{j},q} = |U|^{-q} \cdot \prod_{l=1}^q |\nabla^{j_l} U|.$$

Denote $i = \arg\max_l j_l$. Applying the above estimates of $I_{\vec{j},q}$ with $(J_1, .., J_l), V, p-q$ replaced by $(U, U, .., U), U, -q$ (3.3.22)-(3.3.24), and using $\max(-q, 0) = 0, q \leq K+1$, we obtain

$$|\langle z\rangle^{\sigma+\epsilon/2} T_{\vec{j},q}| \lesssim |\nabla^{j_i} U| \langle z\rangle^{\xi} \prod_{l\neq i} ||U||_{\mathcal{H}^{j_l+d}},$$

where

$$\xi = \sigma + \frac{\epsilon}{2} + \sum_{1\le l\neq i\le q}(-j_l + \sigma + \frac{\epsilon}{2}) + (-q)\sigma + |-q|\frac{\epsilon}{2}$$
$$= -(k+1-j_i) + (-q+q)\sigma + (q+q)\epsilon/2 = -(k-j_i) - 1 + (1+K)\epsilon < -(k-j_i).$$

Using $j_l+d \le K/2+d < K$, $||U||_{\mathcal{H}^{j_l+d}} \lesssim ||U||_{\mathcal{H}^K}$ for $l \neq i$, $\rho_k^{1/2} \lesssim \rho_{j_i}^{1/2}\langle z\rangle^{k-j_i}$, $q, j_i \le k+1$, and applying weighted L^2 estimate to $\nabla^{j_i}J_i$, we establish

$$||\langle z\rangle^{\sigma+\epsilon/2}T_{\vec{j},q}||_{g_k} \lesssim ||U||_{\mathcal{H}^{j_i}}(1+||U||_{\mathcal{H}^K})^{q-1} \lesssim ||U||_{\mathcal{H}^{k+1}}.$$

Combining the estimates for $T_{\vec{j},q}$ with different $\vec{j}, q$, we conclude the proof of (3.3.20).

For (3.3.21), denote $k = l+m$. Applying L^∞ estimate to each term $\nabla^{j_l}U$ in $T_{\vec{j},q}$ and noting that $j_l \le \max(l+1, m) \le K-d$, we prove

$$T_{\vec{j},q} \lesssim \langle z\rangle^{\xi} \lesssim \langle z\rangle^{-k-1/2},$$

where we have used

$$\xi = \sum_{1\le l\le q}(-j_l + \sigma + \frac{\epsilon}{2}) + (-q)\sigma + |-q|\frac{\epsilon}{2} = -(k+1) + q\epsilon < -k - 1/2.$$

We complete the proof. □

3.3.2 L^2 stability analysis of the amplitude

In this section, we estimate the weighted L^2 energy $E_0^2 = (W, W\rho_0)$ (3.3.4). In the following energy estimates, without specification, we will assume that the bootstrap Assumption 3.3.2 holds true, Namely

$$U \ge C_b\bar{U}^{1+\epsilon_2}, \quad C_b = \min_{|z|\le 1}\bar{U}^{-\epsilon_2}(z)/4 > 0,$$
$$\max(E_K, E_0, F_K, F_1) \le 1, \qquad \det(Q) > 0.$$

We will show that the following lemma holds.

Lemma 3.3.6 (Weighted L^2 estimate)**.** *Under the bootstrap assumption 3.3.2, it holds*

$$\frac{1}{2}\frac{d}{d\tau}E_0^2 \le (-\frac{\epsilon}{8} + C\mathcal{E}_1)E_0^2 + C\mathcal{E}_2E_0\,. \tag{3.3.25}$$

or some absolute constant $C > 0$.

Proof. Notice that W vanishes at the origin to the third order so this choice of singular weight induces a well-defined energy. We have by (3.3.1) that

$$\frac{1}{2}\frac{d}{d\tau}E_0^2 = (\mathscr{L}_{\bar{U}}W, W\rho_0) + (\mathscr{N}_U, W\rho_0) + (\mathscr{F}_U + \mathscr{D}_U, W\rho_0)\,.$$

For the leading order linear term, we have via integration by parts that

$$(\mathscr{L}_{\bar{U}}W, W\rho_0) = (d_0 W, W\rho_0)\,,$$

where we calculate the damping

$$\begin{aligned} d_0 &\coloneqq c_U + \frac{1}{2\rho_0}\nabla\cdot(d_{\bar{U}} z\rho_0) + p\bar{U}^{p-1} = -\frac{1}{p-1} + p\bar{U}^{p-1} + \frac{1}{4\rho_0}\nabla\cdot(z\rho_0) \\ &= -\frac{1}{p-1} + \frac{p}{p-1+c_p|z|^2} + \frac{(-6+\epsilon)|z|^{-6+\epsilon} + (\frac{4}{p-1}-\epsilon)c_0|z|^{\frac{4}{p-1}-\epsilon}}{4(|z|^{-6+\epsilon} + c_0|z|^{\frac{4}{p-1}-\epsilon})} \le -\frac{\epsilon}{8}\,. \end{aligned}$$

The last inequality amounts to

$$\begin{aligned} &4p(|z|^{-6+\epsilon} + c_0|z|^{\frac{4}{p-1}-\epsilon}) + ((-6 - \frac{4}{p-1} + \frac{3\epsilon}{2})|z|^{-6+\epsilon} - \frac{\epsilon}{2}c_0|z|^{\frac{4}{p-1}-\epsilon})(p-1+c_p|z|^2) \\ &\le -(p-1)|z|^{-6+\epsilon} - \frac{\epsilon}{2}c_0c_p|z|^{2+\frac{4}{p-1}-\epsilon} + 4pc_0|z|^{\frac{4}{p-1}-\epsilon} \le 0\,. \end{aligned}$$

The last inequality is implied by a weighted AM-GM inequality provided that

$$\Big(\frac{\epsilon c_0 c_p}{2(6+\frac{4}{p-1}-2\epsilon)}\Big)^{6+\frac{4}{p-1}-2\epsilon}(\frac{p-1}{2})^2 \ge \Big(\frac{4pc_0}{8+\frac{4}{p-1}-2\epsilon}\Big)^{8+\frac{4}{p-1}-2\epsilon}\,. \tag{3.3.26}$$

Notice that $\epsilon \le 1/2$ is fixed to be small. We can choose a sufficiently small constant $c_0 > 0$ such that we can conclude the linear estimate

$$\big(\mathscr{L}_{\bar{U}}W, W\rho_0\big) \le -\frac{\epsilon}{8}E_0^2\,. \tag{3.3.27}$$

The nonlinear estimate is more subtle due to the general nonlinearity p. We use Taylor's expansion or Newton-Leibniz's formula twice to derive

$$\mathscr{N}_U = W^2 p(p-1)\int_0^1 (1-\alpha)(\bar{U}+\alpha W)^{p-2}d\alpha\,. \tag{3.3.28}$$

Using Proposition 3.3.5 (3.3.19b) with $(W_1, W_2) = (W, W)$ and $\mathcal{E}_1$ defined in (3.3.11), we obtain

$$||W^2(\bar{U}+\alpha W)^{p-2}||_{\rho_0} \lesssim ||W||_{\mathcal{H}^d}||W||_{\rho_0} \lesssim (E_0 + E_d)E_0 \lesssim \mathcal{E}_1 E_0,$$

which implies

$$|(\mathcal{N}_U, W\rho_0)| \le \|\mathcal{N}_U\|_{\rho_0} E_0 \lesssim \mathcal{E}_1 ||W||_{\rho_0} E_0 \lesssim \mathcal{E}_1 E_0^2 \,. \tag{3.3.29}$$

Finally, we estimate the viscous and residue terms together. We group the terms to make them integrable. Consider a fixed 1D smooth cutoff function χ such that it equals 1 in $[-1,1]$ and 0 outside of $[-2,2]$. We use the notation $\widetilde{f}$ to denote functions only differing from f near the origin, where they equal the residue of f when expanded until its second-order Taylor's expansion at the origin, via the cutoff function χ. For illustrative purposes, we will explicitly write down $\widetilde{f}$ by the expansions

$$f = (f(0) + z^T \nabla f(0) + \frac{1}{2} z^T \nabla^2 f(0) z)\chi(|z|) + \widetilde{f} \,,$$

where $\widetilde{f}$ vanishes to the third order at the origin. By the choice of the modulation parameters in (3.2.11), it's easy to see that[6]

$$\mathscr{F}_U + \mathscr{D}_U = c_W \widetilde{U} - \mathcal{P} z \cdot \nabla \widetilde{U} - v \cdot \widetilde{\nabla U} - H^{p-1} \gamma \widetilde{U} + \widetilde{\mathscr{D}_U} := \widetilde{\mathscr{F}_U} + \widetilde{\mathscr{D}_U} \,. \tag{3.3.30}$$

Each of the terms in $\widetilde{\mathscr{F}_U}$ vanishes to the third order at the origin. Notice that $\rho_k = \rho_0 |z|^{2k}$, for $k = 1, 2, 3$. Recall the definition of $\mathcal{E}_1, \mathcal{E}_2$ from (3.3.11). By Lemma 3.2.1, we have that

$$\|\widetilde{\mathscr{F}_U}\|_{\rho_0} \lesssim \mathcal{E}_2 (1 + E_0 + E_1 + \|\widetilde{\nabla W}\|_{\rho_0}) \,.$$

We can decompose the integral region into the near field $I = [0,1]^d$ and the rest of the outer region I^c to estimate

$$\|\widetilde{\nabla W}\|_{\rho_0} \lesssim (\int_{z \in I} |z|^{\epsilon - d})^{1/2} \sup_{z \in I} |\widetilde{\nabla W}/|z|^3| + \|\widetilde{\nabla W}|z|\|_{\rho_0} \lesssim \Gamma + E_1 \,.$$

Here we recall the definition of Γ, E_k in (3.2.13) and (3.3.4). Combined with Proposition 3.3.3, we have the residue estimate

$$|(\widetilde{\mathscr{F}_U}, W\rho_0)| \lesssim \mathcal{E}_2 E_0 \,. \tag{3.3.31}$$

For the viscous term, we notice as in (3.2.16), we can write

$$\widetilde{\mathscr{D}_U} = \mathrm{tr}(Q \widetilde{S_1}) \,, \quad S_1 = S_{11} + S_{12} + S_{13} + S_{14} \,.$$

[6] Note that $\tilde{U}$ does not denote the perturbation.

We estimate the four terms respectively. We compute

$$\|\widetilde{S_{11}}\|_{\rho_0} = \|\widetilde{\nabla^2 U}\|_{\rho_0} \lesssim 1 + \|\widetilde{\nabla^2 W}\|_{\rho_0} \lesssim 1 + \Gamma + E_2\,,$$

where in the last inequality we use again the decomposition of the integral into the near and far fields. For the remaining three viscous terms, we estimate similarly as follows:

$$\|\widetilde{S_{12}}\|_{\rho_0} = 2|\beta|\|\widetilde{\nabla U \nabla \Theta^T}\|_{\rho_0} \lesssim 2|\beta|\|\widetilde{\nabla W}\widetilde{\nabla \Phi^T}\|_{\rho_0} + (1+\Gamma)^2 \lesssim \Gamma(\Gamma + E_1) + (1+\Gamma)^2\,,$$
$$\|\widetilde{S_{13}}\|_{\rho_0} = \|\widetilde{U\nabla\Theta\nabla\Theta^T}\|_{\rho_0} \lesssim \|\widetilde{W}\widetilde{\nabla\Phi}\widetilde{\nabla\Phi^T}\|_{\rho_0} + (1+\Gamma)^3 \lesssim \Gamma^2(\Gamma + E_0) + (1+\Gamma)^3\,,$$
$$\|\widetilde{S_{14}}\|_{\rho_0} = \|\widetilde{U\nabla^2\Theta}\|_{\rho_0} \lesssim \|\widetilde{W}\widetilde{\nabla^2\Phi}\|_{\rho_0} + (1+\Gamma)^2 \lesssim \Gamma(\Gamma + E_0) + (1+\Gamma)^2.$$

We can collect the viscous estimate by Proposition 3.3.3 and Assumption 3.3.2:

$$|(\widetilde{\mathscr{D}_U}, W\rho_0)| \lesssim \mathcal{E}_2(1+E_0+E_1+E_2+\Gamma)^3 E_0 \lesssim \mathcal{E}_2(1+E_0+E_K)^3 E_0 \lesssim \mathcal{E}_2 E_0\,. \quad (3.3.32)$$

We thereby conclude the proof of Lemma 3.3.6 using (3.3.27), (3.3.29), (3.3.31), and (3.3.32). □

One sees that we already have leading order damping in the L^2 estimates. However, to close the nonlinear estimates, we will need higher order estimates to control the L^∞ norms.

3.3.3 H^1 stability analysis of the phase

We consider the weighted H^1 norm of the phase $F_1^2 = (\nabla\Phi, \nabla\Phi\mathring{\rho}_1)$ (3.3.4). We choose this norm since Φ does not decay at the origin. We will show that the following lemma holds.

Lemma 3.3.7 (Weighted H^1 estimate)**.** *Under the bootstrap assumption 3.3.2, it holds*

$$\frac{1}{2}\frac{d}{d\tau}F_1^2 \le -\frac{1}{8}F_1^2 + C(E_{K-1}+E_0)^2 + C\mathcal{E}_2 F_1\,, \quad (3.3.33)$$

for some absolute constant $C > 0$.

Proof. We have by (3.3.2) that

$$\frac{1}{2}\frac{d}{d\tau}F_1^2 = (\nabla(-\frac{1}{2}\Lambda\Phi) + \nabla\mathscr{N}_\Theta + \nabla\mathscr{F}_\Theta + \nabla\mathscr{D}_\Theta, \nabla\Phi\mathring{\rho}_1)\,.$$

For the leading order linear term, we have via integration by parts that

$$(\partial_i(-\frac{1}{2}\Lambda\Phi), \partial_i\Phi\mathring{\rho}_1) = -\frac{1}{4}(\partial_i\Phi, \partial_i\Phi\mathring{\rho}_1)\,.$$

Therefore we have the linear estimate

$$(\nabla(-\frac{1}{2}\Lambda\Phi), \nabla\Phi\mathring{\rho}_1) = -\frac{1}{4}F_1^2\,. \tag{3.3.34}$$

For the nonlinear estimate, we again use Newton-Leibniz's formula to get

$$\mathcal{N}_\Theta = \delta((\bar{U}+W)^{p-1} - \bar{U}^{p-1}) = \delta(p-1)W\int_0^1 (\bar{U}+\alpha W)^{p-2}d\alpha\,. \tag{3.3.35}$$

It is not difficult to see that $\langle z\rangle^{\sigma+\frac{\epsilon-1}{2}} \in \mathcal{H}^i$ for any $i \geq 0$ since $\epsilon - 1 < 0$. Note that $\mathring{\rho}_1$ (3.1.28) is locally integrable and $\mathring{\rho}_1 \lesssim \langle z\rangle^{\sigma+\frac{\epsilon-1}{2}} g_1$. Applying L^∞ estimate for $W, \bar{U}+\alpha W$ from Corollary 3.3.4, (3.3.13a) in Proposition 3.3.3, and Proposition 3.3.5 (3.3.18) with $W_1 = \langle z\rangle^{\sigma+\frac{\epsilon-1}{2}}, W_2 = W, j_1 = 0, j_2 = 1, k = 1$, we obtain

$$\begin{aligned}||\nabla\mathcal{N}_\Theta||_{\mathring{\rho}_1} &\lesssim ||\nabla\mathcal{N}_\Theta||_{L^\infty(|z|\leq 1)} + ||W_1\nabla\mathcal{N}_\Theta||_{g_1} \lesssim E_0 + E_{K-1} + ||W_1\nabla\mathcal{N}_\Theta||_{\rho_1}\\ &\lesssim E_0 + E_{K-1} + ||W_2||_{\mathcal{H}^{K-1}} \lesssim E_0 + E_{K-1}.\end{aligned}$$

We can collect the nonlinear estimate, via an AM-GM inequality as follows:

$$|(\nabla\mathcal{N}_\Theta, \nabla\Phi\mathring{\rho}_1)| \leq C\|\nabla\mathcal{N}_\Theta\|_{\mathring{\rho}_1}F_1 \leq C(E_0+E_{K-1})^2 + \frac{1}{8}F_1^2\,. \tag{3.3.36}$$

For the residue estimate, we have

$$|(\nabla\mathscr{F}_\Theta, \nabla\Phi\mathring{\rho}_1)| \leq \|\nabla\mathscr{F}_\Theta\|_{\mathring{\rho}_1}F_1 \lesssim \mathcal{E}_2(1+F_1+\|\nabla^2\Phi\|_{\mathring{\rho}_1})F_1 \lesssim \mathcal{E}_2F_1\,. \tag{3.3.37}$$

For the viscous estimate, we have

$$\begin{aligned}|\nabla\mathscr{D}_\Theta\|_{\mathring{\rho}_1} \lesssim \mathcal{E}_2\Big(&\|\frac{\nabla^3U}{U}\|_{\mathring{\rho}_1} + \|\frac{\nabla^2U}{U}\|_{\mathring{\rho}_1}\|\frac{\nabla U}{U}\|_\infty + \|\frac{\nabla U}{U}\|_\infty\|\nabla^2\Theta\|_{\mathring{\rho}_1}\\ &+ (\|\frac{\nabla^2U}{U}\|_\infty + \|\frac{\nabla U}{U}\|_\infty^2)\|\nabla\Theta\|_{\mathring{\rho}_1} + 1 + F_1 + \|\nabla^3\Phi\|_{\mathring{\rho}_1}\Big)\,.\end{aligned} \tag{3.3.38}$$

To estimate the integral $L^2(\mathring{\rho}_1)$, we apply L^∞ estimate in the region $|z| \leq 1$ and (3.3.20) and $\mathring{\rho}_1 \lesssim \langle z\rangle^{\sigma+\epsilon/2}g_1 \lesssim \langle z\rangle^{\sigma+\epsilon/2}g_2$ to the region $|z| \geq 1$:

$$||\nabla^l U/U||_{\mathring{\rho}_1} \lesssim ||\nabla^l U/U||_{L^\infty(|z|\leq 1)} + ||\langle z\rangle^{\sigma+\epsilon/2}\nabla^l U/U||_{g_1} \lesssim 1 + E_0 + E_K \lesssim 1, \quad l = 2, 3.$$

Applying (3.3.21) with $(l, m) = (1, 0), (0, 0)$, we get

$$|\nabla^{l+1}U/U| \lesssim 1. \tag{3.3.39}$$

As a consequence, we can simplify the viscous estimate as follows:

$$|\nabla\mathscr{D}_\Theta\|_{\mathring{\rho}_1} \lesssim \mathcal{E}_2(1+F_1+\|\nabla^2\Phi\|_{\mathring{\rho}_1}+\|\nabla^3\Phi\|_{\mathring{\rho}_1})\,. \tag{3.3.40}$$

Finally, since $\mathring{\rho}_1$ is L^1 integrable and $\mathring{\rho}_i \lesssim \langle z\rangle^{2i-2}\mathring{\rho}_1$ (3.3.17), we can decompose the integral region into $I = [0,1]^d$ and the rest of the outer region I^c as in the L^2 estimate of the amplitude to compute

$$\|\nabla^l \Phi\|_{\mathring{\rho}_1} \lesssim \sup_{z\in I} |\nabla^l \Phi| + \|\nabla^l \Phi |z|^{2l-2}\|_{\mathring{\rho}_1} \lesssim \Gamma + F_l\,, \quad ||\nabla^l \bar{\Theta}||_{\mathring{\rho}_1} \lesssim 1, \quad l = 2,3\,.$$

We use Proposition 3.3.3 and the bootstrap assumption (3.3.8) to further obtain

$$\|\nabla^l \Phi\|_{\mathring{\rho}_1} + ||\nabla \bar{\Theta}||_{\mathring{\rho}_1} \lesssim 1 + F_1 + F_K \lesssim 1.$$

Here we recall the definition of Γ, F_k in (3.2.13) and (3.3.4). Plugging in the estimate in (3.3.40) and combined with (3.3.34), (3.3.36), and (3.3.37), we establish Lemma 3.3.7. □

3.3.4 H^K stability analysis

For the estimate at the highest order, we consider the weighted H^K energies (3.3.4)

$$E_K^2 = (|\nabla^K W|^2, \rho_K)\,, \quad F_K^2 = (|\nabla^K \Phi|^2, U^2\rho_K)\,.$$

In this section, we will establish the following lemma.

Lemma 3.3.8 (Weighted H^K estimate)**.** *Under the bootstrap assumption 3.3.2, we have*

$$\frac{1}{2}\frac{d}{d\tau}(E_K^2 + F_K^2) \le -\frac{\epsilon}{8}(E_K^2 + F_K^2) + \mu_0 \mathcal{E}_1 (E_K + F_K) \tag{3.3.41}$$

for some absolute constant $\mu_0 > 0$.

3.3.4.1 Estimates of the amplitude

Recall the definitions of $\mathscr{L}_{\bar{U}}, \mathscr{N}_U, \mathscr{F}_U$ from (3.3.3). We have

$$\frac{1}{2}\frac{d}{d\tau}E_K^2 = (\nabla^K(\mathscr{L}_{\bar{U}}W) + \nabla^K \mathscr{N}_U + \nabla^K \mathscr{F}_U + \nabla^K \mathscr{D}_U, \nabla^K W \rho_K)\,.$$

For the leading order linear term, we can calculate the damping similarly as in the L^2 estimates. A direct computation yields $\bar{U}^{p-1} \in \mathcal{H}^i$ (3.1.27) for any $i \ge 0$. Using the Leibniz rule, the product rule (3.3.14) in Proposition 3.3.3 with $i + j = K, j \le K-1, m = K-1, n = i+d$, and $g_K \approx \rho_K$ (3.1.28), (3.3.9), we yield

$$||\nabla^K(\bar{U}^{p-1}W) - \bar{U}^{p-1}\nabla^K W||_{\rho_K} \lesssim \sum_{j\le K-1} \nabla^{K-j}||\bar{U}^{p-1}||_{\mathcal{H}^{i+d}}||W||_{\mathcal{H}^j} \lesssim \sum_{j\le K-1} E_j \lesssim \mathcal{E}_1.$$

Therefore, we can compute

$$\nabla^K(\mathscr{L}_{\bar{U}}W) = c_{\bar{U}}\nabla^K W - \frac{1}{2}\sum_i z_i \nabla^K \partial_i W - \frac{K}{2}\nabla^K W + p\bar{U}^{p-1}\nabla^K W + O(\mathcal{R}_{\mathscr{L},K}), \quad ||\mathcal{R}_{\mathscr{L},K}||_{\rho_K} \lesssim \mathcal{E}_1.$$

We can calculate the damping similar to the L^2 case as follows:

$$d_K \coloneqq -\frac{1}{p-1} - \frac{K}{2} + p\bar{U}^{p-1} + \frac{1}{4\rho_K}\nabla\cdot(z\rho_K)$$
$$= \frac{p}{p-1+c_p|z|^2} - \frac{1}{p-1} - \frac{K}{2} + \frac{d + (2K + \frac{4}{p-1} - \epsilon)c_1|z|^{\frac{4}{p-1}-\epsilon-d+2K}}{4(1+c_1|z|^{\frac{4}{p-1}-\epsilon-d+2K})} \leq -\frac{\epsilon}{8},$$

where the last inequality holds for a sufficiently small c_1, which we defer till (3.3.47) where we combine this damping with the estimates of the nonlinear term in the phase equation.

For the nonlinear term, we use Netwon-Leibniz's formula twice as in the L^2 estimate (3.3.28), to derive

$$|(\nabla^K\mathcal{N}_U, \nabla^K W\rho_K)| \lesssim \sup_{\alpha\in[0,1]}(1-\alpha)\|\nabla^K(W^2(\bar{U}+\alpha W)^{p-2})\|_{\rho_K}E_K\,.$$

Since $||f||_{\rho_K} \lesssim ||f||_{\mathcal{H}^K}$ (3.1.28), (3.3.9), using the product estimate (3.3.19a) with $(W_1, W_2) = (W, W)$ and $||f||_{\mathcal{H}^K} \lesssim E_0 + E_K$ (3.3.10), we obtain

$$|(\nabla^K\mathcal{N}_U, \nabla^K W\rho_K)| \lesssim E_K||W||_{\mathcal{H}^K}||W||_{\mathcal{H}^{K-1}} \lesssim E_K(E_K+E_0)(E_0+E_{K-1}) \lesssim E_K\mathcal{E}_1. \tag{3.3.42}$$

Recall $\mathcal{E}_2$ from (3.3.11). For the residue term, we have via integration by parts that

$$|(\nabla^K\mathcal{F}_U, \nabla^K W\rho_K)| \lesssim \mathcal{E}_2(E_K^2 + E_K + (|\nabla^K W|^2, |z\cdot\nabla\rho_K| + |\nabla\rho_K|))\,.$$

Since we have $|\nabla\rho_K|\langle z\rangle \lesssim \rho_K$, we can conclude the residue estimate

$$|(\nabla^K\mathcal{F}_U, \nabla^K W\rho_K)| \lesssim \mathcal{E}_2(E_K + E_K^2) \lesssim \mathcal{E}_1E_K\,. \tag{3.3.43}$$

3.3.4.2 Estimates of the phase

We have

$$\frac{1}{2}\frac{d}{d\tau}F_K^2 = \left(\nabla^K(-\frac{1}{2}\Lambda\Phi) + \nabla^K\mathcal{N}_\Theta + \nabla^K\mathcal{F}_\Theta + \nabla^K\mathcal{D}_\Theta, \nabla^K\Phi U^2\rho_K\right)$$
$$+ \left(\mathscr{L}_{\bar{U}}W + \mathcal{N}_U + \mathcal{F}_U + \mathcal{D}_U, U|\nabla^K\Phi|^2\rho_K\right).$$

Notice that the weight is time-dependent. We remark that it is essential to pair the two linear terms and the two residue terms together to cancel out the leading order term via integration by parts. For the leading order linear term, we have via integration by parts that

$$(\nabla^K(-\frac{1}{2}\Lambda\Phi), \nabla^K\Phi U^2\rho_K) + (\mathscr{L}_{\bar{U}}W + \mathcal{N}_U, U|\nabla^K\Phi|^2\rho_K) = (\mathring{d}_K, |\nabla^K\Phi|^2U^2\rho_K)\,,$$

where we can calculate the damping

$$\mathring{d}_K = \frac{-K}{2} + \frac{1}{4\rho_K}\nabla\cdot(z\rho_K) - \frac{1}{p-1} + \frac{U^p}{U} < d_K + U^{p-1} - \bar{U}^{p-1}\,.$$

Notice that by (3.3.28), (3.3.13a) in Proposition 3.3.3 with $l = 0$, and Corollary 3.3.4, we can further estimate

$$|U^{p-1} - \bar{U}^{p-1}| \lesssim \sup_{0\le\alpha\le 1} |W(\bar{U}+\alpha W)^{p-2}| \lesssim (E_0 + E_{K-1})(1 + E_K + E_0)^{p+2}) \lesssim \mathcal{E}_1. \tag{3.3.44}$$

For the residue term, similarly via integration by parts, we have

$$\begin{aligned}|(\nabla^K \mathcal{F}_\Theta, \nabla^K\Phi U^2\rho_K)| + |(\mathcal{F}_U, U|\nabla^K\Phi|^2\rho_K)| &\lesssim \mathcal{E}_2(F_K + F_K^2 + (\frac{\nabla\cdot((Pz+v)\rho_K)}{2\rho_K}, |\nabla^K\Phi|^2U^2\rho_K))\\ &\lesssim \mathcal{E}_2(F_K + F_K^2) \lesssim \mathcal{E}_1 F_K\,,\end{aligned} \tag{3.3.45}$$

where the inequality is again by the fact that $|\nabla\rho_K|\langle z\rangle \lesssim \rho_K$.

For the nonlinear term, using Newton-Leibniz's formula (3.3.35), we obtain

$$|\nabla^K\mathcal{N}_\Theta| \le I_{0,K} + C\sum_{j\le K-1} I_{0,j}, \quad I_{i,j} = \delta(p-1)\cdot\nabla^j W\cdot\nabla^{K-j}(U+\alpha\bar{W}).$$

Applying (3.3.18) in Proposition 3.3.5 with $(W_1, W_2, j_1, j_2, k) = (U, W, 0, j, K-j)$ and $\bar{U}\in\mathcal{H}^i$ (3.3.16), we obtain

$$||UI_{0,j}||_{\rho_K} \lesssim ||U||_{\mathcal{H}^K}||W||_{\mathcal{H}^{K-1}} \lesssim E_0 + E_{K-1} \lesssim \mathcal{E}_1.$$

Recall $\mathring{\rho}_K = U^2\rho_K$ (3.1.28). For $j \le K-1$, the above estimate implies

$$||I_{0,j}||_{\mathring{\rho}_K} = ||UI_{0,j}||_{\rho_K} \lesssim \mathcal{E}_1, \quad |(I_{0,j}, \nabla^K\Phi\mathring{\rho}_K)| \lesssim ||I_{0,j}||_{\mathring{\rho}_K}F_K \lesssim \mathcal{E}_1F_K. \tag{3.3.46}$$

The term $I_{0,K}$ is trickier and we need to estimate by an AM-GM inequality:

$$((\bar{U}+\alpha W)^{p-2}\nabla^K W, \nabla^K\Phi U^2\rho_K) \le \frac{1}{2}\|(U^{\frac{1}{2}}(\bar{U}+\alpha W)^{\frac{p-2}{2}}\nabla^K\Phi\|^2_{\mathring{\rho}_K} + \frac{1}{2}\|U^{\frac{1}{2}}(\bar{U}+\alpha W)^{\frac{p-2}{2}}\nabla^K W\|^2_{\rho_K}\,,$$

where we pair one of U in U^2 with $\rho_K^{1/2}$ to get $\mathring{\rho}_K^{1/2}$. Applying $U = (1-\alpha)W + \bar{U} + \alpha W$, Newton-Leibniz's rule for $(\bar{U}+\alpha W)^{p-1} - \bar{U}^{p-1}$, Proposition 3.3.3 for W, and Corollary 3.3.4 for $\bar{U}+sW, s\in[0,1]$, which are similar to the estimate of $\mathcal{N}_\Theta$ (3.3.35), we obtain

$$\begin{aligned}U(\bar{U}+\alpha W)^{p-2} &\le (\bar{U}+\alpha W)^{p-1} + C|W(\bar{U}+\alpha W)^{p-2}| \le \bar{U}^{p-1} + C\sup_{s\in[0,1]}|W(\bar{U}+sW)^{p-2}\\ &\le \bar{U}^{p-1} + C(E_{K-1}+E_0) \le (\min\{p-1, c_p\})^{-1}\langle z\rangle^{-2} + C\mathcal{E}_1,\end{aligned}$$

where we extract a decay at the far field for the leading order term. We can calculate the damping

$$\begin{aligned}&\delta(p-1)(\min\{p-1,c_p\})^{-1}\langle z\rangle^{-2}+d_K\\&\leq\frac{\delta(p-1)(p+1)}{\min\{p-1,c_p\}(1+|z|^2)}-\frac{1}{p-1}-\frac{K}{2}+\frac{d+(2K+\frac{4}{p-1}-\epsilon)c_1|z|^{\frac{4}{p-1}-\epsilon-d+2K}}{4(1+c_1|z|^{\frac{4}{p-1}-\epsilon-d+2K})}.\end{aligned}$$

Recall the definition of K in (3.1.29) and similar to the L^2 damping, we can use a weighted AM-GM inequality to conclude for a sufficiently small positive c_1, we have

$$(p-1)\delta(\min\{p-1,c_p\})^{-1/2}\langle z\rangle^{-1/2}+d_K\leq-\frac{\epsilon}{8}. \tag{3.3.47}$$

As a consequence, we collect the linear and nonlinear estimates of the phase, and the linear estimate of the amplitude together as follows:

$$\begin{aligned}&\left(\nabla^K(\mathscr{L}_{\bar{U}}W),\nabla^K W\rho_K\right)+\left(\nabla^K(-\frac{1}{2}\Lambda\Phi+\mathscr{N}_\Theta)+\frac{\mathscr{L}_{\bar{U}}W+\mathscr{N}_U}{U},\nabla^K\Phi U^2\rho_K\right)\\&\qquad\leq-\frac{\epsilon}{8}(E_K^2+F_K^2)+C\mathcal{E}_1(E_K+F_K).\end{aligned} \tag{3.3.48}$$

3.3.4.3 Estimates of the viscous terms

Finally, we estimate the viscous terms. The simpler term can be estimated as follows:

$$(\mathscr{D}_U,U|\nabla^K\Phi|^2\rho_K)\leq\|\frac{\mathscr{D}_U}{U}\|_\infty F_K^2\lesssim\mathcal{E}_1F_K. \tag{3.3.49}$$

The last inequality is derived similarly to the H^1 viscous estimates in (3.3.38), (3.3.39).

We group leading order viscous terms as follows and estimate them together:

$$(\nabla^K\mathscr{D}_U,\nabla^K W\rho_K)+(\nabla^K\mathscr{D}_\Theta,\nabla^K\Phi U^2\rho_K),$$

and we will use integration by parts to cancel out the leading order terms and extract damping. Recall the definition of the viscous terms in (3.2.9). For any tensor f, we define

$$|f|_Q^2=\sum_i(\nabla f_i)^TQ\nabla f_i,$$

where we sum over its scalar entry components f_i.

Notice that $|\nabla\rho_K| \lesssim \rho_K$. We compute the damping of the amplitude using integration by parts and the Cauchy-Schwarz inequality as

$$\begin{aligned}(\nabla^K\Delta_Q U, \nabla^K W\rho_K) &\le C\mathcal{E}_1 E_K - (|\nabla^K W|_Q^2, \rho_K) + C|Q|^{1/2}\|\nabla^K W\|_{\rho_K}\||\nabla^K W|_Q\|_{\rho_K}\\ &\le C\mathcal{E}_1 E_K - \frac{1}{2}(|\nabla^K W|_Q^2, \rho_K)\,.\end{aligned}$$

Using (3.3.13a) from Proposition 3.3.3 and (3.3.16), we get

$$|\nabla U| \lesssim U, \quad |\nabla(U^2\rho_K)| \lesssim |\nabla U|U\rho_K + U^2|\nabla\rho_K| \lesssim U^2\rho_K.$$

Similarly, we compute the damping of the phase as

$$(\nabla^K\Delta_Q\Theta, \nabla^K\Phi U^2\rho_K) \le C\mathcal{E}_1 F_K - \frac{1}{2}(|\nabla^K\Phi|_Q^2, U^2\rho_K)\,.$$

For the four intermediate terms in the viscous terms

$$I_1 = \langle\nabla U, \nabla\Theta\rangle_Q\,, \quad I_2 = U\langle\nabla\Theta, \nabla\Theta\rangle_Q\,, \quad I_3 = \frac{\langle\nabla U, \nabla\Theta\rangle_Q}{U}\,, \quad I_4 = \langle\nabla\Theta, \nabla\Theta\rangle_Q\,,$$

we can simply control their weighted norms using the diffusion term.

We consider the most challenging term I_3. Using the Leibniz rule, (3.3.20) in Proposition 3.3.5, we obtain

$$||I_3||_{\mathring{\rho}_K} \lesssim \sum_{0\le i\le K} ||I_{3,i}||_{\mathring{\rho}_K}, \quad I_{3,i} = (\nabla^i\frac{\nabla U}{U}, \nabla^{K-i+1}\Theta)_Q.$$

For $1 \le i \le K-1$, applying (3.3.21) to $\nabla U/U$ if $i \le K/2+1 < K-d-1$ and (3.3.12) to Θ if $i > K/2$, which implies $K-i+1 \le K/2+1 < K-d-1$, we obtain

$$|I_{3,i}| \lesssim |Q|(\langle z\rangle^{-i}|\nabla^{K-i+1}\Theta| + \langle z\rangle^{-(K-i)}|\nabla^i(\nabla U/U)|).$$

Since $i \le K-1, K-i+1 \le K$, using the estimate (3.3.17) for weights

$$\mathring{\rho}_K^{1/2} \lesssim \langle z\rangle^{i-1}\mathring{\rho}_{K-i+1}^{1/2}, \quad \langle z\rangle^{-(K-i)}\mathring{\rho}_K^{1/2} \lesssim \mathring{\rho}_i^{1/2} \lesssim \langle z\rangle^{\sigma+\epsilon/2}\rho_i^{1/2},$$

and (3.3.20), we obtain

$$||I_{3,i}||_{\mathring{\rho}_K} \lesssim |Q|(||\nabla^{K-i+1}\Theta||_{\mathring{\rho}_{K-i}} + ||\langle z\rangle^{\sigma+\epsilon/2}\nabla^i(\nabla U/U)||_{\rho_i}) \lesssim |Q| \lesssim \mathcal{E}_1.$$

For $I_{3,0}, I_{3,K}$, we use the Cauchy-Schwarz inequality to compute that its ρ_K norm is bounded by

$$\mathcal{E}_1 + |Q|^{1/2}(\|\nabla\Theta\|_\infty\||\nabla^K W|_Q\|_{\rho_K} + \|\frac{\nabla U}{U}\|_\infty\|U|\nabla^K\Phi|_Q\|_{\rho_K})\,.$$

Similarly, we have the estimates for the other three viscous terms I_1, I_2, I_4. Combined with (3.3.39), we can use the Cauchy-Schwarz inequality to derive that

$$\Big(-2\beta\nabla^K\langle\nabla U,\nabla\Theta\rangle_Q - \nabla^K(U\langle\nabla\Theta,\nabla\Theta\rangle_Q),\nabla^K W\rho_K\Big) + \Big(2\nabla^K\frac{\langle\nabla U,\nabla\Theta\rangle_Q}{U} - \beta\nabla^K\langle\nabla\Theta,\nabla\Theta\rangle_Q,\nabla^K\Phi U^2\rho_K\Big)$$
$$\leq C\mathcal{E}_1(E_K+F_K) + \frac{1}{8}((|\nabla^K W|_Q^2,\rho_K) + (|\nabla^K\Phi|_Q^2,U^2\rho_K))\,.$$

Finally, for the last two viscous terms, we use integration by parts to cancel out the leading order terms. Applying estimates similar to the those for I_3 in the above, we can extract the leading order terms, which involve $\nabla^{K+2}U$ or $\nabla^{K+2}\Theta$,

$$-\beta\Big(\nabla^K(U\Delta_Q\Theta),\nabla^K W\rho_K\Big) = -\beta(U\Delta_Q\nabla^K\Theta,\nabla^K W\rho_K) + O(\mathcal{E}_1E_K) + \frac{1}{16}(|\nabla^K\Phi|_Q^2,U^2\rho_K),$$
$$\beta\Big(\nabla^K\frac{\Delta_Q U}{U},\nabla^K\Phi U^2\rho_K\Big) = \beta\Big(\frac{\Delta_Q\nabla^K U}{U},\nabla^K\Phi U^2\rho_K\Big) + O(\mathcal{E}_1F_K) + \frac{1}{16}(|\nabla^K U|_Q^2,\rho_K).$$

Now, we use $U=\bar{U}+W,\Theta=\bar{\Theta}+\Phi$ and integration by parts to cancel out the leading order terms.

$$\begin{aligned}
&-(\nabla^K\Delta_Q\Theta,\nabla^K WU\rho_K) + (\nabla^K\Delta_Q U,\nabla^K\Phi U\rho_K)\\
=&-(\nabla^K\Delta_Q\Phi,\nabla^K WU\rho_K) + (\nabla^K\Delta_Q W,\nabla^K\Phi U\rho_K) + O(\mathcal{E}_1(E_K+F_K))\\
=&\sum_{i,j}\Big(\mathcal{Q}_{ij}(-\partial_i(\partial_j\nabla^K\Phi\cdot\nabla^K W) + \partial_j(\partial_i\nabla^K W\cdot\nabla^K\Phi)),U\rho_K\Big)\\
\leq& C\mathcal{E}_1(E_K+F_K) + \frac{1}{16\beta}((|\nabla^K W|_Q^2,\rho_K) + (|\nabla^K\Phi|_Q^2,U^2\rho_K))\,.
\end{aligned}$$

We notice that the remaining terms from integration by parts are controlled since $|\nabla(U\rho_K)|\lesssim U\rho_K$.

Combining the viscous estimates with the estimates (3.3.42), (3.3.43), (3.3.45), and (3.3.48), we conclude the proof of Lemma 3.3.8.

3.3.4.4 Summary of the H^K estimates

Using (3.3.12a), (3.3.12b) in Proposition 3.3.3, for any $\mu>0$, we obtain

$$\mathcal{E}_1 \leq |\mathcal{Q}| + H^{p-1} + C(\mu)(E_0+F_1) + \mu(E_K+F_K) + (E_0+F_1+E_K+F_K)^2.$$

By Lemma 3.3.8, choosing $\mu<\frac{\epsilon}{16\mu_0}$ and then collecting (3.3.41), (3.3.33), and (3.3.25), we obtain that there exists a sufficiently small constant $1>\nu_1>\nu_2>0$, ν_2 determined after ν_1, such that for the energy

$$E^2 = E_K^2 + F_K^2 + 1/\nu_1 F_1^2 + 1/\nu_2 E_0^2\,, \tag{3.3.50}$$

the following estimate holds

$$\frac{1}{2}\frac{d}{d\tau}E^2 \le -\frac{\epsilon}{16}E^2+C(|Q|+H^{p-1})E+CE^3\,,\ \Longleftrightarrow\ \frac{d}{d\tau}E \le -\frac{\epsilon}{16}E+\mu_1(|Q|+H^{p-1})+\mu_1E^2\,, \tag{3.3.51}$$

for some absolute constant $\mu_1 > 0$. Here, the constant C would depend on ν_1, ν_2. Once we fix ν_1, ν_2, then C becomes a fixed constant μ_1. The estimate holds provided that Assumption 3.3.2 is valid.

3.3.5 Lower bound of the amplitude

We now prove the bootstrap Assumption 3.3.2 by estimating the lower bound of $U\rho$, for the weight $\rho = \bar{U}^{-1-\epsilon_2}$. We will proceed with a maximal principle argument and a barrier argument. Notice that

$$\nabla U = \frac{\nabla(U\rho) - U\nabla\rho}{\rho}\,,\quad \nabla^2 U = \frac{\nabla^2(U\rho) - U\nabla^2\rho}{\rho} - \frac{\nabla(U\rho)\nabla\rho^T + \nabla\rho\nabla(U\rho)^T - 2U\nabla\rho\nabla\rho^T}{\rho^2}\,.$$

We compute by (3.2.7) that

$$\partial_\tau(U\rho) = \mathcal{P}_U(U\rho),\quad \mathcal{P}_U f = A_0 f + A_1\cdot\nabla f + \mathrm{tr}(Q\nabla^2 f). \tag{3.3.52}$$

where the coefficients A_0, A_1 of the parabolic operator $\mathcal{P}_U$ are:

$$A_0 = c_U - H^{p-1}\gamma + U^{p-1} + (\frac{1}{2}z + \mathcal{P}z + \mathcal{V})\cdot\frac{\nabla\rho}{\rho} - \frac{\Delta_Q\rho - 2\beta\langle\nabla\rho,\nabla\Theta\rangle_Q}{\rho} + 2\frac{\langle\nabla\rho,\nabla\rho\rangle_Q}{\rho^2} - \langle\nabla\Theta,\nabla\Theta\rangle_Q -$$

$$A_1 = -(\frac{1}{2}z + \mathcal{P}z + \mathcal{V} + 2\frac{Q\nabla\rho}{\rho} + 2\beta Q),$$

Notice that $\bar{U}$ is the approximate steady state and $|\nabla\rho|\langle z\rangle \lesssim \rho$. We can calculate the damping using the nonlinear estimate (3.3.44) and Lemma 3.2.1 that:

$$A_0 = O(\mathcal{E}_1) - \epsilon_2\frac{z\cdot\nabla\bar{U}}{\bar{U}}\,,\quad A_1 = -(\frac{1}{2}z + \mathcal{P}z) + O(\mathcal{E}_1),\quad |\mathcal{P}| \lesssim \mathcal{E}_1.$$

Next, we define a barrier function $F = \bar{U}^{-4\epsilon_2}$. Since $|z\cdot\nabla F| \lesssim F, |\nabla^i F| \lesssim F, i = 1, 2$, we get

$$\begin{aligned}\mathcal{P}_U F &= (O(\mathcal{E}_1) - \epsilon_2\frac{z\cdot\nabla\bar{U}}{\bar{U}} + \frac{A_1\cdot\nabla F}{F})F + \mathrm{tr}(Q\nabla^2 F)\\ &= (O(\mathcal{E}_1) - \epsilon_2\frac{z\cdot\nabla\bar{U}}{\bar{U}} - \frac{1}{2}\frac{z\cdot\nabla F}{F})F = (O(\mathcal{E}_1) - (\epsilon_2 - 2\epsilon_2)\frac{z\cdot\nabla\bar{U}}{\bar{U}})F = (O(\mathcal{E}_1) + \epsilon_2\frac{z\cdot\nabla\bar{U}}{\bar{U}})F.\end{aligned}$$

- For $|z| \ge 1$, we derive by the form of $\bar{U}$ in (3.1.27) the lower bound $-\frac{z\cdot\nabla\bar{U}}{\bar{U}} \ge \mu_{U,2}$ for some positive constant $\mu_{U,2}$. Recall the definition of $\mathcal{E}_1$ (3.3.11) and $E \lesssim 1$ from

(3.3.50) and Assumption (3.3.8). Since $|Q| \lesssim \mathrm{tr}(Q)$, for some positive constant $\mu_{U,1}$, we have

$$A_0 \geq \mu_{U,2}\epsilon_2 - \mu_{U,1}(\mathrm{tr}(Q) + H^{p-1} + E)\,, \quad \mathcal{P}_U F \leq (\mu_{U,1}(\mathrm{tr}(Q) + H^{p-1} + E) - \mu_{U,2}\epsilon_2)F. \tag{3.3.53}$$

- For $|z| \leq 1$, since ρ is bounded on the interval and we recall the definition of Γ (3.2.13), we can estimate

$$U\rho = \bar{U}^{-\epsilon_2} + W\rho \geq 4C_b - C\Gamma \geq 4C_b - \mu_{U,3}E\,, \tag{3.3.54}$$

for some positive constant $\mu_{U,3}$. Here we use the definition of C_b in Assumption 3.3.2.

Hence, by enforcing $E, |Q| + H^{p-1}$ sufficiently small, we will verify the following bootstrap assumption.

Assumption 3.3.9.

$$A_0 > 0,\ \mathcal{P}_U F < 0, \quad |z| \geq 1, \quad U\rho > 2C_b,\ |z| \leq 1. \tag{3.3.55}$$

Now, we consider $\Omega_c = U\rho + cF$ for $c > 0$. From Corollary 3.3.4 and the choice of ϵ_2, we obtain

$$U\rho \lesssim \langle z\rangle^{\sigma+\epsilon/2}\langle z\rangle^{-\sigma+\epsilon/2} = \langle z\rangle^{\epsilon}, \quad F = \bar{U}^{-4\epsilon_2} \gtrsim \langle z\rangle^{8\epsilon_2/(p-1)} \gtrsim \langle z\rangle^{2\epsilon}.$$

Under the assumption (3.3.55) and 3.3.2, for any $c > 0$, we have

$$\Omega_c(z) > 2C_b, \quad |z| \leq 1, \quad \lim_{|z|\to\infty} \Omega_c = \infty.$$

Using the above estimates of $\mathcal{P}_U$, we get

$$\partial_\tau \Omega_c = \partial_\tau(U\rho) = \mathcal{P}_U(U\rho + cF) - c\mathcal{P}_U F = \mathcal{P}_U\Omega_c - c\mathcal{P}_U F > \mathcal{P}_U\Omega_c.$$

By choosing initial data with $U\rho > 2C_b$ and then applying the maximal principle to the operator $\mathcal{P}_U$ on $|z| \geq 1$, we obtain

$$\Omega_c > 2C_b, \quad U\rho + cF \geq 2C_b, \quad |z| \geq 1.$$

Since c is arbitrary, taking $c \to 0$, we prove $U\rho > 2C_b$ for $|z| \geq 1$, which along with (3.3.55) for $U\rho$ concludes $U\rho \geq 2C_b, \forall z \in \mathbb{R}^d$ and strengthens (3.3.7) in Assumption 3.3.2.

In Section 3.3.6, we prove Assumption 3.3.9.

3.3.6 Bootstrap argument and blowup

In this section, we prove Theorem 3.3.1 by combining previous estimates and use a bootstrap argument.

Recall the ODE of Q from Lemma 3.2.1 and $\mathcal{E}_0, \Gamma$ from (3.2.13)

$$Q_\tau = -(Q_u + \frac{1}{2}Q_d)Q - Q(Q_u^T + \frac{1}{2}Q_d) + O(|Q|\mathcal{E}_0), \quad \mathcal{E}_0 = |Q|(\Gamma + \Gamma^4) + H^{p-1}. \tag{3.3.56}$$

Since the parameters ν_i in the energy E (3.3.50) have been chosen as some absolute constants, under the bootstrap assumption 3.3.2, we get

$$E \lesssim 1, \quad \Gamma \lesssim E_0 + E_K \lesssim E \lesssim 1, \quad \mathcal{E}_0 \lesssim |Q|\Gamma + H^{p-1} \lesssim |Q|E + H^{p-1}. \tag{3.3.57}$$

Taking trace on both side of (3.3.56) and then using $Q = Q_u + Q_u^T + Q_d$,

$$\mathrm{tr}((Q_u + \frac{1}{2}Q_d)Q + Q(Q_u + \frac{1}{2}Q_d)^T) = \mathrm{tr}((Q_u + \frac{1}{2}Q_d + \frac{1}{2}Q_d + Q_u^T)Q) = \mathrm{tr}(Q^2),$$
$$|Q| \approx \mathrm{tr}(Q), \quad \mathrm{tr}(Q^2) = \sum \lambda_{Q,i}^2 \geq \frac{1}{d}(\sum \lambda_{Q,i})^2 = \frac{1}{d}(\mathrm{tr}(Q))^2,$$

where $\lambda_{Q,i}$ is the eigenvalue of Q, and the above estimates, we get for a constant μ_2:

$$\partial_\tau \mathrm{tr}(Q) \leq -\mathrm{tr}(Q^2) + \mu_2(E\mathrm{tr}(Q)^2 + H^{p-1}\mathrm{tr}(Q)) \leq -\frac{1}{d}(\mathrm{tr}Q)^2 + \mu_2(E(\mathrm{tr}Q)^2 + H^{p-1}\mathrm{tr}(Q)). \tag{3.3.58}$$

Recall c_W from (3.2.4). To simplify the nonlinear estimates, in addition to bootstrap assumption 3.3.2, we impose the following assumption

Assumption 3.3.10.

$$|c_W| < \frac{1}{2}\min((p-1)^{-1}, 1), \quad E(\tau) < \min(\frac{1}{4d\mu_1}, \frac{\epsilon}{32\mu_2}), \tag{3.3.59}$$

where μ_1 is the constant in (3.3.50). We denote

$$\epsilon_1 = \mu_2 H^{p-1}, \quad a(\tau) = \exp(\mu_2 \int_0^\tau H^{p-1}(s)ds), \quad \lambda = \frac{\epsilon}{32}. \tag{3.3.60}$$

Consequence of bootstrap assumptions. We perform the energy estimates under the assumptions (3.3.59) and 3.3.2 and show that these estimates can be strengthened.

Using (3.2.2), (3.2.4), we obtain $(p-1)c_U(s) = -1 + (p-1)c_W(s) \leq -\frac{1}{2}$ and

$$H^{p-1}(\tau) \leq H^{p-1}(0)\exp(\int_0^\tau (p-1)c_U(s)ds) \leq H^{p-1}(0)\exp(-\tau/2), \quad -\frac{1}{d} + \mu_2 E(\tau) < -\frac{1}{2d}. \tag{3.3.61}$$

We can solve the ODE of $(\mathrm{tr}(Q))^{-1}$ using the above estimate and (3.3.58) to obtain

$$\partial_\tau E_Q^{-1} \geq \frac{1}{2d} - \mu_2 H^{p-1} E_Q^{-1}, \quad E_Q := \mathrm{tr}(Q).$$

By choose $H^{p-1}(0)$ small enough such that $\exp(2\epsilon_1) < 2$, for any $0 \leq s \leq \tau$, we get

$$a(\tau)a(s)^{-1} \leq e^{\epsilon_1 \int_0^\tau \exp(-s/2)ds} \leq e^{2\epsilon_1} < 2, \quad a(\tau)^{-1}a(s) > \frac{1}{2}, \quad a(0) = 1.$$

Solving the above ODE, we yield

$$\begin{aligned} &E_Q^{-1}(\tau) \geq a(\tau)^{-1}E_Q^{-1}(0) + \frac{1}{2d}\int_0^\tau a(\tau)^{-1}a(s)ds \geq \frac{1}{2}(E_Q^{-1}(0) + \frac{1}{2d}\tau), \\ &E_Q(\tau) \leq \min(2E_Q(0), 4d/\tau). \end{aligned} \tag{3.3.62}$$

Using (3.3.59), the above estimates, and $-\frac{\epsilon}{16} + \mu_1 E < -\frac{\epsilon}{32} = \lambda$ (3.3.60), we obtain

$$\frac{d}{d\tau}E \leq -\lambda E + C(E_Q + H^{p-1}(0)e^{-\tau/2}).$$

Solving the ODE and using (3.3.62), we obtain

$$E(\tau) \leq e^{-\lambda\tau}E(0) + C\int_0^\tau e^{-\lambda(\tau-s)}(\min(E_Q(0), \frac{1}{s}) + H^{p-1}(0)e^{-s/2})ds,$$

where C is some absolute constant and can depend on ϵ, λ. Since $\lambda < 1/2$, by decomposing the integral into $s < \tau/2$ and $s \geq \tau/2$, we obtain

$$\begin{aligned} E(\tau) &\leq e^{-\lambda\tau}(E(0) + CH^{p-1}(0)) + C\Big(E_Q(0)\int_0^{\tau/2} e^{-\lambda(\tau-s)}ds + \int_{\tau/2}^\tau e^{-\lambda(\tau-s)}\min(E_Q(0), 1/\tau)ds\Big) \\ &\leq e^{-\lambda\tau/2}(E(0) + \mu_3 H^{p-1}(0) + \mu_3 E_Q(0)) + \mu_3 \min(E_Q(0), 1/\tau) \end{aligned} \tag{3.3.63}$$

for some absolute constant $\mu_3 > 0$.

Plugging the above estimates and (3.3.57) into Lemma 3.2.1, and using $E \lesssim 1$ (3.3.59), we get for some $\mu_4 > 0$:

$$|c_W(\tau)| < C(E_Q(\tau) + E_Q(\tau)E(\tau) + H^{p-1}(\tau)) < \mu_4(\min(E_Q(0), 1/\tau) + H^{p-1}(0)e^{-\tau/2}). \tag{3.3.64}$$

Continuation of the bootstrap assumptions. For initial data satisfying

$$E(0) < E_*, \quad E_Q(0) < E_*, \quad H^{p-1}(0) < E_*, \tag{3.3.65}$$

with E_* sufficiently small, we obtain from (3.3.62), (3.3.63), (3.3.64) the following estimates

$$E(\tau) \le e^{-\lambda\tau/2}E_*(1+2\mu_3) + \mu_3 \min(E_*, 1/\tau) < E_*(1+3\mu_3), \quad E_Q(\tau) < 2E_*, \quad |c_W| < \mu_4 E_*,$$
$$H^{p-1}(\tau) \le H^{p-1}(0) < E_*, \quad E(\tau) + \mathrm{tr}(Q) + H^{p-1} < (4+3\mu_3)E_*.$$

Therefore, there exists $\nu_3 > 0$ such that for $E_* < \nu_3$, the bootstrap assumption (3.3.59) can be strengthened and continued. Plugging the above estimates into (3.3.53), (3.3.54) we obtain

$$A_0 \ge \mu_{U,2}\epsilon_2 - \mu_{U,3}(3\mu_3+4)E_*, \quad \mathcal{P}_U F \le (\mu_{U,2}(4+3\mu_3)E_* - \mu_{U,2}\epsilon_2)F, \quad U\rho \ge 4C_b - \mu_{U,3}(1+3\mu_3)E_*.$$

By further requiring E_* to be sufficiently small, the Assumption 3.3.9 can be strengthened and continued. The L^∞ estimate in Section 3.3.5 strengthens (3.3.7) in the Assumption 3.3.2. Using the definition (3.3.50) and the above estimate for E, we obtain

$$(E_0 + E_K + F_1 + F_K)(t) \le C(\nu_1, \nu_2)E_*,$$

which strengthens the first inequality in (3.3.8) in Assumption 3.3.2 by further choosing E_* to be small enough.

For the second inequality in (3.3.8), applying the Jacobi's formula $\frac{d}{d\tau}\det(Q(\tau)) = \det(Q(\tau))\mathrm{tr}(Q^{-1}\frac{d}{d\tau}Q)$ to (3.2.15) and using $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, $Q = Q_u + Q_d + Q_u^T$, we obtain

$$\partial_\tau \det(Q) = \det(Q) \cdot \mathrm{tr}(-Q + O(\mathcal{E}_0)).$$

From the above estimates, $|Q|$ and $\mathcal{E}_0$ remain uniformly bounded for all $\tau > 0$. Since $\det(Q(0)) > 0$, we prove $\det(Q) \ge \det(Q(0))e^{-C\tau}$, which strengthens the second inequality in (3.3.8). This concludes the proof of Theorem 3.3.1.

3.4 Refined asymptotics

In this section, building on Theorem 3.3.1, we obtain sharp asymptotics stated in Theorem 3.1.1. In Section 3.4.1, we estimate the sharp blowup rates for the amplitude similarly as in [218]. In Section 3.4.2, we estimate the asymptotics related to the phase and prove L^∞ convergence. In Section 3.4.3, we combine Theorem 3.3.1, Propositions 3.4.1 and 3.4.2 to prove Theorem 3.1.1.

3.4.1 Asymptotics of the amplitude and blowup rate

We use O_{in} and C_{in} to track any constant depending on the norm of the initial data $\mathrm{tr}(Q(0)), \mathrm{tr}(Q^{-1}(0))$. We have the following results for the asymptotics.

Proposition 3.4.1. *Suppose that the initial data (U, Θ, Q, H) satisfy the assumption in Theorem 3.3.1. We have the following asymptotics for the modulation parameters*

$$\left|\frac{H(\tau)^{p-1}}{T-t(\tau)}-1\right| \lesssim C_{in}\langle\tau\rangle^{-1}, \quad \lim_{\tau\to\infty}\frac{\tau}{|\log(T-t(\tau))|}=1, \quad \lim_{t\to T}\frac{\mathbf{R}(t)}{\sqrt{(T-t)|\log(T-t)|}}=I_d. \tag{3.4.1}$$

We consider $\tau \geq 2$. Note that $E_* < 1$. We focus on the asymptotics as $\tau \to \infty$ and the decay rate in τ.

Refined estimate of Q. By inserting (3.3.63) and (3.3.61) into (3.3.58), we get

$$\partial_\tau E_Q \leq -\frac{1}{d}E_Q^2 + C((\frac{1}{\tau}+e^{-\lambda\tau/2})E_Q^2 + E_Q e^{-\tau/2}),$$

for some absolute constant $C > 0$. Since $E_Q > 0$, we arrive at the ODE

$$\partial_\tau E_Q^{-1} \geq \frac{1}{d} - C(\frac{1}{\tau}+e^{-\lambda\tau/2}) - CE_Q^{-1}e^{-\tau/2}.$$

By introducing the integrating factor $a(\tau) = \exp(-CE_*\int_1^\tau e^{-s/2}ds)$, and using the fast convergence $|a(\tau)/a(s)-1| \lesssim E_* e^{-s/2}, a(\tau) \geq e^{-CE_*}$ for $1 \leq s < \tau$, we can solve the above ODE and obtain

$$E_Q^{-1} \geq \frac{1}{d}\tau + O(\log\tau) + E_Q^{-1}(2)e^{-CE_*} \geq \frac{1}{d}\tau + O_{in}(\log\tau).$$

Since $\mathrm{tr}(Q) = \sum \lambda_{Q,i}$, we know that

$$\min(\lambda_{Q,i}) \leq \frac{1}{d}E_Q \leq \frac{1}{\tau} + O_{in}(\frac{\log\tau}{\tau^2}). \tag{3.4.2}$$

Next we estimate $\mathrm{tr}(Q^{-1})$. From (3.3.56), we have by (3.3.57) that

$$\partial_\tau \mathrm{tr}(Q^{-1}) = d - 2\mathrm{tr}(\mathcal{E}_Q Q^{-1}) \leq d + \mu_2(EE_Q + H^{p-1})\mathrm{tr}(Q^{-1}).$$

By the above estimates of E_Q, and the same estimates of E and H^{p-1} in (3.3.63) and (3.3.61), we have that for sufficiently large τ, there exists a μ_5 such that

$$\partial_\tau \mathrm{tr}(Q^{-1}) \leq d + \frac{C}{\tau^2}\mathrm{tr}(Q^{-1}).$$

We conclude that

$$\mathrm{tr}(Q^{-1}) \leq d\tau + O(\log\tau) + \mathrm{tr}(Q^{-1}(2)) \leq d\tau + O_{in}(\log\tau)$$

Using $\mathrm{tr}(Q^{-1}) = \sum_i \lambda_{Q,i}^{-1}$, we obtain

$$\max(\lambda_{Q,i}) \geq \frac{d}{\mathrm{tr}(Q^{-1})} \geq \frac{1}{\tau} + O_{in}(\frac{\log\tau}{\tau^2}). \tag{3.4.3}$$

Combining the above estimates, we obtain

$$\mathrm{tr}(Q^{-1})\mathrm{tr}(Q) \leq d^2 + O_{in}(\frac{\log\tau}{\tau}).$$

Using $\mathrm{tr}(Q^{\alpha}) = \sum \lambda_{Q,i}^{\alpha}, \alpha = 1, -1$, we derive

$$\mathrm{tr}(Q^{-1})\mathrm{tr}(Q) = \sum \lambda_{Q,i} \sum \lambda_{Q,i}^{-1} = d^2 + \sum_{i<j} (\sqrt{\frac{\lambda_{Q,i}}{\lambda_{Q,j}}} - \sqrt{\frac{\lambda_{Q,j}}{\lambda_{Q,i}}})^2.$$

It follows

$$\Big(\sqrt{\frac{\lambda_{Q,i}}{\lambda_{Q,j}}} - \sqrt{\frac{\lambda_{Q,j}}{\lambda_{Q,i}}}\Big)^2 = O_{in}\Big(\frac{\log\tau}{\tau}\Big), \quad \forall i < j, \quad \frac{\max(\lambda_{Q,i})}{\min(\lambda_{Q,i})} = 1 + O_{in}((\frac{\log\tau}{\tau})^{1/2}).$$

Combining the above estimate with (3.4.2) and (3.4.3), we have that each one of the eigenvalue satisfies

$$\lambda_{Q,i} = \frac{1}{\tau} + O_{in}(\tau^{-3/2}\sqrt{\log\tau}) = \frac{1}{\tau} + O_{in}(a_\tau), \quad a_\tau = \tau^{-3/2+\epsilon_3}, \quad \epsilon_3 = \frac{1}{10}.$$

Since Q is symmetric and $Q(\tau) = R(\tau)\Lambda R(\tau)^T$ for $\Lambda = \mathrm{diag}(\lambda_{Q,1}, .., \lambda_{Q,d})$ and some orthogonal matrix R, which satisfies $|R(\tau)| \leq C$ for C independent in τ, the above estimates further imply,

$$Q = R(\tau)\Big(\frac{1}{\tau}I_d + O_{in}(a_\tau)\Big)R(\tau)^T = \frac{1}{\tau}I_d + O_{in}(a_\tau). \tag{3.4.4}$$

Estimate of R and blowup rate. Recall from (3.2.4), (3.2.5)

$$\mathcal{M} = e^{-\tau/2}\mathbf{R}^{-1}, \quad Q = C_W^{p-1}\mathcal{M}\mathcal{M}^T = C_W^{p-1}e^{-\tau/2}\mathbf{R}^{-1}(e^{-\tau/2}\mathbf{R}^{-1})^T = M_Q M_Q^T, \quad M_Q := C_U^{(p-1)/2}\mathbf{R}^{-1}.$$

Note that $\mathbf{R}, \mathcal{M}, M_Q$ are upper triangular matrices. Due to $M_{Q,ii}(0) > 0$ and the non-degeneracy $0 < \det(Q) = \det(M_Q)^2 = \prod M_{Q,ii}^2$ for all τ from (3.3.8), by continuity, we have $M_{Q,ii}(\tau) > 0$, which are the eigenvalues of M_Q, M_Q^T. For each real eigenpair (λ, v) of M_Q^T with $||v||_{l^2}^2 = 1$, we obtain

$$\lambda^2 = \lambda^2||v||_{l^2}^2 = v^T M_Q M_Q^T v = v^T Q v = \tau^{-1}||v||_{l^2}^2 + O_{in}(a_\tau) = \tau^{-1} + O_{in}(a_\tau).$$

Since $M_{Q,ii} > 0$ is a eigenvalue of M_Q, we obtain

$$M_{Q,ii} = \tau^{-1/2}(1+\tau O(a_\tau))^{1/2} = \tau^{-1/2} + O(\tau^{1/2}a_\tau) = \tau^{-1/2} + O((\log\tau)^{1/2}/\tau).$$

Next, we estimate the strictly upper part of M_Q: M_Q^u. Taking the trace, we get

$$\sum_{i\neq j} M_{Q,ij}^2 = \mathrm{tr}(M_Q M_Q^T) - \sum_i M_{Q,ii}^2 = \mathrm{tr}(Q) - \sum_i M_{Q,ii}^2 = d/\tau - d/\tau + O(a_\tau) = O(a_\tau)$$

which implies $M_Q^u = O(a_\tau^{1/2})$. Comparing the strictly upper part $Q = (M_{Q,d} + M_Q^u)(M_{Q,d} + M_Q^u)^T$, we get

$$M_Q^u M_{Q,d} = Q^u - (M_{Q,u}M_{Q,u}^T)^u = O(a_\tau), \quad M_Q^u = O(a_\tau)M_{Q,d}^{-1} = O(a_\tau\tau^{1/2}).$$

Therefore, we conclude,

$$C_U^{(p-1)/2}\mathbf{R}^{-1} = M_Q = \frac{1}{\sqrt{\tau}}I_d + O_{in}(\tau^{1/2}a_\tau) = \frac{1}{\sqrt{\tau}}I_d + O_{in}(\tau^{-1+\epsilon_3}). \tag{3.4.5}$$

Using (3.2.2), we define the blowup time as $T = t(\infty)$. Using (3.2.2) for $t(\tau)$, Lemma 3.2.1 and (3.3.59) for c_W, and (3.3.64), (3.3.6), (3.4.4) for $\mathrm{tr}(Q), \mathcal{E}_0$, we obtain

$$t_\tau = H^{p-1}, \quad (p-1)c_W = \frac{\mu_5}{\tau} + O_{in}(a_\tau), \quad |(p-1)c_W| < \frac{1}{2}, \quad a_\tau = \tau^{-3/2+\epsilon_3}, \quad \mu_5 = \frac{2(1-\beta\delta)dc_p}{p-1}. \tag{3.4.6}$$

Using (3.2.2), (3.2.4), for $\tau \geq 2, s > 0$, we obtain

$$\frac{H^{p-1}(\tau+s)}{H^{p-1}(\tau)} = e^{-s}F(\tau,s), \quad F(\tau,s) := e^{(p-1)\int_\tau^{\tau+s} c_W(z)dz}.$$

Since $|(p-1)c_W(s)| < \min(\frac{1}{2}, C\tau^{-1})$, using $|e^x - 1| \lesssim |x|(e^x+1), \partial_s F(\tau,s) = (p-1)c_W(\tau+s)F(\tau,s), F(\tau,0) = 1$, for $0 \leq z \leq s$, we obtain

$$|F(\tau,s) - 1| \lesssim s/\tau(F(\tau,s) + F(\tau,0)) \lesssim e^{s/2}s/\tau,$$
$$|c_W(\tau+z) - \mu_5\tau^{-1}| \lesssim |(\tau+z)^{-1} - \tau^{-1}| + a_\tau \lesssim z\tau^{-2} + a_\tau,$$

which implies

$$|c_W(\tau+z)F(\tau,z) - \mu_5\tau^{-1}| \lesssim |c_W(\tau+z)(F(\tau,z)-1) + (c_W(\tau+z) - \mu_5\tau^{-1})| \lesssim a_\tau + e^{z/2}z\tau^{-2} + z\tau^{-2},$$
$$|F(\tau,s) - 1 - s\mu_5\tau^{-1}| \lesssim s\max_{0\leq z\leq s}|c_W(\tau+z)F(\tau,z) - \mu_5\tau^{-1}| \lesssim e^{s/2}s^2\tau^{-2} + s\tau^{-3/2+\epsilon_3}.$$

Therefore, integrating $\frac{H^{p-1}(\tau+s)}{H^{p-1}(\tau)}$ for s from 0 to ∞ and using the above estimates, we get

$$\begin{aligned}\frac{T-t(\tau)}{H^{p-1}(\tau)} &= \int_0^\infty \frac{H^{p-1}(\tau+s)}{H^{p-1}(\tau)}ds = \int_0^\infty e^{-s}F(\tau,s)ds = \int_0^\infty (1+\mu_5 s\tau^{-1})e^{-s}ds + O_{in}(\tau^{-3/2+\epsilon_1}) \\ &= 1+\mu_5\tau^{-1}+O_{in}(\tau^{-3/2+\epsilon_1}),\end{aligned} \tag{3.4.7}$$

where we use $\int_0^\infty se^{-s}ds = 1$. Since $(p-1)c_U = (p-1)(\bar{c}_U + c_W) = -1 + O(\tau^{-1})$, we further obtain

$$\log(T-t(\tau)) = (1+O(\tau)^{-1})\log(H^{p-1}) = (1+O(\tau)^{-1})(O_{in}(1)+\int_0^\tau (p-1)c_U) = -\tau+O_{in}(\log(\tau)). \tag{3.4.8}$$

Combining (3.4.5), (3.4.7), (3.4.8), we prove (3.4.1) and Proposition 3.4.1.

3.4.2 Asymptotics of phase and L^∞ convergence

In this section, our goal is to prove the following convergence.

Proposition 3.4.2. *Suppose that the initial data (U,Θ) satisfy the assumption in Theorem 3.3.1. We have*

$$|U(z,\tau)e^{\iota(\Theta-A(\tau))} - \bar{U}^{1+\iota\delta}| \lesssim \min(\langle z\rangle^{\sigma+\epsilon/2}\langle\tau\rangle^{-\epsilon}, \langle\tau\rangle^{\max(-1,2\sigma)}). \tag{3.4.9}$$

where $A(\tau)$ satisfies the following estimate for some constant C_{in} depending on the initial data

$$|A-\bar{A}| \le C_{in}, \quad A(\tau) := \frac{\delta}{p-1}\tau+\Phi(\tau,0), \quad \bar{A}(\tau) := -\frac{\delta\log(T-t(\tau))}{p-1} - \frac{d\beta(1+\delta^2)\log|\log(T-t(\tau))|}{2\mathfrak{b}_*}. \tag{3.4.10}$$

Proof. We will first estimate $\Phi(0)$ and then prove convergence.

Estimate of $\Phi(0)$. Firstly, we compute $A_\tau, \bar{A}_\tau$. Using (3.3.2), we perform similar computations to the proof of Lemma 3.2.1 to compute that

$$\Phi_\tau(0) = -v\cdot\nabla\Phi(0) + \mathscr{D}_\Theta(0) = O(\mathcal{E}_0) + (\beta+\delta)\frac{\kappa_2}{\kappa_0}\mathrm{tr}(Q)\,.$$

Applying (3.4.4) and the estimates (3.3.6), (3.3.57) for $\mathcal{E}_0$, we yield

$$A_\tau(\tau) = \frac{\delta}{p-1} - \frac{d(\beta+\delta)}{2\mathfrak{b}_*\tau} + O_{in}(\tau^{-3/2+\epsilon_1}), \quad \bar{A}_\tau(\tau) = \frac{\delta}{p-1}\frac{t_\tau}{T-t} - \frac{d\beta(1+\delta^2)t_\tau}{2\mathfrak{b}_*|\log(T-t)|(T-t)}\,.$$

Using $t_\tau = H^{p-1}$ (3.4.6), (3.4.7), and (3.4.8), we yield

$$\frac{t_\tau}{T-t(\tau)} = \frac{H^{p-1}}{T-t(\tau)} = 1-\frac{\mu_5}{\tau}+O_{in}(\tau^{-3/2+\epsilon_1}), \qquad \frac{t_\tau}{(T-t)\log(T-t)} = \frac{1}{\tau}+O_{in}(\log(\tau)\tau^{-2}).$$

Using the definition of μ_5 (3.4.6), c_p (3.1.27), we conclude

$$\begin{aligned} A_\tau - \bar{A}_\tau &= (-\frac{d(\beta+\delta)}{2\mathfrak{b}_*} + \frac{\mu_5\delta}{p-1} + \frac{d\beta(1+\delta^2)}{2\mathfrak{b}_*})\frac{1}{\tau} + O_{in}(\tau^{-\frac{3}{2}+\epsilon_3}) \\ &= \frac{\delta}{\tau}(\frac{2(1-\beta\delta)dc_p}{(p-1)^2} - \frac{d(1-\beta\delta)}{2\mathfrak{b}_*}) + O_{in}(\tau^{-\frac{3}{2}+\epsilon_3}). \end{aligned}$$

The first term vanishes due to (3.1.27) for c_p. Since the error term is integrable in $\tau \geq 2$, we conclude the asymptotics of the phase (3.4.10).

L^∞ convergence. Recall $\Theta = \bar{\Theta} + \Phi$ (3.2.4). Integrating (3.3.13b) with $l = 1$, we obtain

$$|\Phi(z) - \Phi(0)| \lesssim \int_0^1 |\nabla\Phi(tz)|dt \lesssim E\int_0^1 \langle tz\rangle^{-1/2}dt \lesssim E\langle z\rangle^{1/2}. \tag{3.4.11}$$

Using $U = \bar{U} + W$ (3.2.4), we decompose

$$J := Ue^{\iota(\Theta-\bar{\Theta}-\Phi(0))} - \bar{U} = Ue^{\iota(\Phi-\Phi(0))} - \bar{U} = We^{\iota(\Phi-\Phi(0))} + \bar{U}(e^{\iota(\Phi-\Phi(0))} - 1) = I + II.$$

Appyling (3.3.13a) to I, (3.4.11) and $|e^{\iota x} - 1| \lesssim \min(|x|, 1)$ to II, and $E \lesssim \langle\tau\rangle^{-1}$ (3.3.6), we prove

$$\begin{aligned} |J| &\lesssim \langle z\rangle^{\sigma+\epsilon/2}E + \langle z\rangle^{\sigma}\min(E\langle z\rangle^{1/2}, 1) \lesssim \min(\langle z\rangle^{\sigma+\epsilon/2}(E+E^\epsilon), E+E^{-2\sigma}) \\ &\lesssim \min(\langle z\rangle^{\sigma+\epsilon/2}\langle\tau\rangle^{-\epsilon}, \langle\tau\rangle^{\max(-1,2\sigma)}). \end{aligned}$$

Since $\Theta - \bar{\Theta} - \Phi(0) + \delta\log\bar{U} = \Theta - A(\tau)$ (3.1.27) and (3.4.10), we get $U^{\iota\delta}J = Ue^{\iota(\Theta-A(\tau))} - \bar{U}^{1+\iota\delta}$. Since $|U^{\iota\delta}| = 1$, the above estimate conclude the proof of (3.4.9). □

3.4.3 Proof of Theorem 3.1.1

In this section, we prove Theorem 3.1.1 with the open set $\mathcal{O}$ prescribed in Remark 3.1.2.

Verification of assumptions in Theorem 3.3.1. We first choose $\nu < 1$ in (3.1.15) for Theorem 3.1.1. Following the proof of Corollary 3.3.4, we obtain $U_0 \lesssim \langle z\rangle^{\sigma+\epsilon/2}$.

Using the definitions (3.3.4) of F_K and (3.1.16) of the norm $\bar{\mathfrak{F}}_K$, and (3.1.28), we get

$$\mathring{\rho}_K = \rho_K U^2 \leq C\langle z\rangle^{2\sigma+\epsilon}\rho_K \geq C(1+|z|^{-d-2K}), \quad F_K \leq C\|\Phi\|_{\bar{\mathfrak{F}}_K}.$$

From (3.1.6), (3.3.50), we obtain $E \leq CE_{in}$. By choosing $\nu = cE_*$ in (3.1.15) with $c > 0$ sufficiently small, the assumptions (3.1.15) implies the assumptions (3.3.5) in Theorem 3.3.1 except for $E_Q < E_*$. Using the definitions of $\mathcal{M}_0, H(0)$ (3.1.13) $\mathcal{R}_0, Q$ (3.2.5), and $C_W(0) = H(0)$ (3.2.4), we obtain

$$\mathrm{tr}(Q(0)) = H(0)^{p-1}\mathrm{tr}(\mathcal{M}\mathcal{M}^T) = H(0)^{p-1}\mathrm{tr}(\mathcal{M}^T\mathcal{M}) = Cu_0(V_0)^{-p}\mathrm{tr}(\nabla^2 u_0(V_0))$$

for some absolute constant C. Therefore, by further choosing c small in $\mu = cE_*$, we obtain $|E_Q| < E_*$ from the last assumption in (3.1.15). We verify the assumptions in Theorem 3.3.1 and can use the results in Theorem 3.3.1, and Propositions 3.4.1 and 3.4.2.

For the time t in Theorem 3.1.1, we use the change of variables $t = t(\tau)$ (3.2.2). Then we only need to prove Theorem 3.1.1 in terms of the self-similar time τ.

Proof of estimates (3.1.9)**,** (3.1.12)**.** Using Theorem 3.3.1, we obtain the estimates (3.3.6), which along with the relation between (U,Θ), (u,θ) and $W = U - \bar{U}, \Phi = \Theta - \bar{\Theta}$ prove (3.1.9) in Theorem 3.1.1.

To obtain (3.1.12) and (3.1.11), we choose $\mu(t(\tau)) = A(\tau), \hat{\mu}(t) = A(\tau) - \bar{A}(\tau)$. Using $\Theta - A = \Theta - \bar{A} - \hat{\mu}$ and the formula of $\bar{A}$ (3.4.10), we obtain

$$J := |\log(T-t)|^{\iota\frac{d\beta(1+\delta^2)}{2b_*}}(T-t)^{\frac{1+\iota\delta}{p-1}}\psi(\mathbf{R}(t)z+\mathcal{V}(t),t)e^{-\iota\hat{\mu}(t)} = (T-t)^{\frac{1}{p-1}}H^{-1}U(z,\tau)e^{\iota(\Theta(z,\tau)-\bar{A}-\hat{\mu})}$$
$$= (T-t)^{\frac{1}{p-1}}H^{-1}U(z,\tau)e^{\iota(\Theta(z,\tau)-A(\tau))}.$$

We denote

$$J_2 = U(z,\tau)e^{\iota(\Theta(z,\tau)-A(\tau))}.$$

Using the limits $H(\tau)/(T-t)^{1/(p-1)} \to 1$ and $\tau/|\log(T-t(\tau))| \to 1$ as $\tau \to \infty$ (3.4.1), and the estimate (3.4.9), we prove

$$|J - \bar{U}^{1+\iota\delta}| \lesssim |(T-t)^{\frac{1}{p-1}}H^{-1} - 1|\cdot|\bar{U}^{1+\iota\delta}| + (T-t)^{\frac{1}{p-1}}H^{-1}|J_2 - \bar{U}^{1+\iota\delta}|$$
$$\lesssim C_{in}(1+\tau)^{\eta} \lesssim C_{in}(1+|\log(T-t(\tau))|)^{\eta},$$

where $\eta = \max(-1, 2\sigma)$.

Proof of rates (3.1.10), (3.1.11). Next, we show that $V(\tau)$ converges as $\tau \to \infty$ in (3.1.10). From Lemma 3.2.1 for $\mathcal{V}$, the decay estimates for Q, E in Theorem 3.3.1, and $\mathbf{R}(\tau) \to 0$ as $\tau \to \infty$ in Proposition 3.4.1, we obtain

$$|\mathbf{R}(\tau)\mathcal{V}(\tau)| \lesssim (1+\tau)^{-2}.$$

Since the upper bound is integrable in τ, using $|V_\tau(\tau)| = |\mathbf{R}(\tau)\mathcal{V}(\tau)|$ (3.2.5), we prove that $V(\tau)$ converges as $\tau \to \infty$. The asymptotics (3.1.11) follows from the definition of μ and (3.4.1).This ends the proof of Theorem 3.1.1.

3.4.4 Proof of Theorem 3.1.5

To prove Theorem 3.1.5, we only need to show that for $\epsilon_0 = \epsilon_0(u_0)$, assumption (3.1.17) implies that $\tilde{u}_0$ is in the open set O in Theorem 3.1.1.

Estimates of $\tilde{V}_0, \tilde{\mathcal{M}}_0, \tilde{H}_0$. Since V_0 is the unique maximizer and $\nabla^2 u_0(V_0) \succ 0$, for δ_1 sufficiently small, we obtain that $\tilde{u}_0$ admits a global non-degenerate maximizer $\tilde{V}_0$ close to V_0 [7] with $|V_0 - \tilde{V}_0| \to 0$ as $||\tilde{u}_0 - u_0||_{L^\infty} \to 0$. Using embedding (3.3.13a) in Proposition 3.3.3, we obtain

$$||u_0 - \tilde{u}_0||_{C^2} \lesssim ||u_0 - \tilde{u}_0||_{\mathcal{H}^K}.$$

Denote

$$\delta_1 := ||u_0 - \tilde{u}_0||_{\mathcal{H}^K}, \quad \delta_2 := |V_0 - \tilde{V}_0|.$$

Using continuity and the above embedding, we obtain

$$\lim_{\epsilon_0 \to 0} |\tilde{V}_0 - V_0| = \lim_{\epsilon_0 \to 0} \delta_2 = 0, \quad |\tilde{u}_0(\tilde{V}_0) - u_0(V_0)| \lesssim \delta_1 + \delta_2, \quad |\nabla^2 \tilde{u}_0(\tilde{V}_0) - \nabla^2 u_0(V_0)| \lesssim \delta_1 + \delta_2. \tag{3.4.12a}$$

Upon choosing $\epsilon_0 > 0$ small, we can define the initial modulation parameters (3.1.13) $\tilde{H}_0, \tilde{\mathcal{M}}_0$ associated with $\tilde{u}_0$ and obtain

$$\lim_{\epsilon_0 \to 0} |\tilde{\mathcal{M}}_0 - \mathcal{M}_0| + |H_0 - \tilde{H}_0| = 0. \tag{3.4.12b}$$

We denote by $\tilde{U}_0$ the rescaled variables for $\tilde{u}_0$:

$$\tilde{U}_0 = \tilde{H}_0 \tilde{u}_0(\tilde{\mathcal{M}}_0^{-1} z + \tilde{V}_0). \tag{3.4.13}$$

[7] Since for any $\delta > 0$, there exists a $r > 0$, such that $u_0(V_0) > u_0(V_0 + z) + \delta, |z| > r$, we obtain $|\tilde{V}_0 - V_0| < r$ when $\delta_1 < \delta/2$.

Verification of assumptions. We show that $(\tilde{u}_0, \tilde{U}_0)$ satisfies assumptions (3.1.15). The implicit constants can depend on u_0.

Firstly, assumptions (3.1.8), (3.1.15) for $\tilde{u}_0$ except for

$$\tilde{U}_0 \bar{U}^{-1-\epsilon_2} > 2C_b \tag{3.4.14}$$

follow from continuity and choosing δ_1 small. Condition (3.4.14) follows from the assumption (3.4.14) for U_0, (3.1.17), the triangle inequality, and choosing ϵ_0 small.

Next, we verify (3.1.16) for $\tilde{U}_0$, i.e.

$$||\tilde{U}_0 - \bar{U}_0||_{\mathfrak{E}_K} < \nu. \tag{3.4.15}$$

Using the definition (3.4.13), we decompose $\tilde{U}_0 - \bar{U}$ as follows

$$\tilde{U}_0 - \bar{U} = \tilde{H}_0\Big(\tilde{u}_0(\tilde{\mathcal{M}}_0^{-1}z + \tilde{V}_0) - u_0(\tilde{\mathcal{M}}_0^{-1}z + \tilde{V}_0)\Big) + \Big(\tilde{H}_0 u_0(\tilde{\mathcal{M}}_0^{-1}z + \tilde{V}_0) - \bar{U}_0\Big) := J_1 + J_2. \tag{3.4.16}$$

For J_1, using a change of variable, (3.4.12a), the assumption (3.1.17) for $u_0 - \tilde{u}_0$, and the embedding (3.3.13a), we obtain

$$||J_1||_{\mathcal{H}^K} \lesssim ||u_0 - \tilde{u}_0||_{\mathcal{H}^K} \lesssim \epsilon_0, \quad \lim_{\epsilon_0 \to 0} \max_{|z| \le 1} |\nabla^i J_1| = 0, \quad \text{for } i = 0, 1, 2, 3. \tag{3.4.17}$$

Denote $H_1 = \tilde{H}_0 H_0^{-1}, \mathcal{M}_1 = \mathcal{M}_0 \mathcal{M}_0^{-1}, V_1 = \mathcal{M}_0(\tilde{V}_0 - V_0)$. For J_2, using $u_0(z) = H_0^{-1} U_0(\mathcal{M}_0(z - V_0))$ (3.1.13) and a change of variable, we obtain

$$\begin{aligned} J_2 &= \tilde{H}_0 H_0^{-1} U_0(\mathcal{M}_0 \mathcal{M}_0^{-1} z + \mathcal{M}_0(\tilde{V}_0 - V_0)) - \bar{U}_0 = H_1 U_0(\mathcal{M}_1 z + V_1) - \bar{U} \\ &= H_1(U_0 - \bar{U})(\mathcal{M}_1 z + V_1) + \Big(H_1 \bar{U}(\mathcal{M}_1 z + V_1) - \bar{U}\Big) := J_{21} + J_{22}. \end{aligned}$$

From (3.4.12b), we obtain that $\mathcal{M}_1 - 1 = o(1), H_1 - 1 = o(1), V_1 = o(1)$. Since $||U_0 - \bar{U}_0||_{\mathfrak{E}_K} < \nu$, by choosing ϵ_0 small enough, we yield

$$||J_{21}||_{\mathfrak{E}_K} < \nu. \tag{3.4.18}$$

Using the smoothness of $\bar{U}$ and the embedding (3.3.13a), we obtain

$$\lim_{\epsilon_0 \to 0} ||J_{22}||_{\mathcal{H}^K} + \max_{|z| \le 1, i \le 3} |\nabla^i J_{22}| = 0. \tag{3.4.19}$$

Next, we show that for f with $\nabla^j f(0) = 0$, $j \le 2$ and $i \le K$, we have

$$||\nabla^i f||_{\rho_i} \lesssim ||f||_{\mathcal{H}^K}. \tag{3.4.20}$$

Using the definition of ρ_K (3.1.28), $\mathfrak{E}_K$ (3.1.6), $g_k, \mathcal{H}^K$ (3.3.9), and embedding (3.3.13a), for f with $\nabla^j f(0) = 0, j \leq 2$, we have

$$||\nabla^i f||_{\rho_i} \lesssim ||\nabla^3 f||_{L^\infty} || |x|^{3-i} \mathbf{1}_{|x|\leq 1}||_{\rho_i} + ||\nabla^i f||_{g_i} \lesssim ||f||_{\mathcal{H}^K}, \quad i \leq 3,$$
$$||\nabla^i f||_{\rho_i} \lesssim ||\nabla^i f||_{L^\infty} ||\mathbf{1}_{|x|\leq 1}||_{\rho_i} + ||\nabla^i f||_{g_i} \lesssim ||f||_{\mathcal{H}^K}, \quad i \leq (d+5)/2.$$

For $\frac{d+5}{2} < i \leq K$, ρ_i and g_i are equivalent and (3.4.20) follows from (3.3.12c). From the definition of $\tilde{U}_0$ and J_i, for $l \leq 2, i = 1, 2$, we obtain $\nabla^l J_{21}(0) = 0, \nabla^l (J_1 + J_{22})(0) = 0$. Applying (3.4.20), we obtain

$$||J_1 + J_{22}||_{\mathfrak{E}_K} \lesssim ||J_1 + J_{22}||_{\mathcal{H}^K},$$

which goes to 0 as $\epsilon_0 \to 0$. Since the inequality (3.4.18) is strict, for ϵ_0 small enough, we prove (3.4.15). Condition (3.1.16) follows from a similar argument, we conclude the proof of Theorem 3.1.5.

Chapter 4

BLOWUPS VIA LOCAL MODULATIONS: KELLER-SEGEL

The Keller-Segel equation, a classical chemotaxis model, and many of its variants have been extensively studied for decades. In this work, we focus on 3D Keller-Segel equation with a quadratic logistic damping term $-\mu\rho^2$ (modeling density-dependent mortality rate) and show the existence of finite-time blowup solutions with nonnegative density and finite mass for any $\mu \in \left[0, \frac{1}{3}\right)$. This range of μ is sharp; for $\mu \geq \frac{1}{3}$, the logistic damping effect suppresses the blowup as shown in [240, 431]. A key ingredient is to construct a self-similar blowup solution to a related aggregation equation as an approximate solution, with subcritical scaling relative to the original model. Based on this construction, we employ a robust weighted L^2 method to prove the stability of this approximate solution, where modulation ODEs are introduced to enforce local vanishing conditions for the perturbation lying in a singular-weighted L^2 space. As a byproduct, we exhibit a new family of type I blowup mechanisms for the classical 3D Keller-Segel equation.

4.1 Introduction

Chemotaxis is a widespread natural phenomenon, and it occurs when organisms, such as body cells or bacteria, detect and move toward chemical signals in their surroundings. A principal mathematical description of chemotaxis is provided by the Keller-Segel system:

$$\begin{cases} \partial_t \rho = \Delta \rho - \nabla \cdot (\rho \nabla c), \\ \Delta c + \rho = 0, \end{cases} \tag{KS}$$

where ρ represents the density of the bacteria and c denotes the concentration of the self-emitted chemical substance. The model captures two key processes: the diffusive effect of random bacterial motion and the directed movement of bacteria toward the highest concentration of the chemical. For a broader introduction to chemotaxis, see [62, 208, 209].

Biologically, beyond the competition between diffusion and bacterial aggregation, the limitation of resources and overcrowding necessitate introducing the bacterial mortality rate that further suppresses the aggregation process. In particular, we consider the following 3D coupled parabolic-elliptic Keller-Segel system with logistic

damping

$$\begin{cases} \partial_t \rho = \Delta \rho - \nabla \cdot (\rho \nabla c) - \mu \rho^2, \\ \Delta c + \rho = 0, \end{cases} \tag{KS-D}$$

where $-\mu\rho^2$ ($\mu \geq 0$) represents the logistic damping rate. For additional background on (KS-D) and related models, we refer the interested readers to [155, 203, 362, 410, 431].

4.1.1 Background

4.1.1.1 Classical Keller-Segel equation: global existence v.s. finite-time blowup

For the classical Keller-Segel equation (KS), the total mass $M(t) = \int \rho(t,x)dx$ is conserved for all time. Additionally, the system (KS) admits a one-parameter family of scaling invariance:

$$\rho(t,x) \to \frac{1}{\lambda^2}\rho\left(\frac{t}{\lambda^2}, \frac{x}{\lambda}\right), \ c(t,x) \to c\left(\frac{t}{\lambda^2}, \frac{x}{\lambda}\right), \quad \forall\, \lambda > 0. \tag{4.1.1}$$

We remark that this scaling invariance also holds for (KS-D).

In two dimensions, the total mass $M = 8\pi$ serves as a crucial quantity determining the global existence and finite-time blowup of the system (KS). Specifically, it has been shown that finite-time blowup occurs whenever $M > 8\pi$ with initial data $\rho_0 \in L^1_+((1+|x|^2), dx)$, while global-in-time existence with a uniform L^∞ bound holds when $M < 8\pi$, see [34, 123].

Nevertheless, the critical mass threshold does not exist for the three-dimensional model (KS). In fact, even with a tiny total mass, there is a radial solution that develops a finite-time singularity found by Nagai [337]. Beyond the radial case, Corrias, Perthame, and Zaag [102] demonstrated that blowup occurs whenever the second moment of the initial density is sufficiently small in comparison to the total mass, while weak solutions exist globally if the initial density has a suitably small $L^{\frac{3}{2}}$ norm. Interested readers can refer to [31, 32, 351, 421] for more results.

4.1.1.2 Classical Keller-Segel equation: singularity formation

Over the last several decades, the singularity formation for the classical Keller-Segel equation (KS) has been well studied. For the 2D case, since the model is in the L^1

critical sense, Naito and Suzuki [338] verified that any finite-time blowup solution to (KS) is of type II[1]. In particular, Raphaël and Schweyer [385] provided a precise construction of a radially stable finite-time blowup solution of the form

$$\rho(t,x) \approx \frac{1}{\lambda^2(t)} U\left(\frac{x}{\lambda(t)}\right), \quad U(x) = \frac{8}{(1+|x|^2)^2}, \tag{4.1.2}$$

where the blowup rate

$$\lambda(t) = \sqrt{T-t}\, e^{-\sqrt{\frac{|\ln(T-t)|}{2}}+O(1)}, \quad t \to T.$$

Subsequently, this result has been extended to several refined scenarios, including nonradial blowup, multi-bubble blowup without collision, and the simultaneous collision of two collapsing bubbles. For further details, we refer the interested reader to [52, 92, 96].

Unlike the 2D case, the 3D Keller-Segel system exhibits a large variety of blowup mechanisms. Notably, there is a countable family of radial self-similar blowup solutions of the form

$$\rho_s(t,x) = \frac{1}{T-t} Q_s\left(\frac{x}{\sqrt{T-t}}\right), \ [2] \tag{4.1.3}$$

which have been identified in [42, 195, 346]. In particular, there is an explicit self-similar solution given by

$$\rho_{s*}(t,x) = \frac{1}{T-t} Q_{s*}\left(\frac{x}{\sqrt{T-t}}\right), \quad \text{with} \quad Q_{s*}(x) = \frac{4(6+|x|^2)}{(2+|x|^2)^2}, \tag{4.1.4}$$

whose radial stability has been verified by Glogić and Schörkhuber [173]. Recently, Li and the third author [274, 275] extended this stability theory to the nonradial setting. Beyond self-similar blowup solutions, other non-self-similar formations have also been identified. For example, Collot, Ghoul, Masmoudi, and Nguyen [95] discovered a collapsing-ring blowup solution, and Nguyen, Nouaili, and Zaag [347] found a type I blowup solution with a log correction on the shrinking rate. Additionally, Hou, Nguyen, and Song [217] recently found a type II blowup solution under axisymmetry, whose local leading-order profile coincides with the rescaled stationary solution of the two-dimensional Keller-Segel equation as given in (4.1.2).

[1] The solution of (KS) exhibits type I blowup at $t = T$ if

$$\limsup_{t\to T} (T-t)\|\rho(t)\|_{L^\infty} < \infty.$$

Otherwise, the blowup is of type II.

[2] It exactly matches the scaling invariance (4.1.1) of the classical Keller-Segel system (KS).

4.1.1.3 Blowup or no blowup with logistic damping

To detect the core mechanism behind the bacterial aggregation, we can rewrite the aggregation term in (KS) or (KS-D) as

$$\nabla \cdot (\rho \nabla \Delta^{-1} \rho) = \nabla \Delta^{-1} \rho \cdot \nabla \rho + \rho^2.$$

In this expression, $\nabla \Delta^{-1} \rho \cdot \nabla \rho$ is the advection term, which, along with the effect of the diffusion term, cannot lead to the finite-time blowup. Consequently, the main factor causing the blowup is ρ^2.

If the logistic damping term $-\mu\rho^2$ in (KS-D) is replaced by a stronger term of the form $-\mu\rho^p$ with $\mu > 0$ and $p > 2$, then the aggregation effect is always suppressed. Precisely, the smooth solution always exists globally in time, regardless of how concentrated the initial density is, see [431].

In the scenario of quadratic damping $-\mu\rho^2$ (corresponding to (KS-D)), Tello and Winkler [431] proved that the global existence is guaranteed for any $\mu > 0$ in two dimensions. For higher dimension $d \geq 3$, they further demonstrated that global existence holds provided $\mu > \frac{d-2}{d}$. Subsequently, Kang and Stevens [240] verified the global existence in the critical case $\mu = \frac{d-2}{d}$ with $d \geq 3$. In the subcritical regime with $\mu \in \left(0, \frac{d-2}{d}\right)$, Fuest [155] partially bridged the gap by proving that finite-time blowup occurs in a bounded domain for $\mu \in \left(0, \frac{d-4}{d}\right)$ in higher dimensions $d \geq 5$. However, to the best of the authors' knowledge, prior to the present work, it remained an *open problem* whether blowup occurs for $\mu \in (\frac{d-4}{d}, \frac{d-2}{d})$ with $d \geq 5$ and $\mu \in (0, \frac{d-2}{d})$ with $3 \leq d \leq 4$.

Regarding the weaker damping term $-\mu\rho^p$ with $1 < p < 2$, for higher dimensions $d \geq 5$, Winkler [452] verified the existence of finite-time blowup solution when $1 < p < \frac{3}{2} + \frac{1}{2d-2}$. Subsequently, for dimensions $d \in \{3, 4\}$, Winkler [453] demonstrated finite-time blowup for $1 < p < \frac{7}{6}$. This range was later significantly extended by Fuest [155], where the author established the finite-time blowup for $1 < p < \frac{3}{2}$ in dimensions $d = 3$ and for $1 < p < 2$ in dimensions $d \geq 4$.

4.1.2 Main result

The principal objective of this work is to establish the *optimal blowup result* to (KS-D) in three dimensions. Specifically, we establish the existence of a smooth

finite-time blowup solution to (KS-D) with nonnegative density and finite mass for any $0 \le \mu < \frac{1}{3}$ in three dimensions.

In the literature, most blowup results for chemotaxis equations are obtained either by tracking the evolution of some appropriate functional (such as the second moment) to obtain a contradiction [34, 102, 123], or by working in the radial setting where the mass accumulation function satisfies an ODE [155]. However, the damping term $-\mu\rho^2$ in (KS-D) destroys the structures that these proofs rely on. In particular, it *leads to a time-decreasing mass* and *breaks the divergence form structure* that the classical Keller-Segel equation (KS) enjoys, thereby making these classical methods too limited to obtain the sharp blowup result for (KS-D).

Motivated by the various singularity formation results, ranging from the nonlinear heat equation [45, 94, 218, 324], nonlinear wave equation [124, 248], nonliear Schrödinger equation [325, 323], incompressible fluids [73, 74, 76, 219, 136, 135], compressible fluids [47, 78, 320], we adopt a strategy based on the direct construction of finite-time blowup solutions via the stability analysis of an approximate blowup solution. And the precise statement of the main theorem is as follows:

Theorem 4.1.1 (Existence of finite-time blowup to (KS-D) with finite mass)**.** *For any fixed $0 \le \mu < \frac{1}{3}$, let j_0 be an integer such that*

$$j_0 \ge J := \frac{3(1-\mu)}{1-3\mu} + 1 > 1, \tag{4.1.5}$$

it then follows that

$$\beta = \beta(j_0, \mu) := \frac{1}{3(1-\mu)} + \frac{1}{2j_0} < \frac{1}{2}. \tag{4.1.6}$$

In addition, under these choices, there exists a radially nonnegative $\rho_0 \in C_0^\infty(\mathbb{R}^3)$ [3] *such that the related smooth solution to (KS-D) exhibits finite-time blowup at $t = T < \infty$. Moreover, the solution satisfies*

$$\rho(t,x) = \frac{1}{T-t}\left[Q\left(\frac{x}{(T-t)^\beta}\right) + \varepsilon\left(t, \frac{x}{(T-t)^\beta}\right)\right], \quad x \in \mathbb{R}^3,\ t \in [0,T), \tag{4.1.7}$$

where

[3] Since $\rho_0 \in C_0^\infty(\mathbb{R}^3)$, the total mass of the system is necessarily finite. Additionally, with the nonnegativity of initial data ρ_0, an argument analogous to [274, Theorem A.1] ensures that the corresponding solution remains nonnegative in its lifespan.

• *(Existence of profile). Q is a unique smooth radial decreasing solution to the equation:*

$$Q + \beta y \cdot \nabla_y Q = \nabla_y Q \cdot \nabla_y \Delta_y^{-1} Q + (1-\mu) Q^2, \tag{4.1.8}$$

with the initial data

$$Q(0) = \frac{1}{1-\mu} \quad \text{and} \quad \partial_r^{(2j_0)} Q(0) = -(2j_0)! < 0. \tag{4.1.9}$$

• *(Smallness of error term). There exists $s = s(\mu) \in \mathbb{Z}_{\geq 1}$ sufficiently large and constants $C > 0$ and $\bar{\epsilon} > 0$ such that the error term ε satisfies*

$$\|\varepsilon(t)\|_{H^s(\mathbb{R}^3)} \leq C\,(T-t)^{\bar{\epsilon}}, \quad \forall\, 0 \leq t < T.$$

Comments on Theorem 4.1.1.

1. *Optimality in blowup results for Keller-Segel equation with logistic damping.*

Theorem 4.1.1 shows the existence of a finite-time blowup solution to 3D Keller-Segel equation with logistic damping (KS-D) for any fixed $0 \leq \mu < \frac{1}{3}$. This result provides the optimal blowup threshold for (KS-D) and completes the long-standing conjecture proposed in [155, 431] as mentioned in Subsection 4.1.1.3.

Moreover, our method is robust enough to extend naturally to higher dimensions $d \geq 3$, resulting in an analogous sharp blowup regime. Specifically, for $\mu \in \left[0, \frac{d-2}{d}\right)$ with $d \geq 3$, there always exists a finite-time blowup solution to (KS-D).

When the damping term is weakened to $-\mu\rho^p$ with $\mu > 0$ and $1 \leq p < 2$, the problem becomes simpler, as this term can also be viewed primarily as a perturbation of the dominant aggregation effect. Concretely, one could use $\frac{1}{T-t} Q\left(\frac{x}{(T-t)^\beta}\right)$ with Q constructed in Theorem 4.1.1 with $\mu = 0$ as an approximate solution. Alternatively, one may use the self-similar solution (4.1.4) of the classical Keller-Segel equation (KS) as an approximate solution. Then, we could apply the stability theory developed in this work or [173, 275] to establish the existence of finite-time blowup for any $\mu > 0$ and $p \in [1, 2)$ in three dimensions. Since this problem is easier to handle, we do not pursue this in detail and only study (KS-D) with essential difficulty.

2. *New blowup mechanism for classical Keller-Segel equation in 3D.*

For 3D classical Keller-Segel equation (KS), by applying Theorem 4.1.1 with $\mu = 0$, we have constructed a countably infinite family of blowup solutions, whose

asymptotic behavior is given by

$$\rho(t,x) \approx \frac{1}{T-t} Q_{\beta_j}\left(\frac{x}{(T-t)^{\beta_j}}\right), \qquad \forall\, t \in (0,T),$$

where $\beta_j = \frac{1}{3} + \frac{1}{2j} < \frac{1}{2}$ for all $j \geq 4$ and each Q_{β_j} solves (4.1.8) with $\beta = \beta_j$. This blowup is an "abnormal" Type I blowup, in the sense that it does not match natural scaling (4.1.1) [4]. Consequently, as a by-product of Theorem 4.1.1, we establish a new family of blowup mechanisms for 3D classical Keller-Segel equation (KS).

3. *Extension of Theorem 4.1.1.*

Theorem 4.1.1 establishes the precise blowup mechanism for the density but does not address its stability. Nevertheless, by following the standard approach outlined in [94, 275], one can construct a finite-codimensional Lipschitz stable manifold of radial initial data such that the corresponding solutions to (KS-D) exhibit blowup dynamics similar to in (4.1.7).

Besides, a natural extension to the nonradial blowup can be expected by using the spectral method from [57, 58, 71, 78]. Alternatively, following the argument in [77], one might introduce a matrix system of modulation ODEs to realize it. Furthermore, noting that the blowup solution constructed in Theorem 4.1.1 is highly localized, one can expect a finite-time blowup solution in a bounded domain by adapting the cut-off technique from [57, 58]. However, as the main focus of this work is on the existence of finite-time blowup solutions to (KS-D), we do not pursue these generalizations in detail here.

In addition, our method is expected to be sufficiently robust to potentially shed light on other models, such as the nonlinear heat equations. As a further motivation for the linear and nonlinear analysis for the Keller-Segel equation (KS-D) in Section 4.4 and Section 4.5, we briefly outline this in Section 4.2 for 1D semilinear heat equation.

4.1.3 Proof strategy and related key highlights.

4.1.3.1 Key idea of the construction: diffusion term as a perturbation.

[4] We take the ansatz $\rho(t,x) = \frac{1}{(T-t)^a} Q\left(\frac{x}{(T-t)^b}\right)$ and plug it into (KS-D). Requiring that all terms in the resulting equation have the same strength, the exponents must satisfy $a = 2b = 1$. This choice precisely coincides with the scaling symmetry (4.1.1) of the equation. We refer to blowup solutions satisfying this form as exhibiting natural self-similar blowup, in analogy with (KS-D).

Motivated by the explicit self-similar blowup solution (4.1.4) to the classical Keller-Segel equation (KS), a direct idea is to seek an analogous self-similar solution matching the natural scaling for the damped system (KS-D). Specifically, since (KS-D) satisfies the scaling invariance (4.1.1), one is led to consider special solutions of form

$$\rho_\mu(t,x) = \frac{1}{T-t} Q_\mu\left(\frac{x}{\sqrt{T-t}}\right),$$

where Q_μ solves

$$Q_\mu + \frac{1}{2} y \cdot \nabla Q_\mu = \Delta Q_\mu + \nabla \cdot (Q_\mu \nabla \Delta^{-1} Q_\mu) - \mu Q_\mu^2. \tag{4.1.10}$$

However, the damping term $-\mu\rho^2$ significantly complicates the problem by breaking the divergence structure of (KS). In particular, the partial mass

$$m_\rho(r) := \int_{B(0,r)} \rho(x)dx, \tag{4.1.11}$$

is invalid to simplify the equation (4.1.10) into a second order local differential equation, resulting (4.1.10) essentially a third-order differential equation for general $\mu \in \left(0, \frac{1}{3}\right)$, which, to the best of authors' knowledge, is quite difficult to analyze and is still open.

Instead, we adopt an alternative approach inspired by [77, 218, 320, 319, 347, 352], where the first step is to neglect the diffusion term and construct a self-similar blowup solution for the following 3D aggregation equation

$$\partial_t \rho = \nabla\rho \cdot \nabla\Delta^{-1}\rho + (1-\mu)\rho^2, \tag{4.1.12}$$

in the form

$$\rho(t,x) = \frac{1}{T-t} Q\left(\frac{x}{(T-t)^\beta}\right), \quad \text{with } \beta < \frac{1}{2}, \tag{4.1.13}$$

where the profile Q satisfies (4.1.8). We then choose (4.1.13) as an approximate solution to the original system (KS-D), which breaks the natural scaling of (KS-D), thus making the diffusion term $\Delta\rho$ subcritical and allowing it to be treated as a perturbation relative to the dominant aggregation term.

Compared to (4.1.10), the equation (4.1.8) is effectively second-order and much more tractable. By restricting to radial solutions and introducing the averaged mass of Q in the ball $B(0,r)$

$$f_Q(r) := \frac{1}{r^3}\int_0^r Q(s)s^2 ds, \tag{4.1.14}$$

(4.1.8) can be reformulated as an ODE system in terms of (Q, f_Q) (see (4.3.3)). The formation of solutions to (4.3.3) is through the dynamical system and a standard phase portrait analysis (See Figure 4.3 and Lemma 4.3.2). These methods have been instrumental in recent developments concerning the construction of smooth profiles for implosion, gravitational collapse, and collapsing-expanding shocks [47, 183, 184, 234, 235, 320]. The regularity of self-similar solutions serves as a fundamental criterion for solution admissibility and plays a crucial role in ensuring the existence of a nontrivial solution in the context of the present problem.

4.1.3.2 Compatibility of $\mu < \frac{1}{3}$: insights from the phase portrait analysis.

One of the main goals of our chapter is to figure out the existence of the smooth nontrivial radial solution to (4.1.8) for some $\beta < \frac{1}{2}$, which can be further simplified into finding a smooth curve connecting

$$P_0 := \left(\frac{1}{1-\mu}, \frac{1}{3(1-\mu)}\right), \tag{4.1.15}$$

[5]

and the origin O under the ODE system for (Q, f_Q) given by (4.3.3). We classify the problem into three cases based on the relative position between $f_Q = \beta$ and P_0, and analyze the corresponding phase portraits:

Case I. *P_0 lies strictly above the line $\{f_Q = \beta\}$: $\beta < \frac{1}{3(1-\mu)}$ (See Figure 4.1).*
We remark here that in *Case I* $\left(\beta < \frac{1}{3(1-\mu)}\right)$, motivated by the related phase portrait (see Figure 4.1), the flow field around P_0 is directed toward P_0. Consequently, P_0 is a stagnant point in this case and $Q(r) \equiv \frac{1}{1-\mu}$ is the only smooth solution to (4.3.3) with $(Q(0), f_Q(0)) = P_0$ that can be obtained. Therefore, to obtain a nontrivial solution to (4.3.3) with $(Q(0), f_Q(0)) = P_0$, it is necessary for β to satisfy

$$\frac{1}{3(1-\mu)} \leq \beta < \frac{1}{2}. \tag{4.1.16}$$

This requires that $\mu < \frac{1}{3}$, which aligns exactly with the *optimal range of* μ provided in Subsection 4.1.1.3 and Theorem 4.1.1.

Case II. *P_0 lies on the line $\{f_Q = \beta\}$: $\beta = \frac{1}{3(1-\mu)}$ (See Figure 4.2).*

Case III. *P_0 lies strictly below the line $\{f_Q = \beta\}$: $\beta > \frac{1}{3(1-\mu)}$ (See Figure 4.3).*

[5] With the singular term $\frac{1}{r}$ appearing on the right-hand side of the ODE system (4.3.3), the choice of initial data $(Q(0), f_Q(0)) = P_0$ should be required to ensure that Q is smooth near $r = 0$.

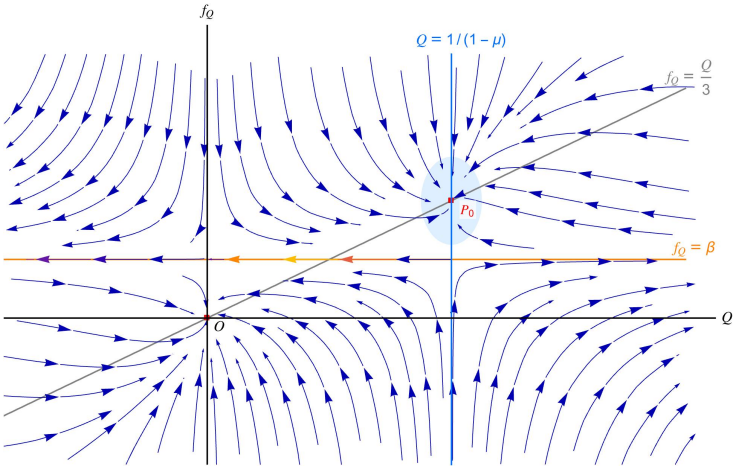


Figure 4.1: Qf_Q-plane with P_0 above the line $f_Q = \beta$. In this case, there is only a trivial solution solving (4.3.3) with $(Q(0), f_Q(0)) = P_0$.

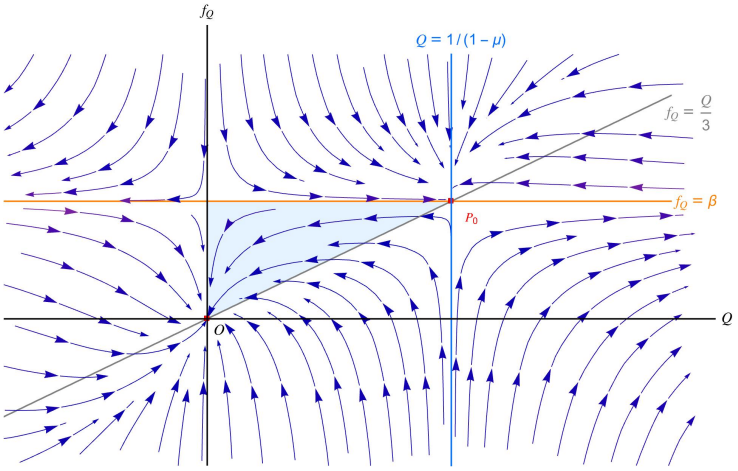


Figure 4.2: Qf_Q-plane with P_0 on the line $f_Q = \beta$. In this case, though it seems possible to have a solution curve connecting between P_0 and the origin O from the phase portrait, with the overwhelming singularity on the right-hand side of (4.3.3), after some careful analysis, the only possible smooth solution starting from P_0 to should be the trivial one: $(Q(r), f_Q(r)) \equiv P_0$.

4.1.3.3 Stability analysis beyond the spectral theory.

After selecting the self-similar blowup solution (4.1.13) of the aggregation equation (4.1.12) as an approximate solution to (KS-D), we then proceed to study the perturbation dynamics near the profile Q under the self-similar coordinate (4.4.2), which constitutes the key to our stability analysis.

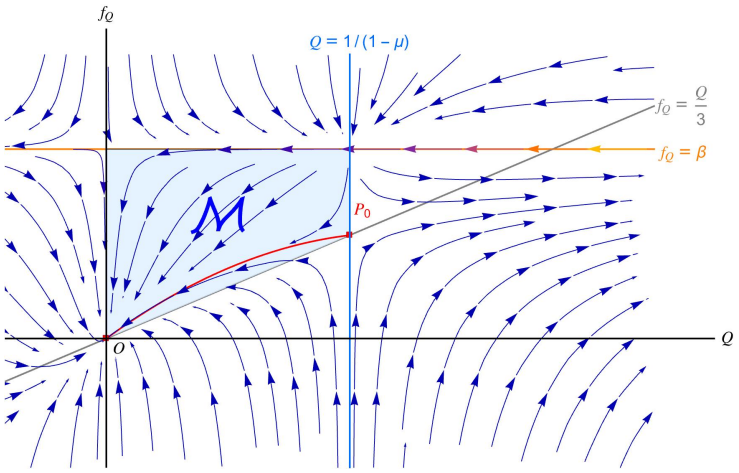


Figure 4.3: Qf_Q-plane with P_0 below $f_Q = \beta$. In this case, we can always choose countably many $\{\beta_j\}_j$ (see Theorem 4.1.1 and Lemma 4.3.2 for more details) such that there exists a smooth solution curve (the red curve in the figure) lying in $\mathcal{M}$ and connecting P_0 and the origin O simultaneously, we will discuss this case in more details later in Lemma 4.3.2.

In the literature, one of the natural approaches to address this problem is to analyze the spectral properties of the associated linearized operator. Such analysis typically relies on compactness arguments [47, 58, 142, 236, 274, 325, 320] or other more intricate operator theory, e.g. studying the precise or approximated spectrum of the linearized operator [95, 217, 275, 324, 347]. For problems with an explicit profile available, studying the spectrum of the linearized operator around the profile yields useful information for stability, e.g. [95, 275, 324]. However, for more complicated dynamics with only implicit or numerically approximate profiles, as is our case, obtaining the spectral information becomes significantly more challenging.

Alternatively, we adopt a robust approach based on energy estimates with a singular weight appropriately chosen. Linear stability can be extracted via enforcing local vanishing conditions of the perturbation at the origin, using only limited information of the profile. The idea was first demonstrated via the seminal works of [73, 76] in the self-similar case, and later on generalized by the second author with collaborators to type I singularities beyond self-similarity [77, 218] with the blowup law automatically inferred. Such a stability argument via singular weights will be further amenable to computer-assisted proofs. In this chapter, we further generalize the idea to the finite-codimensional stability case by studying this open problem.

4.1.4 Structure of the chapter.

In Section 4.2, we briefly discuss the approach for handling the finite-codimensional stability of blowup solutions to the 1D semilinear heat equation by introducing the singular weights. Motivated by this illustrative example, we then focus on the Keller-Segel equation with logistic damping term (KS-D). In Section 4.3, given any fixed $\mu \in \left[0, \frac{1}{3}\right)$, we establish the existence of nontrivial solutions to equation (4.1.8) with an appropriate choice of $\beta \in \left(\frac{1}{3(1-\mu)}, \frac{1}{2}\right)$, the quantitative properties of the approximate profiles will be investigated as well. For the stability analysis, we first examine the coercivity of the linearized operator in the L^2_w sense with w being sufficiently singular at the origin, as detailed in Section 4.4. Finally, in Section 4.5, we will study the nonlinear stability to complete the proof of Theorem 4.1.1.

4.1.5 Notations.

We denote

$$\langle r \rangle = \sqrt{1 + r^2}.$$

And we define a radial smooth cut-off function χ on $B(0,1)$ by

$$\chi(x) = \begin{cases} 1, & \text{if } |y| \leq 1, \\ 0, & \text{if } |y| \geq 2. \end{cases} \tag{4.1.17}$$

Based on this, we define the smooth cut-off function on $B(0,R)$ by

$$\chi_R(x) := \chi\left(\frac{x}{R}\right). \tag{4.1.18}$$

For any given weighted function $w : \mathbb{R}^3 \to \mathbb{R}$, we define $L^2_w(\mathbb{R}^3)$ the weighted inner product by

$$(g_1, g_2)_{L^2_w(\mathbb{R}^3)} = \int_{\mathbb{R}^3} g_1(x) g_2(x) w(x) dx,$$

and the weighted L^2 space $L^2_w(\mathbb{R}^3)$ by the collections of all functions of g satisfying $\|g\|^2_{L^2_w} := (g,g)_{L^2_w} < \infty$.

Furthermore, we define $C^\infty_0(\Omega)$ as the collection of all $C^\infty(\mathbb{R}^3)$ functions with compact support in $\Omega \subset \mathbb{R}^3$.

For any fixed $m \in \mathbb{Z}_{\geq 0}$, we define $\dot{H}^{2m}$ the collection of all functions satisfying

$$\|g\|^2_{\dot{H}^{2m}} = \int_{\mathbb{R}^3} |\Delta^m g|^2 dx < \infty,$$

and define $\dot{H}^{2m+1}$ the collection of all functions satisfying

$$\|g\|^2_{\dot{H}^{2m+1}} = \int_{\mathbb{R}^3} |\nabla \Delta^m g|^2 dx < \infty.$$

In addition, we denote $H^k(\mathbb{R}^3)$ for the collection of all functions with finite $\dot{H}^m$ norm for all $0 \le m \le k$. And H^k_{rad} is the collection of all radial functions lying in H^k. Moreover, wewrite $H^\infty(\mathbb{R}^3) := \cap_{k=0}^\infty H^k(\mathbb{R}^3)$.

If $h(x) = h(|x|)$ is a smooth radial function on $\mathbb{R}^3$, then $h(r)$ can be approximated by the Taylor expansion at the origin:

$$h(r) = \sum_{j=0}^{M} [h]_j r^{2j} + O(r^{2M+2}), \quad \text{with } [h]_j := \left.\frac{\partial_r^{(2j)} h}{(2j)!}\right|_{r=0}, \tag{4.1.19}$$

where $[h]_j$ is the $2j$-th order coefficient of Taylor expansion of $h(r)$ at the origin.

4.2 Motivating example of 1D semilinear heat equation

In this section, to illustrate the ideas of proof more clearly, we first consider the simpler 1D semilinear heat equation as a motivating example. Precisely, we will briefly sketch the high-level idea to study high-order vanishing type-I blowup for the 1D semilinear heat equation under *radial (even symmetric)* setting:

$$u_t = u_{xx} + u^2. \tag{HEAT}$$

4.2.1 Self-similar renormalization and the approximate solution

For any fixed $m \in \mathbb{Z}_{>1}$, we introduce the self-similar coordinate

$$y = \frac{x}{\lambda^{\frac{1}{m}}}, \quad \frac{d\tau}{dt} = \frac{1}{\lambda^2}, \quad \tau\Big|_{t=0} = 0, \quad \frac{\lambda_\tau}{\lambda} = -\frac{1}{2}, \tag{4.2.1}$$

and corresponding renormalization

$$u(t,x) = \frac{1}{\lambda^2} U(\tau, y), \tag{4.2.2}$$

then $\lambda(\tau) = \lambda_0 e^{-\frac{1}{2}\tau}$ and U solves the equation

$$\partial_\tau U = \lambda^{2-\frac{2}{m}} U_{yy} - U - \frac{1}{2m} y \cdot \nabla U + U^2, \tag{4.2.3}$$

where the diffusion term can be regarded as a perturbation since $2 - \frac{2}{m} > 0$. This, in turn, motivates us to find an approximate solution to the equation

$$-U - \frac{1}{2m} y \cdot \nabla U + U^2 = 0. \tag{4.2.4}$$

In particular, this equation can be explicitly solved by

$$U_*(y) = (1 + cy^{2m})^{-1}. \tag{4.2.5}$$

Here $c > 0$ is a constant, and $m > 1$ describes the vanishing order of the next order expansion of U_* near the origin.

4.2.2 Linear stability

We fix $m > 1$ and $c = 1$ in (4.2.5), and plug the ansatz $U = U_* + \varepsilon$ into (4.2.3), it then follows that ε solves

$$\partial_\tau \varepsilon = \lambda^{2-\frac{2}{m}} U_{yy} + \mathcal{L}\varepsilon + \varepsilon^2,$$

where the linearized operator reads

$$\mathcal{L}\varepsilon = -\varepsilon - \frac{1}{2m} y\partial_y \varepsilon + 2U_*\varepsilon. \tag{4.2.6}$$

Next, we introduce a weighted L^2_Θ space with singular weight $\Theta(y) = y^{-4m-4}$ near the origin to extract damping, which is the essential step to close the nonlinear stability. Via the integration by parts, we have the coercivity near the origin

$$(\mathcal{L}\varepsilon, \varepsilon)_{L^2_\Theta} = \left(\left(-1 + 2U_* + \frac{(\Theta y)_y}{4m\Theta} \right) \varepsilon, \varepsilon \right)_{L^2_\Theta} \approx -\frac{3}{4m} (\varepsilon, \varepsilon)_{L^2_\Theta}. \tag{4.2.7}$$

In particular, with careful analysis, there is a small constant $0 < \kappa = \kappa(m, U_*) \ll 1$ such that (4.2.7) can be extended to

$$(\mathcal{L}\varepsilon, \varepsilon)_{L^2_{\Theta+\kappa}} \le -\frac{1}{4m} (\varepsilon, \varepsilon)_{L^2_{\Theta+\kappa}}. \tag{4.2.8}$$

Additionally, we need to introduce the higher Sobolev norm $\dot{H}^{\bar{K}}$ to close the bootstrap argument. Precisely,

$$\begin{aligned} (\mathcal{L}\varepsilon, \varepsilon)_{\dot{H}^{\bar{K}}} &= \left(\left(-1 - \frac{\bar{K}}{2m} + 2U_* + \frac{1}{4m} \right) \partial_y^{\bar{K}} \varepsilon, \partial_y^{\bar{K}} \varepsilon \right)_{L^2} + O(\|\varepsilon\|_{H^{\bar{K}-1}} \|\varepsilon\|_{\dot{H}^{\bar{K}}}) \\ &\le -\frac{2\bar{K} - 4m - 1}{4m} \|\varepsilon\|^2_{\dot{H}^{\bar{K}}} + O(\|\varepsilon\|_{H^{\bar{K}-1}} \|\varepsilon\|_{\dot{H}^{\bar{K}}}), \end{aligned} \tag{4.2.9}$$

where the leading order enjoys damping once we choose $\bar{K} = \bar{K}(m) \gg 1$.

4.2.3 Modulation ODEs and nonlinear stability

With the singular weight $\Theta(y) = |y|^{-4m-4}$ given previously, we further *radially* decompose ε into

$$\varepsilon(\tau, y) = \varepsilon_u(\tau, y) + \varepsilon_s(\tau, y), \qquad \text{with} \qquad \varepsilon_u = \sum_{j=0}^{m} c_j(\tau)\chi(y)y^{2j}, \tag{4.2.10}$$

such that $\varepsilon_s(\tau, y) = O(y^{2m+2})$ near the origin, which yields an ODE system for modulation parameters $\{c_j\}_{j=0}^m$:

$$\begin{cases} \dot{c}_j = \left(1 - \frac{j}{m}\right) c_j + [\varepsilon_u^2]_j + \lambda^{2-\frac{2}{m}}[U_{yy}]_j, & 0 \le j < m-1, \\ \dot{c}_m = [\varepsilon_u^2]_m - 2c_0 + \lambda^{2-\frac{2}{m}}[U_{yy}]_m, & j = m. \end{cases} \tag{4.2.11}$$

Additionally, $\varepsilon_s = O(y^{2m+2})$ solves the equation

$$\partial_\tau \varepsilon_s = \mathcal{L}\varepsilon_s + 2\varepsilon_u\varepsilon_s + \varepsilon_s^2 + G[\lambda, U, \varepsilon_u],$$

with the modulation term $G[\lambda, U, \varepsilon_u] = O(y^{2m+2})$ given by

$$G[\lambda, U, \varepsilon_u] = \left(\lambda^{2-\frac{2}{m}}U_{yy} + \mathcal{L}\varepsilon_u + \varepsilon_u^2\right) - \sum_{j=0}^{K}\left[\lambda^{2-\frac{2}{m}}U_{yy} + \mathcal{L}\varepsilon_u + \varepsilon_u^2\right]_j \chi y^{2j}.$$

Finally, we can use the standard topological argument together with (4.2.7) and (4.2.9) to derive the nonlinear stability with finite codimension $m + 1$.

Remark 4.2.1. *We expect that this nonlinear stability result can be improved to finite-codimension $m-1$, which is two dimensions lower than our previous findings. The key underlying reason is the presence of two degrees of freedom, namely, the choice of the blowup time $T > 0$ and the shrinking rate. These degrees of freedom can be utilized through a matching argument to recover the corresponding unstable directions as in [273, 275].*

Alternatively, one may employ a method of dynamical rescaling to establish stability with finite codimension $m-1$. Specifically, we modify the coordinate (4.2.1) as

$$y = \frac{x}{\mu^{\frac{1}{m}}}, \quad \frac{d\tau}{dt} = \frac{1}{\lambda^2}, \quad \tau\big|_{t=0} = 0, \quad \frac{\lambda_\tau}{\lambda} = -\frac{1}{2} + \frac{1}{2}c_a, \quad \frac{\mu_\tau}{\mu} = -\frac{1}{2} - mc_s.$$

We then define the corresponding renormalization

$$u(t, x) = \frac{1}{\lambda^2}U(\tau, y).$$

Under this coordinate transformation, U satisfies

$$\partial_\tau U = \lambda^2 \mu^{-\frac{2}{m}} U_{yy} + (-1 + c_a)\, U - \left(\frac{1}{2m} + c_s\right) y \partial_y U + U^2,$$

where the parameters (c_a, c_s) are determined by the modulation conditions

$$U(\tau, 0) = U_*(0) \qquad \text{and} \qquad [U(\tau, 0)]_m = [U_*]_m,$$

and we eliminate neutral modes by fixing $c_0 = c_m = 0$. By applying a similar argument of modulation ODEs, we obtain the nonlinear stability with finite codimension $m - 1$. Notably, introducing extra scaling parameters to perturb the scaling symmetry is crucial for extending the argument to the nonradial setting; see previous works by the second author and collaborators [77, 218].

Remark 4.2.2. *Compared with the semilinear heat equation* (HEAT)*, analyzing the Keller-Segel equation with logistic damping* (KS-D) *involves several additional challenges. For example, the profile U_* introduced in* (4.2.5) *is an explicit solution to the first-order and separable local equation* (4.2.4)*. In contrast, for* (KS-D) *with $\mu \in \left(0, \frac{1}{3}\right)$, the associated profile equation* (4.1.8) *is inherently nonlocal and cannot be trivially solved. This nonlocality requires a more delicate analysis, which will be carried out in Section 4.3.*

Moreover, since there is no explicit nontrivial solution to (4.1.8) *with* (4.1.9)*, additional effort is required to derive quantitative properties of the profile Q. Combined with the nonlocal nature of* (KS-D)*, these complexities make the establishment of linear coercivity of the Keller-Segel equation more intricate than in the case of the semilinear heat equation* (HEAT)*. Detailed strategies to handle these obstacles will be presented in Section 4.4.*

4.3 Existence of profile via phase-portrait method

This section is devoted to the existence of smooth profile solving (4.1.8) when $\mu < \frac{1}{3}$ with the appropriate choices of β. Firstly, note that for any radially symmetric function $R(y)$ satisfying $R(y) \to 0$ as $|y| \to \infty$,

$$\nabla \Delta^{-1} R(y) = \frac{y}{|y|^3} \int_0^{|y|} R(s) s^2 ds, \tag{4.3.1}$$

Then, under the radial symmetric assumption, (4.1.8) becomes

$$\left(\frac{1}{r^2} \int_0^r Q(s) s^2 ds - \beta r\right) Q'(r) - Q + (1 - \mu) Q^2 = 0. \tag{4.3.2}$$

If $f_Q \neq \beta$, by introducing f_Q (cf. (4.1.14)), then we obtain an ODE system of $(Q(r), f(r))$ for $r > 0$:

$$\begin{cases} Q'(r) = \frac{1}{\beta - f_Q} \frac{(1-\mu)Q^2 - Q}{r}, \\ f_Q'(r) = \frac{Q - 3f_Q}{r}. \end{cases} \tag{4.3.3}$$

Remark 4.3.1. *For notational simplicity, in the subsequent discussion of this section, we introduce the following notations: we write $f := f_Q$, and $Q_i := [Q]_i$ and $f_i := [f]_i$ for any $i \in \mathbb{Z}_{\geq 0}$, where f_Q is given in (4.3.3), $[Q]_i$ and $[f]_i$ are the $2i$-th order coefficients of Taylor expansion of f and Q at the origin respectively (cf. (4.1.19)).*

Our main result of this section is as follows, which gives the existence of the nontrivial smooth solutions to (4.3.3) and the asymptotic behavior of the related solutions when $r = 0$ and $r \to \infty$.

Lemma 4.3.2. *For any fix $\mu \in [0, \frac{1}{3})$, let $\beta := \beta(j_0, \mu) = \frac{1}{3(1-\mu)} + \frac{1}{2j_0}$ with arbitrarily fixed $j_0 \geq J > 1$ where J is defined in (4.1.5). Then, for any $Q_{j_0} < 0$, there exists a unique smooth solution $(Q(r), f(r))$ to (4.3.3) on $r \geq 0$ with initial conditions:*

$$(Q(0), f(0)) = \left(\frac{1}{1-\mu}, \frac{1}{3(1-\mu)} \right) \quad \text{and} \quad \partial_r^{2j_0} Q(0) = (2j_0)! Q_{j_0}. \tag{4.3.4}$$

Moreover, the solution $(Q(r), f(r))$ satisfies

1. *$(Q(r), f(r))$ is analytic near the origin. Precisely, there exists $\epsilon > 0$ such that $(Q(r), f(r))$ can be expressed as Taylor series*

$$Q(r) = \sum_{j=0}^{\infty} Q_j r^{2j} \quad \text{and} \quad f(r) = \sum_{j=0}^{\infty} f_j r^{2j}, \quad \forall\, r \in [0, \epsilon], \tag{4.3.5}$$

where $\{Q_j\}_{j \geq 0}$ and $\{f_j\}_{j \geq 0}$ satisfy the following recurrence relations

$$f_j = \frac{1}{2j+3} Q_j, \quad \forall\, j \geq 0, \tag{4.3.6}$$

and

$$Q_j = \begin{cases} \frac{1}{1-\mu}, & j = 0, \\ 0, & 0 < j < j_0, \\ \frac{\sum_{i=1}^{j-1} \left(\frac{2i}{2(j-i)+3} + (1-\mu) \right) Q_i Q_{j-i}}{2j \left(\beta - \frac{1}{2j} - \frac{1}{3(1-\mu)} \right)}, & j > j_0. \end{cases} \tag{4.3.7}$$

2. *Both $Q(r), f(r) > 0$ are strictly positive for $r \in [0, \infty)$ and monotonically decrease from $Q_0 = \frac{1}{1-\mu}$ (respectively, $f_0 = \frac{1}{3(1-\mu)}$) to 0 as $r \to +\infty$.*

3. *For any $j \geq 0$, there exists a constant $C = C(\beta, Q, j) > 0$ such that*

$$|\partial_r^{(j)} Q(r)| + |\partial_r^{(j)} f(r)| \leq C \langle r \rangle^{-2-j} \quad \text{for} \quad r \in \overline{\mathbb{R}^+}. \tag{4.3.8}$$

In particular, $Q \in H^\infty(\mathbb{R}^3)$.

Proof. Step 1. Solve (4.3.3) via Taylor expansion. We assume that $(f(r), Q(r))$ can be expanded into the following forms:

$$f(r) = \sum_{j=0}^{\infty} f_j r^{2j}, \quad Q(r) = \sum_{j=0}^{\infty} Q_j r^{2j},$$

which yields that

$$r f'(r) = \sum_{j=0}^{\infty} 2j f_j r^{2j}, \quad r Q'(r) = \sum_{j=0}^{\infty} 2j Q_j r^{2j}.$$

Next, we substitute these expressions into the ODE system (4.3.3) and match coefficients to establish (4.3.6) and (4.3.7). Precisely, we first expand the second equation in (4.3.3) to get that

$$\sum_{j=0}^{\infty} 2j f_j r^{2j} = \sum_{j=0}^{\infty} (Q_j - 3f_j) r^{2j}.$$

which directly induces (4.3.6).

Next, we expand $(\beta - f) r \partial_r Q = (1-\mu) Q^2 - Q$, the first equation in (4.3.3). Using the identity

$$\left(\sum_{j=0}^{\infty} h_j r^{2j} \right) \left(\sum_{j=0}^{\infty} k_j r^{2j} \right) = \sum_{j=0}^{\infty} \left(\sum_{i=0}^{j} k_i h_{j-i} \right) r^{2j},$$

the left-hand side of the equation becomes

$$\begin{aligned} \left(\beta - \sum_{j=0}^{\infty} f_j r^{2j} \right) \left(\sum_{j=0}^{\infty} 2j Q_j r^{2j} \right) &= \beta \sum_{j=0}^{\infty} 2j Q_j r^{2j} - \sum_{j=0}^{\infty} \left(\sum_{i=0}^{j} 2i Q_i f_{j-i} \right) r^{2j} \\ &= \sum_{j=0}^{\infty} \left(2\beta j - \sum_{i=0}^{j} 2i Q_i f_{j-i} \right) r^{2j}, \end{aligned}$$

and the right-hand side of the equation becomes

$$(1-\mu)\sum_{j=0}^{\infty}\left(\sum_{i=0}^{j}Q_iQ_{j-i}\right)r^{2j}-\sum_{j=0}^{\infty}Q_jr^{2j}=\sum_{j=0}^{\infty}\left((1-\mu)\sum_{i=0}^{j}Q_iQ_{j-i}-Q_j\right)r^{2j}.$$

Comparing the coefficients on both sides, we obtain that

$$2\beta jQ_j-\sum_{i=0}^{j}2iQ_if_{j-i}=(1-\mu)\sum_{i=0}^{j}Q_iQ_{j-i}-Q_j,\quad\forall\ j\geq 0.$$

Replacing f_{j-i} by using (4.3.6), we decouple the recurrence relations for $\{Q_j\}_{j=0}^{\infty}$ into

$$2\beta jQ_j-\sum_{i=0}^{j}\frac{2i}{2(j-i)+3}Q_iQ_{j-i}=(1-\mu)\sum_{i=0}^{j}Q_iQ_{j-i}-Q_j,\quad\forall\ j\geq 0.$$

In particular,

$$(1-\mu)Q_0^2-Q_0=0,\quad\text{for}\ \ j=0,$$

and $Q(0)=Q_0=\frac{1}{1-\mu}$ exactly solves this equation. For $j\geq 1$, we isolate the highest index term Q_j to obtain

$$2j\Big(\beta-\frac{1}{2j}-\frac{1}{3(1-\mu)}\Big)Q_j=\sum_{i=0}^{j-1}\left(\frac{2i}{2(j-i)+3}+(1-\mu)\right)Q_iQ_{j-i}.$$

In particular, applying $j=1$ to the equation together with the fact that $\beta<\frac{1}{2}$ and $1+\frac{1}{3(1-\mu)}>1$, it induces

$$2\left(\beta-\left(\frac{1}{2}+\frac{1}{3(1-\mu)}\right)\right)Q_1=0,\quad\Rightarrow\quad Q_1=0.$$

Iteratively, with the choice of $\beta=\frac{1}{3(1-\mu)}+\frac{1}{2j_0}$,

$$\beta-\frac{1}{2j}-\frac{1}{3(1-\mu)}=\frac{1}{2j_0}-\frac{1}{2j}\neq 0,\quad\Rightarrow\quad Q_j=0,\quad\forall\ 2\leq j<j_0.$$

Similarly, with $\partial_r^{2j_0}Q(0)=Q_{j_0}<0$ determined in (4.3.4),

$$\beta-\frac{1}{2j}-\frac{1}{3(1-\mu)}\neq 0,\ \Rightarrow\ Q_j=\frac{\sum_{i=1}^{j-1}\left(\frac{2i}{2(j-i)+3}+(1-\mu)\right)Q_iQ_{j-i}}{2j\left(\beta-\frac{1}{2j}-\frac{1}{3(1-\mu)}\right)},\quad\forall\ j>j_0.$$

Hence, we obtain the recurrence relation (4.3.7).

We remark here that under the previous argument, if $\beta - \frac{1}{2j} - \frac{1}{3(1-\mu)} \neq 0$ for all $j \geq 1$, then $Q_j = 0$ for all $j \geq 1$. In this case, $Q(r) \equiv \frac{1}{1-\mu}$, which is a constant solution and not our focus. Hence, to obtain a non-constant solution, we must give the vanishing condition that $\beta - \frac{1}{2j_0} - \frac{1}{3(1-\mu)} = 0$ for some $j_0 \geq 2$, which exactly matches the choice of β.

Step 2. Analyticity of (Q, f) near the origin. First of all, recalling [184, Lemma B.1], there exists a universal constant $a > 0$ such that for all $L \in \mathbb{N}$,

$$\sum_{\substack{i+j=L \\ i,j>0}} \frac{1}{i^2 j^2} \leq \frac{a}{L^2}, \tag{4.3.9}$$

which, by using the fact that $Lj \geq i$ for any $1 \leq i, j \leq L-1$, yields that

$$\sum_{\substack{i+j=L \\ i,j>0}} \frac{1}{ij^3} \leq \sum_{\substack{i+j=L \\ i,j>0}} \frac{L}{i^2 j^2} \leq \frac{a}{L}. \tag{4.3.10}$$

Next, we are devoted to the estimates of coefficients of the Taylor expansion determined in (4.3.7). Precisely, we claim that there exists $K, \alpha > 0$, such that

$$|Q_j| \leq \frac{K^{j-\alpha}}{j^2}, \quad \forall\, j \geq j_0, \tag{4.3.11}$$

where $K, \alpha > 0$ satisfy

$$|Q_{j_0}| \leq \frac{K^{j_0-\alpha}}{j_0^2} \quad \text{and} \quad \frac{aK^{-\alpha}}{\frac{1}{2j_0} - \frac{1}{2(j_0+1)}} < \frac{1}{2}. \tag{4.3.12}$$

In fact, we proceed by induction. For $1 \leq j \leq j_0$, it automatically holds with (4.3.7) and (4.3.12). We now assume that (4.3.11) holds for any $j_0 \leq j \leq L$, then by using (4.3.7), (4.3.9), (4.3.10), (4.3.11) and (4.3.12), we obtain

$$\begin{aligned}
Q_{L+1} &= \frac{1}{2(L+1)\left(\beta - \frac{1}{2(L+1)} - \frac{1}{3(1-\mu)}\right)} \sum_{\substack{i+j=L+1 \\ i,j>0}} \left(\frac{2i}{2j+3} + (1-\mu)\right) Q_i Q_j \\
&\leq \frac{1}{2(L+1)\left(\frac{1}{2j_0} - \frac{1}{2(j_0+1)}\right)} \sum_{\substack{i+j=L+1 \\ i,j>0}} \left(\frac{2i}{2j+3} + (1-\mu)\right) \frac{K^{L+1-2\alpha}}{i^2 j^2} \\
&\leq \frac{K^{L+1-2\alpha}}{2(L+1)\left(\frac{1}{2j_0} - \frac{1}{2(j_0+1)}\right)} \sum_{\substack{i+j=L+1 \\ i,j>0}} \left(\frac{1}{ij^3} + \frac{1}{i^2 j^2}\right) \leq \frac{K^{L+1-\alpha}}{(L+1)^2},
\end{aligned}$$

which yields the validity of (4.3.11) with $j = L+1$, hence we have closed the induction and verified the claim.

The analyticity of Q at the origin then follows from Cauchy's convergence criterion for power series. Precisely, there exists $\epsilon = \epsilon(K) \ll 1$ such that the series

$$Q(r) = \sum_{j=0}^{\infty} Q_j r^{2j}, \qquad \text{with } Q_j \text{ determined in (4.3.7)},$$

is absolutely convergent on $[0, \epsilon]$. The analyticity of $f(r) = \sum_{j=0}^{\infty} f_j r^{2j}$ on $r \in [0, \epsilon]$ simply follows (4.3.6) and the analyticity on $[0, \epsilon]$ of Q.

Step 3. Extend (Q, f) to $[0, +\infty)$. At this stage, our objective is to extend the previously constructed solution to $\overline{\mathbb{R}^+}$, and verify that both $f(r)$ and $Q(r)$ will asymptotically decay to 0 and the solution curve (see the red curve in Figure 4.3)) remains within $\mathcal{M}$ with

$$\mathcal{M} := \left\{ (Q, f) \middle| \, 0 < Q < Q_0, \ \frac{Q}{3} < f < \beta \right\}.$$

Here, recalling (4.3.3) or Figure 4.3, we observe that both ($Q(r)$ and $f(r)$) are strictly decreasing once the solution curve stays within $\mathcal{M}$, i.e.

$$Q'(r) < 0, f'(r) < 0 \quad \text{if} \quad (Q(r), f(r)) \in \mathcal{M}. \tag{4.3.13}$$

We first claim that there exists $\nu \in (0, \epsilon)$ such that $(Q(r), f(r)) \in \mathcal{M}, \ \forall\, r \in (0, \nu]$. To verify this, since

$$\frac{d^j}{dr^j} Q(0) = 0, \quad \forall\, 1 \le j < 2j_0 \qquad \text{and} \qquad \frac{d^{2j_0}}{dr^{2j_0}} Q(0) = (2j_0)! Q_{j_0} < 0,$$

it follows that both $Q(r)$ and $f(r)$ are strictly decreasing on $[0, \nu]$ for some $0 < \nu \le \epsilon \ll 1$. Consequently, the portion of the solution curve for $r \in [0, \nu]$, which begins at P_0 (4.1.15), remains within the lower-left area of P_0. Furthermore, it continues to be representable as a graph, meaning that there is a smooth function $\mathcal{F}$ with

$$f = \mathcal{F}(Q) \quad \text{on} \quad Q \in \left[Q_0 - \nu_Q, Q_0 \right],$$

for some $0 < \nu_Q \ll 1$. By chain rule, (4.3.3), (4.3.5), (4.3.6) (4.3.7) and $j_0 \ge J > 1$,

$$\begin{aligned} \mathcal{F}'(Q_0) = \frac{df}{dQ}\Big|_{(Q,f)=P_0} &= \lim_{r \to 0^+} \frac{(Q(r) - 3f(r))(\beta - f(r))}{(1-\mu)Q^2(r) - Q(r)} \\ &= \frac{\beta - f_0}{(1-\mu)Q_0} \frac{Q_{j_0} - 3f_{j_0}}{Q_{j_0}} = \frac{1}{2j_0} \left(1 - \frac{3}{2j_0 + 3} \right) < \frac{1}{2j_0} \le \frac{1}{4} < \frac{1}{3}. \end{aligned}$$

Thus, the slope of the solution curve at P_0 is strictly less than that of the line $f = \frac{1}{3}Q$, which contains the lower boundary of $\mathcal{M}$. So by adjusting $0 < \nu \ll 1$ if necessary and using the smoothness of $\mathcal{F}$, we conclude that the curve $\{(Q(r), f(r))\}_{r\in(0,\nu]}$ lies above $f = \frac{1}{3}Q$. This has verified the claim.

Next, we prove that the solution curve remains in $\mathcal{M}$ for all $r \in (0, \infty)$. In fact, ensured by the Cauchy-Lipschitz theory, with the strict decay property of the solution within $\mathcal{M}$ (see (4.3.13)), there are only three possible scenarios:

1. there exists $r_{E1} \in (0, \infty]$, such that the solution exits the region $\mathcal{M}$ at a point $\left(Q_{E1}, \frac{Q_{E1}}{3}\right)$ through the line $f = \frac{Q}{3}$, where $Q_{E1} = Q(r_{E1}) \in (0, Q_0)$;
2. there exists $r_{E2} \in (0, \infty]$, such that the solution exits the region $\mathcal{M}$ at a point $(0, f_{E2})$ through the f-axis, where $f_{E2} = f(r_{E2}) \in (0, f_0)$;
3. the solution escapes $\mathcal{M}$ at the origin O.

We now eliminate scenarios (1) and (2) by contradiction arguments.

For the scenario (1), we assume that it occurs. Since $(Q(r), f(r))$ is strictly decreasing before reaching the line $f = \frac{Q}{3}$, together with Cauchy-Lipschitz theory, the solution curve can be further extended on $Q \in [Q_{E1}, Q_0]$ and represented as $Q \mapsto \mathcal{F}(Q)$ with $\mathcal{F}$ a smooth function.

On the one hand, since $\{(Q(r), f(r))\}_{r\in(0,r_{E_1})} \subset \mathcal{M}$, it follows that $f(r) > \frac{1}{3}Q(r)$ for any $r \in (0, r_{E_1})$ and $f(r), Q(r)$ is strictly decreasing on $r \in (0, r_{E_1})$, thus together with $f(r_{E_1}) = \frac{1}{3}Q(r_{E_1})$,

$$\mathcal{F}'(Q_{E1}) = \lim_{r\to r_{E_1}^-} \frac{f(r) - f(r_{E_1})}{Q(r) - Q(r_{E_1})} \geq \lim_{r\to r_{E_1}^-} \frac{\frac{1}{3}Q(r) - \frac{1}{3}Q(r_{E_1})}{Q(r) - Q(r_{E_1})} = \frac{1}{3}.$$

On the other hand, recalling the ODE system (4.3.3) together with $f(r_{E_1}) = \frac{1}{3}Q(r_{E_1})$ again, we compute

$$\mathcal{F}'(Q_{E1}) = \lim_{r\to r_{E1}^-} \frac{(Q(r) - 3f(r))(\beta - f(r))}{(1-\mu)Q^2(r) - Q(r)} = 0 < \frac{1}{3},$$

which contradicts our assumption and hence we have ruled out the scenarios (1).

Regarding the second scenario (2), by an argument analogous to that of the first case, the solution curve can be extended on $f \in [f_{E2}, f_0]$ and represented as $f \mapsto Q(f)$

with Q a smooth function, and thus

$$\{(Q(r), f(r))\}_{r\in(0,r_{E_2})} = \{(Q(f), f)\}_{f\in(f_{E_2},f_0)} \subset \mathcal{M}. \tag{4.3.14}$$

Recall (4.3.3) again, $Q(f)$ is governed by the equation

$$Q' = \frac{(1-\mu)Q^2 - Q}{(Q-3f)(\beta - f)}, \quad \forall\, f \in [f_{E2}, f_0], \tag{4.3.15}$$

whose local wellposedness near $f = f_{E_2}$ is well established by the Cauchy-Lipschitz theory. Nevertheless, besides the nontrivial solution curve given in (4.3.14), we observe that $\{(0, f)\}_{f\in[f_{E_2},f_0]}$ is also a satisfied solution curve, which leads to a contradiction and hence we have ruled out the scenario (2).

Consequently, we have shown that the solution curve could only exist $\mathcal{M}$ at the origin O, i.e. in scenario (3). To conclude this step, it suffices to verify that

$$\lim_{r\to\infty}(Q(r), f(r)) = (0, 0). \tag{4.3.16}$$

In fact, similar to the previous argument, the solution curve can be further smoothly extended and represented as $Q \mapsto \mathcal{F}(Q)$ on $[0, Q_0]$. Additionally, using (4.3.13) again, the function $r \mapsto Q(r)$ is one-to-one and thus r can be expressed by $r = \mathcal{R}(Q)$ on $Q \in [0, Q_0]$, with $\mathcal{R}$ solving

$$\mathcal{R}'(Q) = \mathcal{R}(Q)\frac{\beta - \mathcal{F}(Q)}{(1-\mu)Q^2 - Q},$$

which follows from the inverse function theorem and (4.3.3). Then, a direct computation together with (4.3.4), (4.3.6), (4.3.13) and the fact that $\mathcal{R}(\frac{Q_0}{2}) \in (0, \infty)$ yields that

$$\begin{aligned}\lim_{Q\to 0^+} \ln \mathcal{R}(Q) &= \ln \mathcal{R}\left(\frac{Q_0}{2}\right) + \lim_{Q\to 0^+}\int_{\frac{Q_0}{2}}^{Q} \frac{\beta - \mathcal{F}(\tilde{Q})}{(1-\mu)\tilde{Q}^2 - \tilde{Q}}\, d\tilde{Q} \\ &> \ln \mathcal{R}\left(\frac{Q_0}{2}\right) + (\beta - f_0)\lim_{Q\to 0+}\int_{\frac{Q_0}{2}}^{Q}\left(\frac{1}{\tilde{Q} - \frac{1}{1-\mu}} - \frac{1}{\tilde{Q}}\right) d\tilde{Q} = +\infty,\end{aligned}$$

and thus we have verified (4.3.16) and finished this step.

Step 4. Uniqueness. Assume that there exist two smooth solutions $(Q^{(1)}, f^{(1)})$ and $(Q^{(2)}, f^{(2)})$ solving the system (4.3.3) with the same initial conditions (4.3.4), then

by the C^∞ continuity of Q at the origin and a similar argument as in *Step 1* and *Step 3*, it can be verified that $(Q^{(i)}, f^{(i)})$ satisfies

$$Q_0^{(1)} = Q_0^{(2)} = \frac{1}{1-\mu}, \quad Q_j^{(1)} = Q_j^{(2)} = 0 \quad (\text{with } 1 \le j < j_0 - 1), \quad Q_{j_0}^{(1)} = Q_{j_0}^{(2)} < 0, \tag{4.3.17}$$

and both $Q^{(i)}(r)$ and $f^{(i)}(r)$ are strictly decreasing on $\mathbb{R}^+$ for both $i = 1, 2$. Next, we study the difference between these two solutions. Denote

$$(\triangle Q, \triangle f) := (Q^{(1)} - Q^{(2)}, f^{(1)} - f^{(2)}),$$

which solves the following ODE system:

$$\begin{cases} (\triangle Q)' = \frac{1}{\beta - f^{(1)}} \frac{(1-\mu)(Q^{(1)}+Q^{(2)})\triangle Q - \triangle Q}{r} - \frac{(1-\mu)(Q^{(2)})^2 - Q^{(2)}}{r} \frac{\triangle f}{(\beta - f^{(1)})(\beta - f^{(2)})}, \\ (\triangle f)' = \frac{\triangle Q - 3\triangle f}{r}. \end{cases}$$

We define

$$\mathcal{E}(r) := |\triangle Q(r)|^2 + |\triangle f(r)|^2,$$

and compute

$$\begin{aligned} \mathcal{E}' &= I + II, \\ \text{with} \quad I &= 2\frac{\triangle f \triangle Q - 3(\triangle f)^2}{r} + 2\frac{(1-\mu)(Q^{(1)} + Q^{(2)}) - 1}{(\beta - f^{(1)})r}(\triangle Q)^2, \\ II &= -2\frac{(1-\mu)(Q^{(2)})^2 - Q^{(2)}}{r} \frac{1}{(\beta - f^{(2)})(\beta - f^{(1)})} \triangle f \triangle Q, \end{aligned}$$

By using Cauchy inequality, strictly decreasing fact of $Q^{(i)}(r)$ and $f^{(i)}(r)$, (4.3.4) and (4.3.6), we bound I by

$$I \le \left(1 + 2\frac{(1-\mu)(Q^{(1)}(0) + Q^{(2)}(0)) - 1}{(\beta - f^{(1)}(0))}\right)\frac{\mathcal{E}(r)}{r} = (4j_0 + 1)\frac{\mathcal{E}(r)}{r}.$$

Applying condition (4.3.17) along with the Cauchy's inequality, there exists constant $M = M(j_0, \mu, Q^{(1)}, Q^{(2)}) > 0$ such that

$$II \le M|\triangle f \triangle Q| \le M\mathcal{E}(r),$$

which yields that

$$\mathcal{E}'(r) = I + II \le (4j_0 + 1)\frac{\mathcal{E}(r)}{r} + M\mathcal{E}(r).$$

Applying Gronwall's inequality,

$$0 \le \mathcal{E}(r) \le \mathcal{E}(\delta) \left(\frac{r}{\delta}\right)^{4j_0+1} e^{M(r-\delta)}, \qquad \text{uniformly for } r, \delta > 0.$$

Since $(Q^{(i)}, f^{(i)})$ (with $i = 1, 2$) satisfy the same initial conditions (4.3.17), it follows that

$$\partial_r^j \mathcal{E}(r)\Big|_{r=0} = 0, \quad \forall\, 0 \le j \le 4j_0 + 2,$$

which implies that

$$\lim_{\delta \to 0+} \frac{\mathcal{E}(\delta)}{\delta^{4j_0+1}} e^{-M\delta} = 0.$$

By the squeezing theorem, it follows that $\mathcal{E}(r) \equiv 0$ for any $r \ge 0$. Consequently, we conclude $Q^{(1)} \equiv Q^{(2)}$ and this completes the uniqueness.

Step 5. Decay estimate of Q and its derivative. Based on the discussion in *Step 3*, both $Q(r) \searrow 0$ and $f(r) \searrow 0$ as $r \to \infty$. Next for any fixed $\nu_0 > 0$, recalling the choice of $\beta = f_0 + \frac{1}{2j_0}$, there exists $M_0 = M_0(\nu_0) \gg 1$ such that

$$0 < (1-\nu_0)Q(r) \le Q(r) - (1-\mu)Q^2(r) \le Q(r), \quad \forall\, r \ge M_0,$$

$$\text{and} \quad -2j_0 \le -\frac{1}{\beta - f} \le -\frac{1}{\beta} < 0.$$

From the ODE (4.3.3) satisfied by $Q(r)$, these inequalities imply

$$-\frac{2j_0 Q}{r} \le Q'(r) \le -\frac{(1-\nu_0)Q}{\beta r}, \qquad \forall\, r \ge M_0,$$

hence

$$Q(r) \le Q(M_0) \left(\frac{r}{M_0}\right)^{-\frac{1}{\beta}(1-\nu_0)}, \quad \forall\, r \ge M_0.$$

In particular, since $\beta < \frac{1}{2}$, we can choose $\nu_0 > 0$ sufficiently small such that $\frac{1-\nu_0}{\beta} < 2$. This allows us to find a constant $C = C(\beta, M_0) > 0$ such that

$$Q(r) \le C \langle r \rangle^{-2}, \quad \forall\, r \ge 0. \tag{4.3.18}$$

Moreover, using the second equation in (4.3.3), there exists a constant $C = C(\beta, M_0) > 0$ such that

$$0 < f(r) = \frac{1}{r^3} \int_0^r Q(s) s^2 ds \le \frac{C}{r^3} \int_0^r \frac{s^2}{\langle s \rangle^2} ds \le C \langle r \rangle^{-2}, \quad \forall\, r \ge 0. \tag{4.3.19}$$

For the derivatives of $Q(r)$ and $f(r)$, applying ∂_r^{j-1} (with $j \geq 1$) to (4.3.3) and using the base decay estimates (4.3.18) and (4.3.19), a straightforward induction then shows that

$$|\partial_r^{(j)} Q(r)| \leq C \langle r \rangle^{-2-j}, \tag{4.3.20}$$

for some constant $C = C(\beta, M_0, j) > 0$.

Finally, combining the C^∞ smoothness of Q, the above decay bound on $\partial_r^{(j)} Q$ (cf. (4.3.18) and (4.3.20)), and the standard identity $Dg = \frac{x}{|x|} \cdot \nabla g$ for radial function g, we can verify that $Q \in H^\infty(\mathbb{R}^3)$. This completes the proof of the lemma. □

Based on Lemma 4.3.2, one naturally obtains the existence of a smooth profile satisfying (4.1.8). More precisely, we have the following result.

Corollary 4.3.3 (Existence of smooth self similar profile)**.** *Under the assumptions of Lemma 4.3.2, for any fixed $[Q]_{j_0} < 0$, there exists a unique smooth radial symmetric solution $Q \in H^\infty(\mathbb{R}^3)$ to (4.1.8) satisfying*

$$Q(0) = \frac{1}{1-\mu}, \quad \text{and } \partial_r^{(2j_0)} Q(0) = (2j_0)![Q]_{j_0} < 0. \tag{4.3.21}$$

In particular, $Q(r)$ is strictly radially decreasing with respect to $r \geq 0$ and satisfies the decay estimates (4.3.8).

4.4 Linear theory

Starting from this section, we fix $\mu \in [0, \frac{1}{3})$ and a positive integer $j_0 \geq \frac{3(1-\mu)}{1-3\mu} + 1 > 1$, then we have

$$\beta = \frac{1}{3(1-\mu)} + \frac{1}{2j_0} \in \left(0, \frac{1}{2}\right). \tag{4.4.1}$$

We now introduce the self-similar coordinate:

$$y = \frac{x}{\lambda^{2\beta}}, \quad \frac{d\tau}{dt} = \frac{1}{\lambda^2}, \quad \tau\Big|_{t=0} = 0, \quad \frac{\lambda_\tau}{\lambda} = -\frac{1}{2}, \tag{4.4.2}$$

and the corresponding renormalization

$$\rho(t,x) = \frac{1}{\lambda^2} \Psi(\tau, y). \tag{4.4.3}$$

The equation (KS-D) can be mapped into the renormalization system of Ψ:

$$\partial_\tau \Psi = \lambda^{2-4\beta} \Delta \Psi - \Psi - \beta y \cdot \nabla \Psi + \nabla \Psi \cdot \nabla \Delta^{-1} \Psi + (1-\mu)\Psi^2, \tag{4.4.4}$$

where, from (4.4.2), the function λ is

$$\lambda(\tau) = \lambda_0 e^{-\frac{1}{2}\tau}, \quad \forall\, \tau \geq 0, \tag{4.4.5}$$

with initial data given by $\lambda\Big|_{\tau=0} = \lambda_0 > 0$.

Next, we consider a solution to the above equation in perturbative form

$$\Psi = Q + \varepsilon,$$

where Q is a solution to the PDE (4.1.8), corresponding to the fixed parameters μ and j_0, whose existence is established in Lemma 4.3.2. Direct computation shows the residue ε solves

$$\partial_\tau \varepsilon = \mathcal{L}\varepsilon + \lambda^{2-4\beta}\Delta\Psi + N(\varepsilon), \tag{4.4.6}$$

where $\mathcal{L}$ is the linear operator

$$\mathcal{L}\varepsilon = -(\varepsilon + \beta y \cdot \nabla\varepsilon) + \nabla\varepsilon \cdot \nabla\Delta^{-1}Q + \nabla Q \cdot \nabla\Delta^{-1}\varepsilon + 2(1-\mu)Q\varepsilon, \tag{4.4.7}$$

and $N(\varepsilon)$ is the nonlinear term

$$N(\varepsilon) = \nabla\varepsilon \cdot \nabla\Delta^{-1}\varepsilon + (1-\mu)\varepsilon^2. \tag{4.4.8}$$

For the linear operator $\mathcal{L}$ given in (4.4.7), we will prove its coercivity in the sense of L^2_w for some weight w sufficiently singular at the origin. Precisely, we have the following result:

Proposition 4.4.1 (Coercivity of $\mathcal{L}$ in L^2_w)**.** *For $\mathcal{L}$ given in (4.4.7), there exists a weight defined by*

$$w(y) := |y|^{-A} + B, \tag{4.4.9}$$

where $(A, B) \in \mathbb{Z}_+ \times \mathbb{R}_+$ satisfy $A \gg 4j_0 + 3$ and $\frac{A}{4} \in \mathbb{Z}_{>0}$, such that $\mathcal{L}$ satisfies the coercivity in the following sense:

$$(\mathcal{L}g, g)_{L^2_w} \leq -\frac{1}{8}\|g\|^2_{L^2_w}, \quad \forall\, g \in L^2_w(\mathbb{R}^3) \quad \textit{radially symmetric}. \tag{4.4.10}$$

Remark 4.4.2. *Due to the limited quantitative properties of Q available from Lemma 4.3.2 and the nonlocal structure of the operator $\mathcal{L}$, it is difficult to identify the optimal value of A explicitly, in contrast with the situation for the heat equation discussed in Section 4.2. Nevertheless, the existence of some A satisfying (4.4.10) is sufficient for our subsequent analysis.*

Proof of Proposition 4.4.1. First of all, for any $g \in L^2_w$, by integration by parts, (4.3.1) and the fact that $y \cdot \nabla w = -A|y|^{-A}$,

$$-\beta \int (y \cdot \nabla g) g w dy = \frac{\beta}{2}(-A+3) \int g^2 |y|^{-A} dy + \frac{3B\beta}{2} \int g^2 dy,$$

and

$$\int \left(\nabla g \cdot \nabla \Delta^{-1} Q\right) g w dy = -\frac{1}{2} \int Q g^2 w dy + \frac{A}{2} \int g^2 f_Q |y|^{-A} dy, \tag{4.4.11}$$

with which, recalling the definition of $\mathcal{L}$ (see (4.4.7)), we split $(\mathcal{L}g, g)_{L^2_w}$ into the following three parts:

$$(\mathcal{L}g, g)_{L^2_w} = I_{SI} + I_{LO} + I_{NLO}, \tag{4.4.12}$$

where

$$I_{SI} := \left(-1 + \frac{\beta}{2}(-A+3)\right) \int g^2 |y|^{-A} dy + \frac{A}{2} \int g^2 f_Q |y|^{-A} dy + \left(\frac{3}{2} - 2\mu\right) \int g^2 Q |y|^{-A} dy,$$
$$I_{LO} := \left(-B + \frac{3B\beta}{2}\right) \int g^2 dy + \left(\frac{3}{2} - 2\mu\right) B \int g^2 Q dy, \quad I_{NLO} := \int \left(\nabla Q \cdot \nabla \Delta^{-1} g\right) g w dy.$$

Here I_{SI} denotes the sum of local terms involving the singular weight $|y|^{-A}$, I_{LO} denotes the sum of local terms without singular weight, and I_{NLO} denotes the sum of nonlocal terms.

Estimate of I_{SI}. By Lemma 4.3.2, Q and f_Q are both radially decreasing. Moreover, recalling the initial conditions of $(Q(0), f_Q(0))$ given in (4.3.4) and the definition of β in (4.4.1), I_{SI} can be estimated by

$$\begin{aligned} I_{SI} &\le \left(-1 + \frac{\beta}{2}(-A+3)\right) \int g^2 |y|^{-A} dy + \left(\frac{A}{2} f_Q(0) + \left(\frac{3}{2} - 2\mu\right) Q(0)\right) \int g^2 |y|^{-A} dy \\ &= \left(-1 + \frac{\frac{1}{3(1-\mu)} + \frac{1}{2j_0}}{2}(-A+3) + \frac{A}{6(1-\mu)} + \left(\frac{3}{2} - 2\mu\right) \frac{1}{1-\mu}\right) \int g^2 |y|^{-A} dy \\ &= \left(1 + \frac{-A+3}{4j_0}\right) \int g^2 |y|^{-A} dy, \end{aligned} \tag{4.4.13}$$

where the coercivity becomes linearly stronger as $A \gg 1$ grows.

We remark here that the derivation of (4.4.13) is rather delicate. Precisely, one might naturally expect that the term $g + \beta y \cdot \nabla g$ alone provides enhanced coercivity with increasing A. However, this gain is offset by the contribution $\frac{A}{2} \int g^2 f_Q |y|^{-A} dy$,

which greatly diminishes the coercivity originating from the scaling term. Nevertheless, with the estimate given in (4.4.13), we find that this essential obstacle can be almost exactly controlled via the fact that $\beta - f_Q(r) \geq \beta - f_Q(0) = \frac{1}{2j_0} > 0$.

Estimate of I_{LO}. Again, by Lemma 4.3.2, Q is radially decreasing and $\left\| |y|^{-\frac{1}{2}} \partial_r Q \right\|_{L^2} < \infty$, so there exists $R_1 = R_1(Q) \gg 1$ such that

$$\frac{3}{2} Q(R_1) \leq \frac{1}{1000} \quad \text{and} \quad \left\| |y|^{-\frac{1}{2}} \partial_r Q \right\|_{L^2(|x| \geq R_1)} \leq \frac{1}{5000}. \tag{4.4.14}$$

Noting that $\beta < \frac{1}{2}$, we may now estimate the lower-order term I_{LO} as follows:

$$\begin{aligned} I_{LO} &\leq \left(-B + \frac{3}{4} B \right) \int g^2 dy + \left(\frac{3}{2} - 2\mu \right) B \left(\int_{|y| \leq R_1} g^2 Q dy + \int_{|y| \geq R_1} g^2 Q dy \right) \\ &\leq -\frac{B}{4} \int g^2 dy + \left(\frac{3}{2} - 2\mu \right) BQ(0) R_1^A \int_{|y| \leq R_1} g^2 |y|^{-A} dy + \left(\frac{3}{2} - 2\mu \right) BQ(R_1) \|g\|_{L^2}^2 \\ &= -\frac{3B}{16} \|g\|_{L^2}^2 + \left(\frac{3}{2} - 2\mu \right) BQ(0) R_1^A \|g\|_{L^2_{|y|^{-A}}}^2. \end{aligned} \tag{4.4.15}$$

Given the limited qualitative information about Q, the term involving Q cannot be directly controlled, especially the integral of $\int_{|y| \leq R_1} \int g^2 Q dy$, where $Q(y) = O(1)$ on $B(0, R_1)$. To address this, we will absorb this term using the previous coercivity (4.4.13) by choosing $0 < B \ll 1$ sufficiently small.

Estimate of I_{NLO}. By Cauchy's inequality, the nonlocal term involving in I_{NLO} can be controlled pointwise by

$$\begin{aligned} &\frac{1}{r^2} \int_0^r |g(s)| s^2 ds = \frac{1}{4\pi r^2} \int_{B(0,r)} |g(y)| dy = \frac{1}{4\pi r^2} \int_{B(0,r)} |g(y)| |y|^{-\frac{A}{2}} |y|^{\frac{A}{2}} dy \\ &\leq \frac{1}{4\pi r^2} \left(\int_{B(0,r)} g^2(y) |y|^{-A} dy \right)^{\frac{1}{2}} \left(\int_{B(0,r)} |y|^A dy \right)^{\frac{1}{2}} = \sqrt{\frac{1}{4\pi(A+3)}} r^{\frac{A}{2} - \frac{1}{2}} \|g\|_{L^2_{|y|^{-A}}}, \end{aligned} \tag{4.4.16}$$

and this induces that

$$\begin{aligned} &\left| \int \left(\nabla Q \cdot \nabla \Delta^{-1} g \right) g |y|^{-A} dy \right| = \int |\partial_r Q| \left(\frac{1}{|y|^2} \int_0^{|y|} |g(s)| s^2 ds \right) |g| |y|^{-A} dy \\ &\leq \sqrt{\frac{1}{4\pi(A+3)}} \|g\|_{L^2_{|y|^{-A}}} \int |\partial_r Q| |g| |y|^{-\frac{A}{2} - \frac{1}{2}} dy \leq \sqrt{\frac{1}{4\pi(A+3)}} \left\| |y|^{-\frac{1}{2}} \partial_r Q \right\|_{L^2} \|g\|_{L^2_{|y|^{-A}}}^2, \end{aligned} \tag{4.4.17}$$

which, by choosing $A \gg 1$ sufficiently large, can be absorbed in (4.4.13). Similarly, repeating the previous argument, together with the choice of $R_1 \gg 1$ (see (4.4.14)), the L^2 component of I_{NLO} can be controlled by:

$$\begin{aligned} B\int \left(\nabla Q\cdot\nabla\Delta^{-1}g\right) g dy &\le B\|g\|_{L^2}\left(\int_{|y|\le R_1}|\partial_r Q||y|^{-\frac12}|g|dy+\int_{|y|\ge R_1}|\partial_r Q||y|^{-\frac12}|g|dy\right)\\ &\le \frac{B}{100}\|g\|_{L^2}^2+50BR_1^A\left\||\partial_r Q|y|^{-\frac12}\right\|_{L^2}^2\|g\|^2_{L^2_{|y|^{-A}}}+50B\left\||\partial_r Q|y|^{-\frac12}\right\|^2_{L^2(|y|\ge R_1)}\|g\|_{L^2}^2\\ &\le \frac{B}{50}\|g\|_{L^2}^2+50BR_1^A\left\||\partial_r Q|y|^{-\frac12}\right\|_{L^2}^2\|g\|^2_{L^2_{|y|^{-A}}}. \end{aligned} \tag{4.4.18}$$

The idea of handling this term is similar to I_{LO}, and we can use the smallness of $0<B\ll 1$ to absorb $B\int_{|y|\le R_1}\left(\nabla Q\cdot\nabla\Delta^{-1}g\right)gdy$ by the coercivity (4.4.13). Hence combining (4.4.17) and (4.4.18) yields

$$I_{NLO}\le\left(\sqrt{\frac{1}{4\pi(A+3)}}\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}+50BR_1^A\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}^2\right)\|g\|^2_{L^2_{|y|^{-A}}}+\frac{B}{50}\|g\|_{L^2}^2. \tag{4.4.19}$$

Conclusion of coercivity. Combining the estimates (4.4.13), (4.4.15) and (4.4.19), we obtain that

$$\begin{aligned} (\mathcal{L}g,g)_{L^2_w}\le -\frac18 B\|g\|_{L^2}^2+\left(1+\frac{-A+3}{4j_0}+\sqrt{\frac{1}{4\pi(A+3)}}\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}\right)\int g^2|y|^{-A}dy\\ +\left(\left(\frac32-2\mu\right)BR_1^AQ(0)+50BR_1^A\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}^2\right)\|g\|^2_{L^2_{|y|^{-A}}}, \end{aligned}$$

with $R_1=R_1(Q)>0$ chosen as in (4.4.14). Next, we choose $A\gg 4j_0+4$ with $\frac{A}{4}\in\mathbb{Z}_{>0}$ such that

$$1+\frac{-A+3}{4j_0}\le -1,\quad \text{and}\quad \sqrt{\frac{1}{4\pi(A+3)}}\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}\le\frac{1}{100}, \tag{4.4.20}$$

then we choose $0<B=B(A,R_1,\mu)\ll 1$ sufficiently small such that

$$\left(\frac32-2\mu\right)BR_1^AQ(0)+50BR_1^A\left\||y|^{-\frac12}\partial_r Q\right\|_{L^2}^2\le\frac{1}{100},$$

so that we obtain that

$$(\mathcal{L}g,g)_{L^2_w}\le -\frac18\|g\|^2_{L^2_{|y|^{-A}}}-\frac18 B\|g\|_{L^2}^2=-\frac18\|g\|^2_{L^2_w},$$

which completes the proof of (4.4.10). □

In addition, to study the nonlinear stability of Q in Section 4.5, it is necessary to establish the coercivity properties of the linearized operator $\mathcal{L}$ in a higher order Sobolev space $\dot{H}^{2K+4}$ with $K = \frac{A}{4}$. The precise statement is as follows:

Proposition 4.4.3 (Coercivity of $\mathcal{L}$ in $\dot{H}^{2K+4}$)**.** *Let $A > 0$ be as given in Proposition 4.4.1, and define $K := \frac{A}{4} \in \mathbb{Z}_{\geq 1}$. Then there exists constant $C = C(Q, A, B, j_0, \mu) > 0$ such that*

$$(\mathcal{L}g, g)_{\dot{H}^{2K+4}} \leq -\frac{1}{8}\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2}, \quad \forall\, g \in H^{2K+4}_{rad}. \tag{4.4.21}$$

Proof. By the definition of $\mathcal{L}$ given in (4.4.7),

$$(\mathcal{L}g, g)_{\dot{H}^{2K+4}} = I + II + III + IV, \tag{4.4.22}$$

with

$$I := -\left(\Delta^{K+2}\left((g + \beta y \cdot \nabla g)\right), \Delta^{K+2} g\right)_{L^2}, \quad II := \left(\Delta^{K+2}\left(\nabla g \cdot \nabla \Delta^{-1} Q\right), \Delta^{K+2} g\right)_{L^2},$$
$$III := \left(\Delta^{K+2}\left(\nabla Q \cdot \nabla \Delta^{-1} g\right), \Delta^{K+2} g\right)_{L^2}, \quad IV := \left(\Delta^{K+2}\left(2(1-\mu) Q g\right), \Delta^{K+2} g\right)_{L^2}.$$

First, observe that by scaling invariance, the term I can be computed as

$$I = -\frac{1}{2}\frac{d}{d\lambda}\bigg|_{\lambda=1} \left\|\Delta^{K+2}\left(\lambda g(\lambda^\beta y)\right)\right\|^2_{L^2} = -\left(\frac{\beta}{2}(4K+5) + 1\right)\|\Delta^{K+2} g\|^2_{L^2}. \tag{4.4.23}$$

For the term II, recalling (4.3.1), we rewrite $\nabla \Delta^{-1} Q = r f_Q \mathbf{e}_r$ with $\mathbf{e}_r = \frac{y}{|y|}$. Then by Lemma 4.3.2, [47, Lemma A.4], integration by parts, Gagliardo-Nirenberg inequality, and Young's inequality, there exists a uniform constant $C = C(K, Q) > 0$ such that

$$\begin{aligned} II \leq & \left(\left(\nabla \Delta^{K+2} g \cdot \nabla \Delta^{-1} Q\right), \Delta^{K+2} g\right)_{L^2} + 2(K+2)\left((\partial_r (r f_Q))\Delta^{K+2} g, \Delta^{K+2} g\right)_{L^2} \\ & + C\left\|r f_Q\right\|_{W^{2K+4,\infty}} \|g\|_{H^{2K+3}} \|g\|_{\dot{H}^{2K+4}} \\ \leq & 2(K+2)\left((\partial_r (r f_Q))\Delta^{K+2} g, \Delta^{K+2} g\right)_{L^2} + \frac{1}{100}\|g\|^2_{\dot{H}^{K+2}} + C\|g\|^2_{L^2}. \end{aligned} \tag{4.4.24}$$

Since $Q(r)$ is radially decreasing by Lemma 4.3.2,

$$f_Q(r) = \frac{1}{r^3}\int_0^r Q(s) s^2 ds \geq \frac{1}{r^3} Q(r) \int_0^r s^2 ds = \frac{1}{3} Q(r),$$

which yields that

$$\partial_r (r f_Q(r)) = Q(r) - 2 f_Q(r) \leq 3 f_Q(r) - 2 f_Q(r) = f_Q(r), \quad \forall\, r > 0.$$

Hence, by the decreasing fact of f_Q given by Lemma 4.3.2, (4.4.24) can be improved to

$$II \le \left(2(K+2)f_Q(0) + \frac{1}{100}\right)\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2}. \tag{4.4.25}$$

For the term III, following the argument in (4.4.16), we first compute

$$\int \left(\nabla\Delta^{K+2}Q\cdot\nabla\Delta^{-1}g\right)\Delta^{K+2}g\,dy = \int \left|\partial_r\Delta^{K+2}Q\right|\left(\frac{1}{r^2}\int_0^r |g(s)|s^2ds\right)\left|\Delta^{K+2}g\right|dy$$

$$\le \left(\int |\partial_r\Delta^{K+2}Q||y|^{-\frac{1}{2}}|\Delta^{K+2}g|dy\right)\|g\|_{L^2} \le \left\||y|^{-\frac{1}{2}}\partial_r\Delta^{K+2}Q\right\|_{L^2}\|g\|_{L^2}\|g\|_{\dot{H}^{2K+4}},$$

which, by Lemma 4.3.2, Gagliardo-Nirenberg inequality, and Young's inequality, yields that there exists $C = C(K,Q) > 0$ such that

$$III \le \int \left(\nabla\Delta^{K+2}Q\cdot\nabla\Delta^{-1}g\right)\Delta^{K+2}g\,dy + C\left(\sum_{j=1}^{2K+4}\|\nabla^{2K+5-j}Q\|_{L^\infty}\|\nabla^{j+1}\Delta^{-1}g\|_{L^2}\right)\|g\|_{\dot{H}^{2K+4}}$$

$$\le C\|g\|_{H^{2K+3}}\|g\|_{\dot{H}^{2K+4}} \le \frac{1}{100}\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2}. \tag{4.4.26}$$

Finally, for IV, by Lemma 4.3.3, Gagliardo-Nirenberg inequality and Young's inequality again,

$$IV \le 2(1-\mu)\int Q|\Delta^{K+2}g|^2dy + C\sum_{j=0}^{2K+3}\|\nabla^{2K+4-j}Q\|_{L^\infty}\|\nabla^j g\|_{L^2}\|\Delta^{K+2}g\|_{L^2}$$

$$\le \left(2(1-\mu)Q(0) + \frac{1}{100}\right)\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{H^{2K+3}}. \tag{4.4.27}$$

Consequently, combining (4.4.23), (4.4.25), (4.4.26), (4.4.27), along with the choice of $\beta < \frac{1}{2}$ from (4.4.1), Lemma 4.3.2, the selection of A from (4.4.20) and $K := \frac{A}{4}$, one finds that

$$(\mathcal{L}g,g)_{\dot{H}^{2K+4}} \le \left(-\frac{\beta}{2}(4K+5) + \frac{2K+4}{3(1-\mu)} + 1 + \frac{3}{100}\right)\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2}$$

$$= \left(-\frac{A+8}{4j_0} + \frac{3\beta}{2} + 1 + \frac{3}{100}\right)\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2} \le -\frac{1}{8}\|g\|^2_{\dot{H}^{2K+4}} + C\|g\|^2_{L^2},$$

and thus, we have completed the proof of (4.4.21). □

4.5 Nonlinear stability

In this section, building upon the coercivity properties of the operator $\mathcal{L}$ established in Section 4.4, we proceed to analyze the nonlinear stability of Q. Then, based on this nonlinear stability, we will conclude this section by constructing a finite-time blowup solution to (KS-D) for any $\mu \in \left[0, \frac{1}{3}\right)$ with finite mass and nonnegative density.

4.5.1 Ansatz and modulation

Recalling the ansatz $\Psi = Q+\varepsilon$, where ε is *radially symmetric* and solves the equation (4.4.6), we further decompose ε as motivated by Proposition 4.4.1:

$$\varepsilon(\tau, y) = \varepsilon_u(\tau, y) + \varepsilon_s(\tau, y), \text{ with } \varepsilon_u(\tau, y) = \sum_{j=0}^{K} c_j(\tau)\chi(y)|y|^{2j}, \tag{4.5.1}$$

where χ is a radial smooth cut-off function on $B(0,1)$ defined in (4.1.17), and $K = \frac{A}{4} \in \mathbb{Z}$ is as in Proposition 4.4.3. The coefficients $\{c_j\}_{j=0}^{K}$ are chosen such that the component ε_s satisfies the vanishing condition at the origin

$$\partial_r^{2j}\varepsilon_s(\tau, 0) = 0, \qquad \forall\ 0 \le j \le K. \tag{4.5.2}$$

In particular, we split ε_u as

$$\varepsilon_u(\tau) = \varepsilon_{u,real} + \varepsilon_{u,fake}\ , \tag{4.5.3}$$

where

$$\begin{cases} \varepsilon_{u,real}(\tau, y) := \sum_{j=0}^{j_0} c_j(\tau)\chi(y)|y|^{2j}, \\ \varepsilon_{u,fake}(\tau, y) := \sum_{j=j_0+1}^{K} c_j\chi(y)|y|^{2j}. \end{cases} \tag{4.5.4}$$

Remark 4.5.1. *Recalling that the singular weight introduced in Proposition 4.4.1 might not be sharp, the decomposition (4.5.1) allows for stable components to be included in ε_u. One can distinguish the unstable and stable parts as in (4.5.4). For further intuition regarding decomposition (4.5.4), we refer the readers to the modulation ODE system (4.5.9).*

4.5.1.1 Coefficients of the Taylor expansion of $\mathcal{L}\varepsilon_u$

With the vanishing condition (4.5.2), one can derive the modulation equations for $\{c_j\}_{j=0}^{K}$. A crucial step in this process is to compute the Taylor expansion at the origin of each term appearing in (4.4.6). As a preliminary, we focus in this section on determining the Taylor coefficients of $\mathcal{L}\varepsilon_u$ at the origin.

Specifically, recalling the definition of $\mathcal{L}$ from (4.4.7), and Taylor expansions of

(Q, f_Q) at the origin given in (4.3.5), we obtain for any $r \in (0, \epsilon]$,

$$\begin{aligned}\mathcal{L}\varepsilon_u &= -(\varepsilon_u + \beta r\partial_r\varepsilon_u) + f_Q\,(r\partial_r\varepsilon_u) + r\partial_r Q\left(\frac{1}{r^3}\int_0^r \varepsilon_u(s)s^2 ds\right) + 2(1-\mu)Q\varepsilon_u \\ &= -\sum_{j=0}^{K} c_j r^{2j} - \beta\sum_{j=0}^{K} 2jc_j r^{2j} + \left(\sum_{j=0}^{\infty}[f_Q]_j r^{2j}\right)\left(\sum_{j=0}^{K} 2jc_j r^{2j}\right) \\ &\quad + \left(\sum_{j=0}^{\infty} 2j[Q]_j r^{2j}\right)\left(\sum_{j=0}^{K}\frac{c_j}{2j+3}r^{2j}\right) + 2(1-\mu)\left(\sum_{j=0}^{\infty}[Q]_j r^{2j}\right)\left(\sum_{j=0}^{K} c_j r^{2j}\right).\end{aligned}$$

Applying (4.3.6), (4.3.7) and $[Q]_{j_0} = -1$, this yields that

$$[\mathcal{L}\varepsilon_u]_j = -c_j - 2\beta jc_j + 2j[f_Q]_0 c_j + 2(1-\mu)[Q]_0 c_j = \frac{j_0 - j}{j_0}c_j, \qquad \forall\, 0 \le j < j_0,$$

and

$$\begin{aligned}[\mathcal{L}\varepsilon]_{j_0} &= -c_{j_0} - 2\beta j_0 c_{j_0} + 2j_0[f_Q]_0 c_{j_0} + 2j_0[Q]_{j_0}\frac{c_0}{3} + 2(1-\mu)[Q]_0 c_{j_0} + 2(1-\mu)[Q]_{j_0}c_0 \\ &= \sigma_{0,j_0}c_0, \qquad \text{where} \qquad \sigma_{0,j_0} = -\left(\frac{2}{3}j_0 + 2(1-\mu)\right).\end{aligned}$$

Similarly, for $j_0 + 1 \le j \le K$, there exists $\{\sigma_{i,j}\} = \{\sigma_{i,j}(\mu, j_0)\} \subset \mathbb{R}$ such that

$$[\mathcal{L}\varepsilon]_j = \frac{j_0 - j}{j_0}c_j + \sum_{i=0}^{j-1}\sigma_{i,j}c_i, \qquad \forall\, j_0 + 1 \le j \le K.$$

In summary, $[\mathcal{L}\varepsilon_u]_j$, the $2j$-th order coefficient of the Taylor expansion of $\mathcal{L}\varepsilon_u$ at the origin, satisfies

$$[\mathcal{L}\varepsilon_u]_j = \begin{cases} \frac{j_0-j}{j_0}c_j, & 0 \le j < j_0, \\ \frac{j_0-j}{j_0}c_j + \sigma_{0,j_0}c_0, & j = j_0, \\ \frac{j_0-j}{j_0}c_j + \sum_{i=0}^{j-1}\sigma_{i,j}c_i, & j_0 + 1 \le j \le K. \end{cases} \tag{4.5.5}$$

4.5.1.2 Modulation equations.

Next, we derive the modulation equations for $\{c_j\}_{j=0}^K$ under the restriction given in (4.5.2), building on the previous preliminary results. Concretely, with the decomposition of ε in (4.5.1) and the equation (4.4.6),

$$\partial_\tau\varepsilon_s = \mathcal{L}\varepsilon_s + N(\varepsilon_s) + \nabla\varepsilon_u\cdot\nabla\Delta^{-1}\varepsilon_s + \nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u + 2(1-\mu)\varepsilon_u\varepsilon_s + G[\lambda, \Psi, \varepsilon_u], \tag{4.5.6}$$

where the *modulation term* $G[\lambda, \Psi, \varepsilon_u]$ is defined as

$$G[\lambda, \Psi, \varepsilon_u] := \lambda^{2-4\beta}\Delta\Psi - \sum_{j=0}^{K} \dot{c}_j \chi(r) r^{2j} + \mathcal{L}\varepsilon_u + N(\varepsilon_u). \tag{4.5.7}$$

Since $\varepsilon_s = O(r^{2K+2})$ at the origin as indicated in (4.5.2), and by the definition of $\mathcal{L}$ given in (4.4.7), $\mathcal{L}\varepsilon_s = O(r^{2K+2})$ at the origin, the following holds

$$\partial_\tau \varepsilon_s - \left(\mathcal{L}\varepsilon_s + N(\varepsilon_s) + \nabla\varepsilon_u \cdot \nabla\Delta^{-1}\varepsilon_s + \nabla\varepsilon_s \cdot \nabla\Delta^{-1}\varepsilon_u + 2(1-\mu)\varepsilon_u\varepsilon_s \right) = O(r^{2K+2}),$$

which requires the modulation term $G[\lambda, \Psi, \varepsilon_u] = O(r^{2K+2})$ at the origin. Noting that $G[\lambda, \Psi, \varepsilon_u]$ can be expanded as follows:

$$\begin{aligned} G[\lambda, \Psi, \varepsilon_u] &= \lambda^{2-4\beta}\Delta\Psi - \sum_{j=0}^{K} \dot{c}_j \chi r^{2j} + \mathcal{L}\varepsilon_u + N(\varepsilon_u) \\ &= \lambda^{2-4\beta} \sum_{j=0}^{K} [\Delta\Psi]_j \chi r^{2j} - \sum_{j=0}^{K} \dot{c}_j \chi r^{2j} + \sum_{j=0}^{K} [\mathcal{L}\varepsilon_u]_j \chi r^{2j} + \sum_{j=0}^{K} [N(\varepsilon_u)]_j \chi r^{2j} \\ &\quad + \lambda^{2-4\beta} \left((\Delta Q + \varepsilon_u) - \sum_{j=0}^{K} [\Delta Q + \Delta\varepsilon_u]_j \chi r^{2j} \right) + \lambda^{2-4\beta} \left(\Delta\varepsilon_s - [\Delta\varepsilon_s]_K \chi r^{2K} \right) \\ &\quad + \left(\mathcal{L}\varepsilon_u - \sum_{j=0}^{K} [\mathcal{L}\varepsilon_u]_j \chi r^{2j} \right) + N(\varepsilon_u) - \sum_{j=0}^{K} [N(\varepsilon_u)]_j \chi r^{2j}, \end{aligned} \tag{4.5.8}$$

where (4.5.8) has fewer orders of vanishing at the origin. Hence (4.5.8) should be imposed to vanish at the origin, then together with (4.5.5), this yields the following modulation equations onto the coefficients $\{c_j\}_{j=0}^{K}$:

$$\begin{cases} \dot{c}_j = \frac{j_0 - j}{j_0} c_j + \lambda^{2-4\beta}[\Delta\Psi]_j + [N(\varepsilon_u)]_j, & 0 \le j < j_0, \\ \dot{c}_{j_0} = \sigma_{0,j_0} c_0 + \lambda^{2-4\beta}[\Delta\Psi]_{j_0} + [N(\varepsilon_u)]_j, & j = j_0, \\ \dot{c}_j = \frac{j_0 - j}{j_0} c_j + \sum_{i=0}^{j-1} \sigma_{i,j} c_i + \lambda^{2-4\beta}[\Delta\Psi]_j + [N(\varepsilon_u)]_j, & j_0 < j \le K. \end{cases} \tag{4.5.9}$$

In conclusion, we obtain an ODE-PDE system in terms of $\left(\{c_j\}_{j=0}^{K}, \varepsilon_s\right)$, which couples equations (4.5.6) and (4.5.9). Moreover, the modulation term $G[\lambda, \Psi, \varepsilon_u]$ in (4.5.6) can be further expressed as follows:

$$\begin{aligned} G[\lambda, \Psi, \varepsilon_u] &= \lambda^{2-4\beta} \left((\Delta Q + \varepsilon_u) - \sum_{j=0}^{K} [\Delta Q + \Delta\varepsilon_u]_j \chi r^{2j} \right) + \lambda^{2-4\beta} \left(\Delta\varepsilon_s - [\Delta\varepsilon_s]_K \chi r^{2K} \right) \\ &\quad + \left(\mathcal{L}\varepsilon_u - \sum_{j=0}^{K} [\mathcal{L}\varepsilon_u]_j \chi r^{2j} \right) + N(\varepsilon_u) - \sum_{j=0}^{K} [N(\varepsilon_u)]_j \chi r^{2j}. \end{aligned} \tag{4.5.10}$$

4.5.2 Bootstrap argument.

We take $\delta_g \ll 1$ satisfying

$$\delta_g \ll \min\left\{\frac{1}{8}, 2-4\beta, \frac{1}{j_0}, \frac{1}{|\sigma_{0,j_0}|}, A^{-1}, B\right\}, \tag{4.5.11}$$

and introduce the function

$$d_{real}(\tau) := \sqrt{\frac{1}{2}\sum_{j=1}^{j_0-1} c_j^2 + \frac{10}{\delta_g^2}c_0^2 + \frac{1}{4|\sigma_{0,j_0}|}c_{j_0}^2}, \tag{4.5.12}$$

which describes the behavior of $\varepsilon_{u,real}$. The main result is as follows:

Proposition 4.5.2 (Bootstrap). *There exists $0 < \delta_g \ll 1$ satisfying (4.5.11) such that there exists $(\delta_0, \delta_1, \delta_2, \delta_3, \delta_{j_0+1}, \delta_{j_0+2}..., \delta_K)$ with*

$$\delta_0 \ll \delta_3 \ll \delta_{j_0+1} \ll \delta_{j_0+2} \ll \delta_{j_0+3} \ll ... \ll \delta_K \ll \delta_1 \ll \delta_2 \ll \delta_3^{\frac{1}{2}} \ll \delta_g \ll Q(2), \tag{4.5.13}$$

such that for any $\lambda(0) = \lambda_0$, $\varepsilon_s(0) \in L^2_w \cap \dot{H}^{2K+6}_{rad}(\mathbb{R}^3)$ and $\{c_j(0)\}_{j=j_0+1}^K$ [6] satisfying

$$|\lambda_0| + |\lambda_0|^{2-4\beta} + \|\varepsilon_s(0)\|_{L^2_w(\mathbb{R}^3)} + \|\varepsilon_s(0)\|_{\dot{H}^{2K+4}(\mathbb{R}^3)} + \sum_{j=j_0+1}^{K} |c_j(0)| \le \delta_0, \tag{4.5.14}$$

there exists $\{c_j(0)\}_{j=0}^{j_0}$ such that the radial solution to (4.4.4) with initial datum

$$(\Psi_0, \lambda_0) = \left(Q + \sum_{j=0}^{j_0} c_j(0)\chi|y|^{2j} + \sum_{j=j_0+1}^{K} c_j(0)\chi|y|^{2j} + \varepsilon_{s0}, \lambda_0\right),$$

globally exists and can be split into (4.5.1) satisfying the following for all $\tau \ge 0$:

- *(Control the L^2 base norm of stable part ε_s)*

$$\|\varepsilon_s(\tau)\|_{L^2_w} \le \delta_1 e^{-\frac{1}{2}\delta_g \tau}. \tag{4.5.15}$$

- *(Control the higher regularity of stable part ε_s via $\dot{H}^{2K+4}$ norm)*

$$\|\varepsilon_s(\tau)\|_{\dot{H}^{2K+4}} \le \delta_2 e^{-\frac{1}{2}\delta_g \tau}. \tag{4.5.16}$$

- *(Control the real unstable part $\varepsilon_{u,real}$)*

$$d_{real}(\tau) \le \delta_3 e^{-\frac{7}{10}\delta_g \tau}. \tag{4.5.17}$$

[6] With this regularity assumption on initial data, together with (4.4.2) and (4.4.3), by the standard fixed point arguments as in [336] or [274, Appendix A], the local wellposedness can be ensured and the related solution to (KS-D) satisfying $\rho \in C_t^0\left([0, T_{max}), H_x^{2K+6}(\mathbb{R}^3)\right) \cap C_t^1\left([0, T_{max}), H_x^{2K+4}(\mathbb{R}^3)\right)$.

• *(Control the fake unstable part $\varepsilon_{u,fake}$)*

$$|c_j(\tau)| \leq \delta_j e^{-\frac{7}{10}\delta_g \tau}, \quad \forall\ j_0+1 \leq j \leq K. \tag{4.5.18}$$

Remark 4.5.3. *In what follows, the estimates* (4.5.27), (4.5.32), (4.5.35), *and* (4.5.36) *yield the requirements on* $\{\delta_i\}$ *summarized in* (4.5.13). *Notably, the constants* $C > 0$ *appearing below, which might vary from line to line if necessary, are all independent of* $\{\delta_i\}$. *Therefore, we can choose appropriate* $\{\delta_i\}$ *in a manner to close Proposition 4.5.2.*

Remark 4.5.4. *Proposition 4.5.2 is the center of the chapter. As in [94, 325], Proposition 4.5.2 will be proven via contradiction using the topological argument as follows: given* $\big(\lambda_0, \varepsilon_{u,fake}(0), \varepsilon_s(0)\big)$ *satisfying* (4.5.14), *we assume that for any* $\varepsilon_{u,real}(\tau, y) = \sum_{j=0}^{j_0} c_j(\tau)\chi(y)|y|^{2j}$ *satisfying* (4.5.17) *when* $\tau = 0$, *the exit time*

$$\tau^* = \sup\ \Big\{\tau \geq 0 : \big(d_{real}, \{c_j\}_{j=j_0+1}^{N}, \varepsilon_s\big) \text{ satisfy (4.5.15)-(4.5.18) simultaneously on } [0,\tau]\Big\} \tag{4.5.19}$$

is finite. We then look for a contradiction under the assumptions of Proposition 4.5.2. Subsequently, we study the flow satisfying (4.5.15)-(4.5.18) *on* $[0, \tau^*]$. *Specifically, we will show that the bounds in* (4.5.15), (4.5.16), *and* (4.5.18) *can be further improved within the bootstrap regime. This implies that the only possible scenario for the solution to exit the bootstrap regime is for the real unstable condition* (4.5.17) *to fail as* $\tau > \tau^*$. *Moreover, using the outer-going flux property of* d_{real} *at the exit time* $\tau = \tau^*$, *we conclude a contradiction via Brouwer's fixed point theorem.*

4.5.3 A priori estimates

4.5.3.1 Preliminaries

In the section, as preliminaries, we are devoted to the estimate of $[\Delta\Psi]_j$ with $0 \leq j \leq K$, which is the coefficient of Taylor expansion of $\Delta\Psi$ with the $2j$-th order at the origin. And the main result is as follows:

Lemma 4.5.5. *Under the bootstrap assumption of Proposition 4.5.2, as for the function* Ψ *defined in* (4.4.3), *there exists a uniform constant* $C = C(K) > 0$ *such that the quantity* $[\Delta\Psi]_j := \frac{1}{(2j)!}\partial_r^{2j}\Delta\Psi(0)$ *satisfies*

$$\big|[\Delta\Psi]_j\big| \leq \begin{cases} C(K), & 0 \leq j < K, \\ C(K) + C(K)\|\varepsilon_s\|_{H^{2K+4}}, & j = K. \end{cases} \tag{4.5.20}$$

Proof. Recall the decomposition of Ψ defined in (4.5.1),

$$[\Delta\Psi]_j = \frac{1}{(2j)!}\partial_r^{2j}[\Delta\Psi](0) = \frac{1}{(2j)!}\partial_r^{2j}\Delta Q(0)+\frac{1}{(2j)!}\partial_r^{2j}\Delta\varepsilon_u(0)+\frac{1}{(2j)!}\partial_r^{2j}\Delta\varepsilon_s(0),$$

where, by bootstrap assumptions (4.5.17) and (4.5.18), there exists a constant $C = C(K) > 0$ such that

$$\left|\frac{1}{(2j)!}\partial_r^{2j}[\Delta Q](0)\right| + \left|\frac{1}{(2j)!}\partial_r^{2j}(\Delta\varepsilon_u)(0)\right| \le C(K), \quad \forall\, 0 \le j \le 2K.$$

For $\partial_r^{2j}\Delta\varepsilon_s(\tau,0)$, note that $\varepsilon_s = O(r^{2K+2})$ at the origin,

$$\partial_r^{2j}\Delta\varepsilon_s(0) = 0, \quad \forall\, 0 \le j \le K-1.$$

In addition, for $\partial_r^{2K}\Delta\varepsilon_s(\tau,0)$, by Sobolev inequality, there exists $C = C(K) > 0$ such that

$$\begin{aligned}|\partial_r^{2K}\Delta\varepsilon_s(\tau,0)| &\le C\left|[\varepsilon_s(\tau)]_{K+1}\right| \le C|\nabla^{2K+2}\varepsilon_s(\tau,0)| \\ &\le C\|\nabla^{2K+2}\varepsilon_s(\tau)\|_{L^\infty} \le C\|\varepsilon_s(\tau)\|_{H^{2K+4}}. \end{aligned} \tag{4.5.21}$$

Consequently, combining all of the arguments above immediately yields (4.5.20). □

Lemma 4.5.6 (Estimates of terms related to ε_u)**.** *Under the bootstrap assumptions of Proposition 4.5.2, for the function ε_u defined in (4.5.1), $N(\varepsilon_u)$ is compactly supported on $B(0,2)$ and satisfies*

$$\sum_{j=0}^{K}\left|[N(\varepsilon_u)(\tau)]_j\right| + \left\|N(\varepsilon_u)(\tau)\right\|_{H^{2K+4}} \le C\sum_{j=0}^{K}|c_j(\tau)|^2, \quad \forall\, \tau\in[0,\tau^*], \tag{4.5.22}$$

for some uniform constant $C = C(K) > 0$. In addition, $\mathcal{L}\varepsilon_u \in H^{2K+4}$ and satisfies

$$\sum_{j=0}^{K}\left|[\mathcal{L}\varepsilon_u(\tau)]_j\right| + \left\|\mathcal{L}\varepsilon_u(\tau)\right\|_{H^{2K+4}} \le C\sum_{j=0}^{K}|c_j(\tau)|, \quad \forall\, \tau\in[0,\tau^*], \tag{4.5.23}$$

for some uniform constant $C = C(K,Q) > 0$.

Proof. Recalling the definition of ε_u given in (4.5.1), and $N(\varepsilon_u) = \nabla\Delta^{-1}\varepsilon_u \cdot \nabla\varepsilon_u + (1-\mu)\varepsilon_u^2$, we check

$$\begin{aligned} N(\varepsilon_u) &= (r\partial_r\varepsilon_u)\frac{1}{r^3}\int_0^r \varepsilon_u(s)s^2 ds + (1-\mu)\varepsilon_u^2 \\ &= \sum_{j_1,j_2=0}^{K}\left(\frac{2j_1 c_{j_1}c_{j_2}}{2j_2+3} + (1-\mu)c_{j_1}c_{j_2}\right) r^{2j_1+2j_2} \\ &= \sum_{j=0}^{2K}\left(\sum_{j_1+j_2=j}\frac{2j_1 c_{j_1}c_{j_2}}{2j_2+3} + (1-\mu)c_{j_1}c_{j_2}\right) r^{2j}, \quad \forall\, r \le \frac{1}{8}. \end{aligned}$$

This yields

$$\left|[N(\varepsilon_u)]_i\right| = \left|\sum_{j_1+j_2=j}\frac{2j_1 c_{j_1}c_{j_2}}{2j_2+3} + (1-\mu)c_{j_1}c_{j_2}\right| \le C(K)\sum_{j=0}^{K}|c_j|^2, \quad \forall\, 0 \le i \le K.$$

In addition, by using Gagliardo-Nirenberg inequality and supp $\varepsilon_u \in B(0,2)$,

$$\|N(\varepsilon_u)\|_{H^{2K+4}} \le C\|\varepsilon_u\|^2_{H^{2K+5}} \le C\sum_{j=0}^{K}|c_j|^2,$$

which concludes (4.5.22). To verify (4.5.23), it is a direct result of (4.5.5) that there exists a constant $C = C(K,Q) > 0$ such that

$$\left|[\mathcal{L}\varepsilon_u]_i\right| \le C(K,Q)\sum_{j=0}^{K}|c_j|, \quad \forall\, 0 \le i \le K.$$

Additionally, recall (4.3.1), Lemma 4.3.2 and supp $\varepsilon_u \subset B(0,2)$, similar to the argument in (4.4.26) and (4.4.27), by using Gagliardo-Nirenberg inequality, there exists a constant $C = C(K,Q) > 0$ such that

$$\begin{aligned} \left\|\mathcal{L}\varepsilon_u\right\|_{H^{2K+4}} &\le \left\| -(\varepsilon_u + \beta y\cdot\nabla\varepsilon_u) + \nabla\varepsilon_u\cdot\nabla\Delta^{-1}Q + 2(1-\mu)Q\varepsilon_u\right\|_{H^{2K+4}(B(0,2))} \\ &\quad + \|\nabla\Delta^{-1}\varepsilon_u\cdot\nabla Q\|_{H^{2K+4}} \\ &\le C\|Q\|_{H^{2K+5}}\|\varepsilon_u\|_{H^{2K+5}} \le C\sum_{j=0}^{K}|c_j|. \end{aligned}$$

This completes the proof of (4.5.23). □

4.5.3.2 L^2_w estimate of ε_s.

This section is to improve the $L^2_w(\mathbb{R}^3)$ estimate of ε_s under the bootstrap assumptions of Proposition 4.5.2. And the main result is as follows:

Lemma 4.5.7 (L^2_w estimate of ε_s)**.** *Under the assumptions of Proposition 4.5.2, there exists a constant $C = C(Q, A, B, \delta_g) > 0$ such that*

$$\frac{d}{d\tau}\|\varepsilon_s\|^2_{L^2_w} \le -\delta_g\|\varepsilon_s\|^2_{L^2_w} + C\left(\lambda^{4-8\beta} + d^2_{real} + \sum_{j=j_0+1}^{K}|c_j|^2\right) \quad \forall\, \tau \in [0, \tau^*]. \tag{4.5.24}$$

In particular, the bootstrap assumption (4.5.15) can be improved to

$$\|\varepsilon_s(\tau)\|_{L^2_w} \le \frac{1}{2}\delta_1 e^{-\frac{\delta_g}{2}\tau}, \quad \forall\, \tau \in [0, \tau^*].$$

The proof of Lemma 4.5.7 is lengthy; thus, we separate the L^2_w estimate of the modulation term $G[\lambda, \Psi, \varepsilon_u]$ defined in (4.5.7). The estimate is as follows:

Lemma 4.5.8 (Estimate on modulation term $G[\lambda, \Psi, \varepsilon_u]$)**.** *Under the bootstrap assumption of Proposition 4.5.2, as for $G[\lambda, \Psi, \varepsilon_u]$ defined in (4.5.7), there exists a uniform constant $C = C(Q, A, B) > 0$ such that*

$$\|G[\lambda, \Psi, \varepsilon_u]\|_{L^2_w} \le C\lambda^{2-4\beta} + C\sum_{j=0}^{K}|c_j|. \tag{4.5.25}$$

Proof of Lemma 4.5.8. Recall (4.5.10), $G[\lambda, \Psi, \varepsilon_u] = O(r^{2K+2})$ at the origin, then by applying Taylor expansion with the form of integral residue at the origin onto $G[\lambda, \Psi, \varepsilon_u]$,

$$\begin{aligned}\big\|G[\lambda, \Psi, \varepsilon_u]\big\|^2_{L^2_w(|y|\le 1)} &\le C\int_{|y|\le 1}|G[\lambda, \Psi, \varepsilon_u]|^2\frac{1}{|y|^{4K}}dy \\ &\le C\big\|D^{2K}G[\lambda, \Psi, \varepsilon_u]\big\|^2_{L^\infty} \le C\|G[\lambda, \Psi, \varepsilon_u]\|^2_{H^{2K+2}}.\end{aligned}$$

For the case with $|y| \ge 1$,

$$\big\|G[\lambda, \Psi, \varepsilon_u]\big\|_{L^2_w(|y|\ge 1)} \le C\big\|G[\lambda, \Psi, \varepsilon_u]\big\|_{L^2(|y|\ge 1)} \le C\|G[\lambda, \Psi, \varepsilon_u]\|_{H^{2K+2}}.$$

Consequently, together with the definition of $G[\lambda, \Psi, \varepsilon_u]$ in (4.5.10), Lemma 4.5.5, (4.5.21), and Lemma 4.5.6, this concludes that

$$\begin{aligned}
&\|G[\lambda,\Psi,\varepsilon_u]\|_{L^2_w} \le C\|G[\lambda,\Psi,\varepsilon_u]\|_{H^{2K+2}} \\
&\le C\lambda^{2-4\beta}\left(\|\Delta Q+\varepsilon_u\|_{H^{2K+2}}+\|\Delta\varepsilon_s\|_{H^{2K+2}}+\sum_{j=0}^{K}\left|[\Delta Q+\varepsilon_u]_j\right|+[\Delta\varepsilon_s]_K\right) \\
&\quad + C\|\mathcal{L}\varepsilon_u\|_{H^{2K+2}}+\sum_{j=0}^{K}\left|[\mathcal{L}\varepsilon_u]_j\right|+C\|N(\varepsilon_u)\|_{H^{2K+2}}+C\sum_{j=0}^{K}\left|[N(\varepsilon_u)]_j\right| \\
&\le C\lambda^{2-4\beta}\left(1+\|\varepsilon_s\|_{H^{2K+4}}\right)+C\left(\sum_{j=0}^{K}|c_j|+|c_j|^2\right)\le C\lambda^{2-4\beta}+C\sum_{j=0}^{K}|c_j|.
\end{aligned}$$

for some uniform constant $C>0$, and thus we have concluded the proof. □

Next, we are devoted to the proof of Lemma 4.5.7.

Proof of Lemma 4.5.7. Recall that ε_s solves the equation (4.5.6),

$$\frac{1}{2}\frac{d}{d\tau}\|\varepsilon_s\|^2_{L^2_w}=(\partial_\tau\varepsilon_s,\varepsilon_s)_{L^2_w}=I+II+III+IV,$$

with

$$\begin{aligned}
&I:=(\mathcal{L}\varepsilon_s,\varepsilon_s)_{L^2_w}, \quad II:=\left(\nabla\varepsilon_u\cdot\nabla\Delta^{-1}\varepsilon_s+\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u+2(1-\mu)\varepsilon_u\varepsilon_s,\varepsilon_s\right)_{L^2_w}, \\
&III:=(N(\varepsilon_s),\varepsilon_s)_{L^2_w}, \quad IV:=(G[\lambda,\Psi,\varepsilon_u],\varepsilon_s)_{L^2_w}.
\end{aligned}$$

Step 1. Estimate of the small linear term II. The small linear term II can be written as $II:=II_1+II_2+II_3$ with

$$II_1:=\left(\nabla\varepsilon_u\cdot\nabla\Delta^{-1}\varepsilon_s,\varepsilon_s\right)_{L^2_w}, \quad II_2:=\left(\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u,\varepsilon_s\right)_{L^2_w}, \quad II_3:=(2(1-\mu)\varepsilon_u\varepsilon_s,\varepsilon_s)_{L^2_w}.$$

First of all, repeating (4.4.17), together with the fact that $\operatorname{supp}\varepsilon_u\subset B(0,2)$, there exists $C=C(A)>0$ such that II_1 can be controlled by

$$II_1\le C\left\||y|^{-\frac{1}{2}}\partial_r\varepsilon_u\right\|_{L^2}\|\varepsilon_s\|^2_{L^2_w}\le C\|\nabla\varepsilon_u\|_{L^\infty}\|\varepsilon_s\|^2_{L^2_w}.$$

For II_2, direct computation shows

$$\begin{aligned}
II_2&=\int\left(\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u\right)\varepsilon_s w dy=\frac{1}{2}\int\nabla\varepsilon_s^2\cdot\nabla\Delta^{-1}\varepsilon_u w dy \\
&=-\frac{1}{2}\int\varepsilon_s^2\varepsilon_u w dy-\frac{1}{2}\int\varepsilon_s^2\nabla\Delta^{-1}\varepsilon_u\cdot\nabla w dy\le C\|\varepsilon_u\|_{L^\infty}\|\varepsilon_s\|^2_{L^2_w}, \qquad (4.5.26)
\end{aligned}$$

where the last inequality holds by Cauchy-Schwarz inequality together with the pointwise estimate

$$|\nabla\Delta^{-1}\varepsilon_u\cdot\nabla w| = \left|\left(\frac{1}{r^2}\int_0^r \varepsilon_u(s)s^2ds\right)(\partial_r w)\right| \le C(A)\|\varepsilon_u\|_{L^\infty} w.$$

For II_3, it can be directly controlled by

$$II_3 \le \left|(2(1-\mu)\varepsilon_u\varepsilon_s,\varepsilon_s)_{L^2_w}\right| \le 2(1-\mu)\|\varepsilon_u\|_{L^\infty}\|\varepsilon_s\|^2_{L^2_w}.$$

In sum, the small linear term II can be estimated by

$$II \le C\|\varepsilon_u\|_{W^{1,\infty}}\|\varepsilon_s\|^2_{L^2_w}.$$

Step 2. Estimate of the nonlinear term III. We begin by performing integration by parts and using a similar argument in (4.5.26), it then follows that there exists $C = C(A) > 0$ such that

$$III = \left(\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_s + (1-\mu)\varepsilon_s^2,\varepsilon_s\right)_{L^2_w} \le C\|\varepsilon_s\|_{L^\infty}\|\varepsilon_s\|^2_{L^2_w}.$$

Step 3. Summary of the estimates. Combining all of the estimates above, (4.4.10), (4.5.11), (4.5.13) and (4.5.25), and applying Young's inequality, we obtain that there exists a constant $C = C(Q, A, B, \delta_g) > 0$ such that

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}\|\varepsilon_s\|^2_{L^2_w} &\le -\frac{1}{8}\|\varepsilon_s\|^2_{L^2_w} + C\left(\|\varepsilon_u\|_{W^{1,\infty}} + \|\varepsilon_s\|_{L^\infty}\right)\|\varepsilon_s\|^2_{L^2_w} + C\left(\lambda^{2-4\beta} + \sum_{j=0}^K |c_j|\right)\|\varepsilon_s\|_{L^2_w}\\ &\le -\frac{\delta_g}{2}\|\varepsilon_s\|^2_{L^2_w} + C\left(\lambda^{4-8\beta} + \sum_{j=0}^K |c_j|^2\right),\end{aligned}$$

which concludes (4.5.24). Hence, recalling the bootstrap assumptions (4.5.15), (4.5.16), (4.5.17) and (4.5.18), and by Gronwall's inequality,

$$\begin{aligned}\|\varepsilon_s(\tau)\|^2_{L^2_w} &\le e^{-\delta_g\tau}\|\varepsilon_s(0)\|^2_{L^2_w} + C\int_0^\tau e^{-\delta_g(\tau-s)}\left(\lambda^{4-8\beta}(s) + d^2_{real}(s) + \sum_{j=j_0+1}^K |c_j(s)|^2\right)ds\\ &\le e^{-\delta_g\tau}\|\varepsilon_s(0)\|^2_{L^2_w} + C\int_0^\tau e^{-\delta_g(\tau-s)}\left(\lambda_0^{4-8\beta}e^{-(2-4\beta)s} + \left(\delta_3^2 + \sum_{j=j_0+1}^K \delta_j^2\right)e^{-\frac{7}{5}\delta_g s}\right)ds\\ &\le \left[\delta_0^2 + C\left(\lambda_0^{4-8\beta} + \delta_3^2 + \sum_{j=j_0+1}^K \delta_j^2\right)\right]e^{-\delta_g\tau}, \quad \forall\,\tau\in(0,\tau^*],\end{aligned} \tag{4.5.27}$$

for some constant $C = C(Q, A, B, \delta_g) > 0$ independent on each δ_i. Ttogether with (4.5.13) and (4.5.14), this yields that

$$\|\varepsilon_s(\tau)\|_{L^2_w} \leq \frac{1}{2}\delta_1 e^{-\frac{1}{2}\delta_g \tau}, \quad \forall\, \tau \in [0, \tau^*],$$

hence we have concluded the proof. □

4.5.3.3 $\dot{H}^{2K+4}$ estimate of ε_s.

This section is to improve the $\dot{H}^{2K+4}(\mathbb{R}^3)$ estimate of ε_s under the bootstrap assumptions of Proposition 4.5.2. The main result is as follows:

Lemma 4.5.9 ($\dot{H}^{2K+4}$ estimate of ε_s)**.** *Under the assumptions of Proposition 4.5.2, there exists a constant $C = C(Q, A, B, \delta_g) > 0$ such that the following inequality hold uniformly in $\tau \in [0, \tau^*]$:*

$$\frac{d}{d\tau}\|\varepsilon_s\|^2_{\dot{H}^{2K+4}} \leq -\frac{1}{16}\|\varepsilon_s\|^2_{\dot{H}^{2K+4}} + C\|\varepsilon_s\|^2_{L^2} + C\left(\|\varepsilon_s\|^4_{H^{2K+4}} + \lambda^{4-8\beta} + \sum_{j=0}^{K}|c_j|^2\right). \tag{4.5.28}$$

In particular, the bootstrap assumption (4.5.16) can be improved to

$$\|\varepsilon_s(\tau)\|_{\dot{H}^{2K+4}} \leq \frac{1}{2}\delta_2 e^{-\frac{\delta_g}{2}\tau}, \quad \forall\, \tau \in [0, \tau^*].$$

Proof. Recall that ε_s solves the equation (4.5.6),

$$\frac{1}{2}\frac{d}{d\tau}\|\varepsilon_s\|^2_{\dot{H}^{2K+4}} = I + II + III + IV,$$

with

$$\begin{aligned}
I &:= \left(\Delta^{K+2}\mathcal{L}\varepsilon_s, \Delta^{K+2}\varepsilon_s\right)_{L^2},\\
II &:= \left(\Delta^{K+2}\left(\nabla\varepsilon_u \cdot \nabla\Delta^{-1}\varepsilon_s + \nabla\varepsilon_s \cdot \nabla\Delta^{-1}\varepsilon_u + 2(1-\mu)\varepsilon_u\varepsilon_s\right), \Delta^{K+2}\varepsilon_s\right)_{L^2},\\
III &:= \left(\Delta^{K+2}N(\varepsilon_s), \Delta^{K+2}\varepsilon_s\right)_{L^2}, \quad IV := \left(\Delta^{K+2}G[\lambda, \Psi, \varepsilon_u], \Delta^{K+2}\varepsilon_s\right)_{L^2}.
\end{aligned}$$

Estimate of small linear term II. For the small linear term II, since $\varepsilon_u \in C_0^\infty(\mathbb{R}^3)$ with support in $B(0, 2)$, together with integration by parts, Sobolev inequality, Gagliardo-Nirenberg inequality and (4.3.1), there exists uniform constant $C =$

$C(K,\mu)>0$ such that

$$\begin{aligned}
II &\le \left(\Delta^{K+2}\left(\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u\right),\Delta^{K+2}\varepsilon_s\right)_{L^2}+\left(\Delta^{K+2}\left(\nabla\varepsilon_u\cdot\nabla\Delta^{-1}\varepsilon_s+2(1-\mu)\varepsilon_u\varepsilon_s\right),\Delta^{K+2}\varepsilon_s\right)_{L^2}\\
&\le \left(\nabla\Delta^{K+2}\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_u,\Delta^{K+2}\varepsilon_s\right)_{L^2}+C\sum_{j=1}^{2K+4}\|\nabla^{2K+5-j}\varepsilon_s\|_{L^2}\|\nabla^{j+1}\Delta^{-1}\varepsilon_u\|_{L^\infty}\|\varepsilon_s\|_{\dot H^{2K+4}}\\
&\quad + C\|\nabla^{2K+5}\varepsilon_u\|_{L^2}\|\nabla\Delta^{-1}\varepsilon_s\|_{L^\infty(B(0,2))}+C\sum_{j=1}^{2K+4}\|\nabla^{2K+5-j}\varepsilon_u\|_{L^\infty}\|\nabla^{j+1}\Delta^{-1}\varepsilon_s\|_{L^2}\\
&\le C\|\varepsilon_u\|_{H^{2K+6}}\|\varepsilon_s\|^2_{H^{2K+4}}. \qquad (4.5.29)
\end{aligned}$$

Estimate of nonlinear term III. For III, by integration by parts, there exists a uniform constant $C=C(K,\mu)>0$ such that

$$\begin{aligned}
III &= \left(\Delta^{K+2}\left(\nabla\varepsilon_s\cdot\nabla\Delta^{-1}\varepsilon_s\right),\Delta^{K+2}\varepsilon_s\right)_{L^2}+(1-\mu)\left(\Delta^{K+2}(\varepsilon_s^2),\Delta^{K+2}\varepsilon_s\right)_{L^2}\\
&\le C\left(\|\varepsilon_s\|_{L^\infty}+\|\nabla^2\Delta^{-1}\varepsilon_s\|_{L^\infty}\right)\|\varepsilon_s\|^2_{H^{2K+4}}+C\sum_{j=2}^{2K+4}\|\varepsilon_s\|_{H^{2K+6-j}}\|\varepsilon_s\|_{H^j}\|\varepsilon_s\|_{\dot H^{2K+4}}\\
&\quad + C\sum_{j=1}^{2K+3}\|\varepsilon_s\|_{H^{2K+5-j}}\|\varepsilon_s\|_{H^{j+1}}\|\varepsilon_s\|_{\dot H^{2K+4}}\\
&\le C\|\varepsilon_s\|^3_{H^{2K+4}}. \qquad (4.5.30)
\end{aligned}$$

Estimate of the modulation term IV. For the modulation term IV, recalling (4.5.10), (4.5.21), (4.5.22) and (4.5.23), there exists a uniform constant $C>0$ such that

$$\left\|G[\lambda,\Psi,\varepsilon_u]-\lambda^{2-4\beta}\Delta\varepsilon_s\right\|_{\dot H^{2K+4}}\le C\lambda^{2-4\beta}+C\sum_{j=0}^{K}|c_j|,$$

which, by the non-negativity of λ (see (4.4.5)), yields that

$$\begin{aligned}
IV &= \left(\Delta^{K+2}G[\lambda,\Psi,\varepsilon_u],\Delta^{K+2}\varepsilon_s\right)_{L^2}\\
&= \lambda^{2-4\beta}\left(\Delta^{K+2}\Delta\varepsilon_s,\Delta^{K+2}\varepsilon_s\right)_{L^2}+\left(\Delta^{K+2}\left(G[\lambda,\Psi,\varepsilon_u]-\lambda^{2-4\beta}\Delta\varepsilon_s\right),\Delta^{K+2}\varepsilon_s\right)_{L^2}\\
&= -\lambda^{2-4\beta}\|\nabla\Delta^{K+2}\varepsilon_s\|_{L^2}+\left(\Delta^{K+2}\left(G[\lambda,\Psi,\varepsilon_u]-\lambda^{2-4\beta}\Delta\varepsilon_s\right),\Delta^{K+2}\varepsilon_s\right)_{L^2}\\
&\le C\left(\lambda^{2-4\beta}+\sum_{j=0}^{K}|c_j|\right)\|\varepsilon_s\|_{\dot H^{2K+4}}. \qquad (4.5.31)
\end{aligned}$$

Conclusion of the estimates. Noting that I has been handled in Proposition 4.4.3, combining (4.4.21), (4.5.30), (4.5.29) and (4.5.31) concludes that

$$\begin{aligned}\frac{d}{d\tau}\frac{1}{2}\|\varepsilon_s\|^2_{\dot{H}^{2K+4}} \leq -\frac{1}{8}\|\varepsilon_s\|^2_{\dot{H}^{2K+4}} + C\|\varepsilon_s\|^2_{L^2}\\ + C\left(\|\varepsilon_s\|^2_{H^{2K+4}} + \lambda^{2-4\beta} + \sum_{j=0}^{K}|c_j|\right)\|\varepsilon_s\|_{H^{2K+4}},\end{aligned}$$

which, together with Young's inequality, yields (4.5.28). Hence, recalling the bootstrap assumptions (4.5.15), (4.5.16), (4.5.17) and (4.5.18), by Gronwall's inequality,

$$\begin{aligned}&\|\varepsilon_s(\tau)\|^2_{\dot{H}^{2K+4}}\\ &\leq \|\varepsilon_s(0)\|^2_{\dot{H}^{2K+4}}e^{-\frac{1}{16}\tau} + C\left(\delta_1^2+\delta_2^4+\delta_3^2+\delta_0^{4-8\beta}+\sum_{j=0}^{K}\delta_j^2\right)\int_0^\tau e^{-\frac{1}{16}(\tau-s)}e^{-\delta_g s}ds\\ &\leq C\left(\delta_0^2+\delta_1^2+\delta_2^4+\delta_3^2+\delta_0^{4-8\beta}+\sum_{j=0}^{K}\delta_j^2\right)e^{-\delta_g\tau}, \quad \forall\,\tau\in[0,\tau^*]. \qquad (4.5.32)\end{aligned}$$

Combining the assumptions on the coefficients (4.5.11) and (4.5.13), this yields that

$$\|\varepsilon_s(\tau)\|_{\dot{H}^{2K+4}} \leq \frac{1}{2}\delta_2 e^{-\frac{1}{2}\delta_g\tau}, \qquad \forall\,\tau\in[0,\tau^*],$$

and hence we have concluded the result. □

4.5.3.4 Estimate of $\varepsilon_{u,fake}$.

This section is dedicated to improving the bootstrap assumption (4.5.18) under the bootstrap assumptions of Proposition 4.5.2. The main result is stated as follows:

Lemma 4.5.10 (Estimate of $\varepsilon_{u,fake}$)**.** *Under the assumptions of Proposition 4.5.2, for any $j_0+1\leq j\leq K$, there exists a constant $C=C(\mu,K,j_0,\beta)>0$ such that*

$$\frac{d}{d\tau}c_j^2 \leq -\frac{1}{j_0}c_j^2 + C\lambda^{8-4\beta} + C\sum_{i=0}^{j-1}|c_i|^2 + C\sum_{i=0}^{K}|c_i|^4, \quad \forall\,\tau\in[0,\tau^*]. \qquad (4.5.33)$$

In particular, the bootstrap assumption (4.5.18) can be improved to

$$|c_j(\tau)| \leq \frac{1}{2}\delta_j e^{-\frac{7\delta_g}{10}\tau}, \quad \forall\,\tau\in[0,\tau^*]. \qquad (4.5.34)$$

Proof. Recalling (4.5.9), c_j solves the ODE

$$\dot{c}_j = \frac{j_0-j}{j_0}c_j + \lambda^{2-4\beta}[\Delta\Psi]_j + \sum_{i=0}^{j-1}\sigma_{i,j}c_i + [N(\varepsilon_u)]_j, \qquad \text{for any } j_0<j\leq K.$$

Then by Young's inequality, (4.5.20) and (4.5.22), there exists $C = C(K, j_0, \mu) > 0$ such that

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}c_j^2 = c_j\dot{c}_j &= c_j\left(\frac{j_0 - j}{j_0}c_j + \lambda^{2-4\beta}[\Delta\Psi]_j + \sum_{i=0}^{j-1}\sigma_{i,j}c_i + [N(\varepsilon_u)]_j\right)\\ &= \frac{j_0 - j}{j_0}c_j^2 + \lambda^{2-4\beta}[\Delta\Psi]_j c_j + \sum_{i=0}^{j-1}\sigma_{i,j}c_i c_j + [N(\varepsilon_u)]_j c_j\\ &\le -\frac{1}{2j_0}c_j^2 + C\lambda^{8-4\beta} + C\sum_{i=0}^{j-1}|c_i|^2 + C\sum_{i=0}^{K}|c_i|^4.\end{aligned}$$

By Gronwall's inequality, (4.5.11) (4.5.13), (4.5.14), and bootstrap assumption (4.5.17) and (4.5.18), there exists a constant $C = C(K, j_0, \beta) > 0$ such that

$$\begin{aligned}c_j^2(\tau) &\le e^{-\frac{1}{j_0}\tau}c_j^2(0) + C\int_0^\tau e^{-\frac{1}{j_0}(\tau-s)}\left[\lambda_0^{4-8\beta}e^{-(2-4\beta)s} + \left(\delta_3^2 + \sum_{i=j_0+1}^{j-1}\delta_i^2\right)e^{-\frac{7}{5}\delta_g s}\right.\\ &\qquad\left. + \left(\delta_3^4 + \sum_{i=j_0+1}^{K}\delta_i^4\right)e^{-\frac{14}{5}\delta_g s}\right]ds\\ &\le e^{-\frac{1}{j_0}\tau}c_j^2(0) + C\left(\lambda_0^{4-8\beta} + \delta_3^2 + \sum_{j_0+1\le i\le j-1}\delta_i^2 + \delta_3^4 + \sum_{i=j_0+1}^{K}\delta_i^4\right)e^{-\frac{7}{5}\delta_g\tau}\\ &\le C\left(\delta_0^2 + \lambda_0^{4-8\beta} + \delta_3^2 + \sum_{i=j_0+1}^{j-1}\delta_i^2\right)e^{-\frac{7}{5}\delta_g\tau} \le \frac{1}{4}\delta_j^2 e^{-\frac{7}{5}\delta_g\tau},\quad \forall\,\tau\in[0,\tau^*],\end{aligned} \tag{4.5.35}$$

which concludes the proof. □

4.5.4 Control of real unstable part.

Proof of Proposition 4.5.2.

Improvement of the bootstrap assumptions. Arguing by contradiction, suppose that there exists an initial triple $(\lambda_0, \varepsilon_{u,fake}(0), \varepsilon_s(0))$ satisfying (4.5.14) such that for any $\varepsilon_{u,fake}(0)$ with $d_{real}(0) \le \delta_3$, the exit time τ^* defined in (4.5.19) is always finite. Then, by Lemma 4.5.7, Lemma 4.5.9 and Lemma 4.5.10, combining with the continuity of the solution to system (4.4.4) with respect to τ, there exists $0 < \epsilon_{\tau^*} \ll 1$ such that (4.5.15), (4.5.16) and (4.5.18) hold for all $\tau \in [0, \tau^* + \epsilon_{\tau^*}]$.

Outer-going flux property. In light of the preceding argument, the only possible mechanism by which the solution exits the bootstrap regime is the failure of the bootstrap assumption for the real unstable component $\varepsilon_{u,real}$, specifically when this assumption ceases to hold on $[0,\tau]$ for some $\tau > \tau^*$. Moreover, by the local well-posedness of the system (KS-D), it follows that the mapping $\tau \mapsto d_{real}(\tau)$ is continuous. If we define the set

$$B(\tau) := \left\{ \{c_j\}_{j=0}^{j_0} \in \mathbb{R}^{j_0+1} : \sqrt{\frac{1}{2}\sum_{j=1}^{j_0-1} c_j^2 + \frac{10}{\delta_g^2}c_0^2 + \frac{1}{4|\sigma_{0,j_0}|}c_{j_0}^2} \leq \delta_3 e^{-\frac{7}{10}\delta_g \tau} \right\},$$

then by the contradiction assumption, the following holds

$$\begin{cases} \{c_j(\tau)\}_{j=0}^{j_0} \in B(\tau), & \forall\, \tau \in [0,\tau^*], \\ \{c_j(\tau)\}_{j=0}^{j_0} \in \partial B(\tau), & \tau = \tau^*. \end{cases}$$

Recalling the modulation ODE system satisfied by the coefficients $\{c_j\}_{j=0}^{j_0}$ as given in (4.5.9), and invoking Young's inequality in conjunction with Lemma 4.5.5 and Lemma 4.5.6, there exists a constant $C = C(\delta_g, |\sigma_{0,j_0}|, K, j_0) > 0$ independent of $\{\delta_i\}$, such that the quantity d_{real} defined in (4.5.12) satisfies

$$\begin{aligned}
&\frac{d}{d\tau} d_{real}^2(\tau) = \sum_{j=1}^{j_0-1} c_j \dot{c}_j + \frac{20}{\delta_g^2} c_0 \dot{c}_0 + \frac{1}{2|\sigma_{0,j_0}|} c_{j_0} \dot{c}_{j_0} \\
&= \sum_{j=1}^{j_0-1} \frac{j_0 - j}{j_0} c_j^2 + \frac{20}{\delta_g^2} c_0^2 + \frac{1}{2} c_0 c_{j_0} + \lambda^{2-4\beta} \left(\sum_{j=1}^{j_0-1} c_j [\Delta\Psi]_j + \frac{20}{\delta_g^2} c_0 [\Delta\Psi]_0 + \frac{1}{2|\sigma_{0,j_0}|} c_{j_0} [\Delta\Psi]_{j_0} \right) \\
&\quad + \sum_{j=1}^{j_0-1} [N(\varepsilon_u)]_j c_j + \frac{20}{\delta_g^2} c_0 [N(\varepsilon_u)]_0 + \frac{1}{2|\sigma_{0,j_0}|} c_{j_0} [N(\varepsilon_u)]_{j_0} \\
&\geq -C\lambda^{4-8\beta} \sum_{j=0}^{j_0} \left| [\Delta\Psi]_j \right|^2 - \frac{\delta_g^2}{160} c_{j_0}^2 - C \sum_{j=0}^{K} \left| [N(\varepsilon_u)]_j \right|^2 \\
&\geq -C\lambda^{4-8\beta} - \frac{\delta_g^2 |\sigma_{0,j_0}|}{40} d_{real}^2 - C d_{real}^4 - C \sum_{j=j_0+1}^{K} |c_j|^4.
\end{aligned}$$

Consequently, under the contradiction assumption that $e^{\frac{7}{10}\delta_g \tau^*} d_{real}(\tau^*) = \delta_3$, and in light of bootstrap assumptions (4.5.14), (4.5.17), and (4.5.18), together with coefficient restrictions (4.5.11) and (4.5.13), we deduce the outer-going flux property

on $\partial B(\tau^*)$:

$$\begin{aligned}
&\frac{d}{d\tau}\left(e^{\frac{7}{5}\delta_g\tau}d_{real}^2(\tau)\right)\Big|_{\tau=\tau^*} = \left(\frac{7}{5}\delta_g e^{\frac{7}{5}\delta_g\tau}d_{real}^2(\tau) + e^{\frac{7}{5}\delta_g\tau}\frac{d}{d\tau}d_{real}^2(\tau)\right)\Big|_{\tau=\tau^*}\\
&\geq \frac{7}{5}\delta_g e^{\frac{7}{5}\delta_g\tau^*}d_{real}^2(\tau^*) - C\lambda_0^{4-8\beta}e^{\frac{7}{5}\delta_g\tau^*}e^{-(2-4\beta)\tau^*} - \frac{\delta_g^2|\sigma_{0,j_0}|}{40}e^{\frac{7}{5}\delta_g\tau^*}d_{real}^2(\tau^*)\\
&\qquad - Ce^{\frac{7}{5}\delta_g\tau^*}d_{real}^4(\tau^*) - C\sum_{j=j_0+1}^{K}e^{\frac{7}{5}\delta_g\tau^*}|c_j(\tau^*)|^4\\
&\geq \frac{7}{5}\delta_g\delta_3^2 - C\lambda_0^{4-8\beta} - \frac{\delta_g^2|\sigma_{0,j_0}|}{40}\delta_3^2 - C\delta_3^4 - C\sum_{j=j_0+1}^{K}\delta_j^4 \geq \delta_g\delta_3^2 > 0. \qquad (4.5.36)
\end{aligned}$$

Brouwer's topological argument. Firstly, for any $\delta > 0$, we define the following convex subset on $\mathbb{R}^{j_0+1}$:

$$\mathcal{B}_{real}(0,\delta) := \left\{\{c_j\}_{j=0}^{j_0} \in \mathbb{R}^{j_0+1} : \sqrt{\frac{1}{2}\sum_{j=1}^{j_0-1}c_j^2 + \frac{10}{\delta_g^2}c_0^2 + \frac{1}{4|\sigma_{0,j_0}|}c_{j_0}^2} < \delta\right\}.$$

By the local well-posedness theory for (KS-D) and contradiction assumption (4.5.19), we establish the continuity of the map $\varepsilon_{u,real}(0) \mapsto \varepsilon_{u,real}(\tau^*(\varepsilon_{u,real}(0)))$, where $\tau^*(\varepsilon_{u,real}(0))$ denotes the exit time associated with $\big(\lambda_0, \varepsilon_{u,real}(0), \varepsilon_{u,fake}(0), \varepsilon_s(0)\big)$ satisfying the initial condition (4.5.14). Accordingly, we define a mapping $\Phi : \overline{\mathcal{B}_{real}(0,1)} \mapsto \partial\mathcal{B}_{real}(0,1)$ as follows:

$$\begin{aligned}
\{d_j\}_{j=0}^{j_0} \in \overline{\mathcal{B}_{real}(0,1)} &\mapsto \{c_j(0)\}_{j=0}^{j_0} := \{\delta_3 d_j\}_{j=0}^{j_0} \in \overline{\mathcal{B}_{\mathbb{R}^{j_0+1}}(0,\delta_3)}\\
&\mapsto \varepsilon_{u,real}(0,y) = \sum_{j=0}^{j_0}c_j(0)\chi(y)|y|^{2j}\\
&\mapsto \varepsilon_{u,real}(\tau^*,y) = \sum_{j=0}^{j_0}c_j(\tau^*)\chi(y)|y|^{2j}\\
&\mapsto \{d_j(\tau^*)\}_{j=0}^{j_0} := \left\{\frac{1}{\delta_3}e^{\frac{7}{10}\delta_g\tau^*}c_j(\tau^*)\right\}_{j=0}^{j_0} \in \partial\mathcal{B}_{real}(0,1),
\end{aligned}$$

whose continuity is ensured by the local well-posedness of (KS-D). In particular, when $\{d_j\}_{j=0}^{j_0} \in \partial\mathcal{B}_{real}(0,1)$, by the outer-going flux property of the flow as indicated in (4.5.36), $\tau = 0$ is exactly the exit time, and it leads to $\Phi = Id$ on the boundary $\partial\mathcal{B}_{real}(0,1)$. However, with the continuity of Φ on the nonempty compact convex set $\overline{\mathcal{B}_{real}(0,1)}$, we conclude that $-\Phi$ has a fixed point on the boundary $\partial\mathcal{B}_{real}(0,1)$, which contradicts to the assertion that $\Phi = Id$ on the boundary. Consequently, the proof of Proposition 4.5.2 has been completed. □

4.5.5 Existence of finite blowup solution to (KS-D) with finite mass.

This section is devoted to the study of the existence of the finite-time blowup solution to (KS-D) for any fixed $0 \le \mu < \frac{1}{3}$ and concludes the chapter.

Proof of Theorem 4.1.1. Firstly, for any fixed $\mu \in \left[0, \frac{1}{3}\right)$, we choose Q constructed in Lemma 4.3.2 and fix the weight w as specified in Proposition 4.4.1. Then as indicated in Proposition 4.5.2, there exist $\delta_g \ll 1$ and $\{\delta_i\}$ satisfying (4.5.11) and (4.5.13), such that the result shown in Proposition 4.5.2 holds.

Construction of nonnegative initial data with finite mass. Firstly, since $Q \in H^{2K+4}$ as guaranteed by Lemma 4.3.2, we can choose $\lambda_0 \ll \delta_0$ sufficiently small and $R_2 \gg 1$ sufficiently large such that the following estimates hold:

$$\|(\chi_{R_2} - 1)Q\|_{L^2_w} + \|(\chi_{R_2} - 1)Q\|_{\dot{H}^{2K+4}} \ll \delta_0, \tag{4.5.37}$$

where χ_{R_2} is a cut-off function on $B(0, R_2)$ defined by (4.1.18). It then follows that there exists $\{c_j(0)\}_{j=0}^{j_0}$ such that Proposition 4.5.2 holds with the initial datum

$$\big(\lambda_0, \varepsilon_s(0,y), \varepsilon_{u,real}(0,y), \varepsilon_{u,fake}(0,y)\big) := \left(\lambda_0, (\chi_{R_2} - 1)Q, \sum_{j=0}^{j_0} c_j(0)\chi(y)|y|^{2j}, 0\right). \tag{4.5.38}$$

In other words, there exists $\varepsilon_{u,real}(0,y)$ such that the solution to (4.4.4) with initial data

$$\Psi_0 = Q + \varepsilon_s(0) + \varepsilon_{u,real}(0) + \varepsilon_{u,fake}(0) = \chi_{R_2}Q + \sum_{j=0}^{j_0} c_j(0)\chi(y)|y|^{2j} \subset C_0^\infty(\mathbb{R}^3),$$

globally exists in τ and satisfies (4.5.15)-(4.5.18) on $[0, +\infty)$, and the corresponding solution $\Psi(\tau)$ can be decomposed into $\Psi(\tau) = Q + \varepsilon(\tau)$ with

$$\begin{aligned}\|\varepsilon(\tau)\|_{H^{2K+4}} &\le \|\varepsilon_u(\tau)\|_{H^{2K+4}} + \|\varepsilon_s(\tau)\|_{H^{2K+4}} \\ &\le C\left(\delta_1 + \delta_2 + \delta_3 + \sum_{j=j_0+1}^{K} \delta_j\right) e^{-\frac{1}{2}\delta_g \tau}, \quad \forall\, \tau > 0,\end{aligned}$$

for some universal $C > 0$. In addition, in view of the fact that Q is radially decreasing, together with the condition (4.5.13), we obtain that

$$\begin{cases} \Psi_0(y) \ge Q(2) - C\sum_{j=0}^{j_0} |c_j(0)| \ge \frac{1}{2}Q(2) > 0, & \forall\, |y| \le 2, \\ \Psi_0(y) = \chi_{R_2}(y)Q(y) \ge 0, & \forall\, |y| \ge 2, \end{cases}$$

which yields the non-negativity of Ψ_0. Now, returning to the original equation (KS-D), and recalling (4.4.2), (4.4.3) and (4.4.5), the nonnegativity of initial density ρ_0 can be guaranteed:

$$\rho_0(x) = \frac{1}{\lambda_0^2}\Psi_0\left(\frac{x}{\lambda_0^{2\beta}}\right) \in C_0^\infty(\mathbb{R}^3).$$

Finite time blowup. Recalling (4.4.2), we know that

$$\frac{d\lambda}{dt} = \frac{d\lambda}{d\tau}\frac{d\tau}{dt} = -\frac{\lambda}{2}\frac{1}{\lambda^2} = -\frac{1}{2\lambda}, \quad \text{with } \lambda(0) = \lambda_0,$$

which implies that $\lambda(t)$ blows up at $T = \lambda_0^2$ and $\lambda(t)$ is given by

$$\lambda(t) = \sqrt{T-t} = \sqrt{\lambda_0^2 - t}, \quad \forall\, t \in [0, T).$$

Consequently, the solution to (KS-D) with initial data ρ_0 is of form

$$\rho(t,x) = \frac{1}{T-t}\,(Q+\varepsilon)\left(t, \frac{x}{(T-t)^\beta}\right),$$

with

$$\|\varepsilon(\tau)\|_{H^{2K+4}} \leq \epsilon_0 (T-t)^{\frac{\delta_g}{2}},$$

for some $0 < \epsilon_0 \ll 1$. Consequently, we have concluded the proof of Theorem 4.1.1.

□

Chapter 5

KOLMOGOROV-ARNOLD NETWORK

In this chapter, we present Kolmogorov-Arnold Network (KAN), a novel deep learning architecture inspired by the Kolmogorov-Arnold representation theorem, mostly based on our works [293, 292, 446]. Partially motivated by the need for symbolic computation of singularities, KANs offer a powerful tool for science, especially when interpretability is desired. Compared to prevalent Multi-Layer Perceptrons (MLPs) with *fixed* activation functions on *nodes* ("neurons"), KANs have *learnable* activation functions on *edges* ("weights"). KANs learn interpretable 1D functions on their edges whose connection graph is also simple enough to be explained. Through examples in mathematics, KANs are shown to be useful "collaborators" helping scientists (re)discover mathematical and physical laws. Moreover, KANs are shown to be more accurate and have faster scaling laws than MLPs in function fitting and PDE solving, both theoretically and empirically.

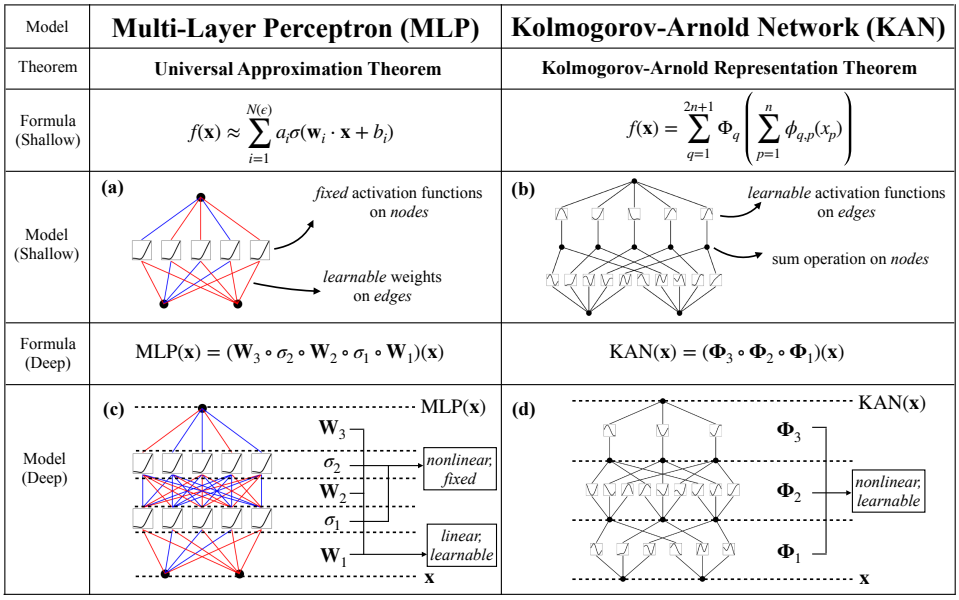


Figure 5.1: Multi-Layer Perceptrons (MLPs) vs. Kolmogorov-Arnold Networks (KANs)

5.1 Introduction

Multi-layer perceptrons (MLPs) [190, 108, 206], also known as fully-connected feedforward neural networks, are foundational building blocks of today's deep learning models. The importance of MLPs can never be overstated, since they are the default models in machine learning for approximating nonlinear functions, due to their expressive power guaranteed by the universal approximation theorem [206]. However, MLPs often lack interpretability, which makes them less useful for tasks when interpretability is key, e.g., when we want to extract symbolic formulas from datasets. In science, symbolic functions are prevalent, e.g., $E = mc^2$ (energy-mass relation), $r = \frac{a}{1+e\cos\theta}$ (ellipse), $p = e^{-\frac{E}{kT}}/Z$ (Boltzman distribution). Although MLPs can numerically approximate these functions to a reasonable accuracy, they cannot reveal symbolic structures of these equations.

Therefore, we need a representation theorem that is more aligned with symbolic representations than the universal approximation theorem. In our search, the good old Kolmogorov-Arnold representation theorem (KA theorem) came to our attention. Although the KA theorem has long been considered irrelevant for learning [172] because the theorem does not guarantee smoothness, we are more optimistic about the smoothness of deeper representations. For example, as we will show, $f(x_1, x_2, x_3, x_4) = \exp(\sin(x_1^2+x_2^2)+\sin(x_3^2+x_4^2))$ can be smoothly represented by a three-layer network, but a two-layer network that attempts to fit this function leads to pathological representations.

Unsurprisingly, the possibility of using Kolmogorov-Arnold representation theorem to build neural networks has been studied [422, 255, 283, 261, 266, 145, 333]. However, most work has stuck with the original depth-2 width-$(2n + 1)$ representation, and many did not have the chance to leverage more modern techniques (e.g., back propagation) to train the networks. Our contribution lies in generalizing the original Kolmogorov-Arnold representation to arbitrary widths and depths, revitalizing and contextualizing it in today's deep learning world, as well as using empirical experiments to highlight its potential for AI + Science due to its accuracy and interpretability.

Named after the two great mathematicians, Andrey Kolmogorov and Vladimir Arnold, this new type of network is called the *Kolmogorov-Arnold Network* (KAN). Like MLPs, KANs have fully-connected structures. However, while MLPs place fixed activation functions on *nodes* ("neurons"), KANs place learnable activation functions on *edges* ("weights"), as illustrated in Figure 5.1. Each learnable weight

parameter in an MLP is replaced by a learnable 1D function (parametrized as a spline) in a KAN. KANs' nodes simply sum incoming signals without applying any non-linearities.

Although interpretability is our initial motivation to develop KANs, KANs demonstrate impressive accuracy and fast scaling laws as well, both theoretically and empirically. Despite their elegant mathematical interpretation, KANs are nothing more than combinations of splines and MLPs, leveraging their respective strengths and avoiding their respective weaknesses. Splines are accurate for low-dimensional functions but suffer from the curse of dimensionality (COD) problem. MLPs, On the other hand, suffer less from COD thanks to their ability to learn features and compositional structure, but are less accurate than splines in low dimensions. KANs have MLPs on the outside and splines on the inside, combining the best of two things into one.

The chapter is organized as follows: is organized as follows: In Section 5.2, we introduce the KAN architecture, analyze the network's approximation ability, and propose two training techniques to make KANs interpretable and accurate. In Section 5.3, we show that KANs are interpretable and can be used for scientific discoveries. We use a knot theory example from mathematics to demonstrate that KANs can be helpful "collaborators" for scientists. In Section 5.4, we show that KANs are more accurate than MLPs for data fitting and PDE solving with better scaling laws. We discuss the superiority of KANs learning high-frequency functions in Section 5.5. We discuss related works, follow-up works inspired by our work, and draw conclusions in Section 5.6.

5.2 Kolmogorov–Arnold Networks (KAN)

Multi-Layer Perceptrons (MLPs) are inspired by the universal approximation theorem. We instead focus on the Kolmogorov-Arnold representation theorem, which can be realized by a new type of neural network called Kolmogorov-Arnold networks (KAN). We review the Kolmogorov-Arnold theorem in Section 5.2.1, to inspire the design of Kolmogorov-Arnold Networks in Section 5.2.2. Section 5.2.3 provides mathematical description of KANs' expressive power. Section 5.2.5 and Section 5.2.4 propose techniques to make KANs accurate and interpretable.

5.2.1 Kolmogorov-Arnold representation theorem

Vladimir Arnold and Andrey Kolmogorov established that if f is a multivariate continuous function on a bounded domain, then f can be written as a finite composition

of continuous functions of a single variable and the binary operation of addition. More specifically, for a smooth $f : [0,1]^n \to \mathbb{R}$,

$$f(\mathbf{x}) = f(x_1, \cdots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right), \tag{5.2.1}$$

where $\phi_{q,p} : [0,1] \to \mathbb{R}$ and $\Phi_q : \mathbb{R} \to \mathbb{R}$. In a sense, they showed that the only true multivariate function is addition, since every other function can be written using univariate functions and sum. One might naively consider this great news for machine learning: learning a high-dimensional function boils down to learning a polynomial number of 1D functions. However, these 1D functions can be non-smooth and even fractal, so they may not be learnable in practice [375]. Because of this pathological behavior, the Kolmogorov-Arnold representation theorem was regarded as theoretically sound but practically useless [375].

However, we are more optimistic about the usefulness of the Kolmogorov-Arnold theorem for machine learning. First of all, we need not stick to the original Eq. (5.2.1) which has only two-layer non-linearities and a small number of terms $(2n+1)$ in the hidden layer: we will generalize the network to arbitrary widths and depths. Deeper and wider networks potentially have stronger expressive power with smooth functions. Moreover, most functions in science and daily life are often smooth and have sparse compositional structures [282], potentially facilitating smooth Kolmogorov-Arnold representations.

5.2.2 KAN architecture

Suppose we have a supervised learning task consisting of input-output pairs $\{x_i, y_i\}$, where we want to find f such that $y_i \approx f(x_i)$ for all data points. Eq. (5.2.1) implies that we are done if we can find appropriate univariate functions $\phi_{q,p}$ and Φ_q. This inspires us to design a neural network which explicitly parametrizes Eq. (5.2.1). Since all functions to be learned are univariate functions, we can parametrize each 1D function as a B-spline curve, with learnable coefficients of local B-spline basis functions [1]. Now we have a prototype of KAN, whose computation graph is exactly specified by Eq. (5.2.1) and illustrated in Figure 5.1 (b) (with the input dimension $n = 2$), appearing as a two-layer neural network with activation functions placed on edges instead of nodes (simple summation is performed on nodes), and with width $2n + 1$ in the middle layer.

[1] Details in Appendix 5.7.1 and illustrated in Figure 5.13 right.

As mentioned, such a network is known to be too simple to approximate any function arbitrarily well in practice with smooth splines! We therefore generalize our KAN to be wider and deeper. The key insight comes from the analogy between MLPs and KANs. In MLPs, once we define a layer (which is composed of a linear transformation and nonlinearties), we can stack more layers to make the network deeper. To build deep KANs, we should first answer: "what is a KAN layer?" It turns out that a KAN layer with n_{in}-dimensional inputs and n_{out}-dimensional outputs can be defined as a matrix of 1D functions

$$\mathbf{\Phi} = \{\phi_{q,p}\}, \qquad p = 1, 2, \cdots, n_{\text{in}}, \qquad q = 1, 2 \cdots, n_{\text{out}}, \tag{5.2.2}$$

where the functions $\phi_{q,p}$ have trainable parameters (parameterized as B-splines, see Appendix 5.7.1), as detaild below. In the Kolmogov-Arnold theorem, the inner functions form a KAN layer with $n_{\text{in}} = n$ and $n_{\text{out}} = 2n + 1$, and the outer functions form a KAN layer with $n_{\text{in}} = 2n + 1$ and $n_{\text{out}} = 1$. So the Kolmogorov-Arnold representations in Eq. (5.2.1) are simply compositions of two KAN layers. Now it becomes clear what it means to have deeper Kolmogorov-Arnold representations: simply stack more KAN layers!

The shape of a general KAN is represented by an integer array

$$[n_0, n_1, \cdots, n_L], \tag{5.2.3}$$

where n_i is the number of nodes in the i^{th} layer of the computational graph. We denote the i^{th} neuron in the l^{th} layer by (l, i), and the activation value of the (l, i)-neuron by $x_{l,i}$. Between layer l and layer $l + 1$, there are $n_l n_{l+1}$ activation functions: the activation function that connects (l, i) and $(l + 1, j)$ is denoted by

$$\phi_{l,j,i}, \quad l = 0, \cdots, L - 1, \quad i = 1, \cdots, n_l, \quad j = 1, \cdots, n_{l+1}. \tag{5.2.4}$$

The pre-activation of $\phi_{l,j,i}$ is simply $x_{l,i}$; the post-activation of $\phi_{l,j,i}$ is denoted by $\tilde{x}_{l,j,i} \equiv \phi_{l,j,i}(x_{l,i})$. The activation value of the $(l + 1, j)$ neuron is simply the sum of all incoming post-activations:

$$x_{l+1,j} = \sum_{i=1}^{n_l} \tilde{x}_{l,j,i} = \sum_{i=1}^{n_l} \phi_{l,j,i}(x_{l,i}), \qquad j = 1, \cdots, n_{l+1}. \tag{5.2.5}$$

In matrix form, this reads

$$\mathbf{x}_{l+1} = \underbrace{\begin{pmatrix} \phi_{l,1,1}(\cdot) & \phi_{l,1,2}(\cdot) & \cdots & \phi_{l,1,n_l}(\cdot) \\ \phi_{l,2,1}(\cdot) & \phi_{l,2,2}(\cdot) & \cdots & \phi_{l,2,n_l}(\cdot) \\ \vdots & \vdots & & \vdots \\ \phi_{l,n_{l+1},1}(\cdot) & \phi_{l,n_{l+1},2}(\cdot) & \cdots & \phi_{l,n_{l+1},n_l}(\cdot) \end{pmatrix}}_{\mathbf{\Phi}_l} \mathbf{x}_l, \tag{5.2.6}$$

where $\mathbf{\Phi}_l$ is the function matrix corresponding to the l^{th} KAN layer. A general KAN network is a composition of L layers: given an input vector $x_0 \in \mathbb{R}^{n_0}$, the output of KAN is

$$\text{KAN}(\mathbf{x}) = (\mathbf{\Phi}_{L-1} \circ \mathbf{\Phi}_{L-2} \circ \cdots \circ \mathbf{\Phi}_1 \circ \mathbf{\Phi}_0)\mathbf{x}. \tag{5.2.7}$$

We can also rewrite the above equation to make it more analogous to Eq. (5.2.1), assuming output dimension $n_L = 1$, and define $f(\mathbf{x}) \equiv \text{KAN}(\mathbf{x})$:

$$f(\mathbf{x}) = \sum_{i_{L-1}=1}^{n_{L-1}} \phi_{L-1,i_L,i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \cdots \left(\sum_{i_2=1}^{n_2} \phi_{2,i_3,i_2} \left(\sum_{i_1=1}^{n_1} \phi_{1,i_2,i_1} \left(\sum_{i_0=1}^{n_0} \phi_{0,i_1,i_0}(x_{i_0}) \right) \right) \right) \cdots \right), \tag{5.2.8}$$

which is quite cumbersome. In contrast, our abstraction of KAN layers and their visualizations are cleaner and intuitive. The original Kolmogorov-Arnold representation Eq. (5.2.1) corresponds to a 2-Layer KAN with shape $[n, 2n+1, 1]$. Notice that all the operations are differentiable, so we can train KANs with back propagation. For comparison, an MLP can be written as interleaving of affine transformations $\mathbf{W}$ and non-linearities σ:

$$\text{MLP}(\mathbf{x}) = (\mathbf{W}_{L-1} \circ \sigma \circ \mathbf{W}_{L-2} \circ \sigma \circ \cdots \circ \mathbf{W}_1 \circ \sigma \circ \mathbf{W}_0)\mathbf{x}. \tag{5.2.9}$$

It is clear that MLPs treat linear transformations and nonlinearities separately as $\mathbf{W}$ and σ, while KANs treat them all together in $\mathbf{\Phi}$. In Figure 5.1 (c) and (d), we visualize a three-layer MLP and a three-layer KAN, to clarify their differences. We use k-th order B-splines to parameterize the nonlinearities, and implementation details of KANs are left in Appendix 5.7.1.

Remark: Complexities. Assuming a KAN with depth L, width N, grid size G, spline order k. The model has $O(N^2GL)$ parameters. Suppose a training batch has size B, memory usage is $O(2^kBN^2GL)$, the number of operations is $O(2^kBN^2GL)$ both for forward and backward runs. The 2^k factor is due to the recursive computation of order k splines.

5.2.3 KAN's approximation abilities and scaling laws

Recall that in Eq. (5.2.1), the 2-Layer width-$(2n+1)$ representation may be non-smooth. However, deeper representations may bring the advantages of smoother activations. To facilitate an approximation analysis, we still assume smoothness of activations, but allow the representations to be arbitrarily wide and deep, as in Eq. (5.2.7). To emphasize the dependence of our KAN on the finite set of grid points, we use $\mathbf{\Phi}_l^G$ and $\Phi_{l,i,j}^G$ below to replace the notation $\mathbf{\Phi}_l$ and $\Phi_{l,i,j}$ used in Eq. (5.2.5) and (5.2.6).

Theorem 5.2.1 (Approximation theory, KAN)**.** *Let* $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. *Suppose that a function* $f(\mathbf{x})$ *admits a representation*

$$f = (\mathbf{\Phi}_{L-1} \circ \mathbf{\Phi}_{L-2} \circ \cdots \circ \mathbf{\Phi}_1 \circ \mathbf{\Phi}_0)\mathbf{x}\,, \tag{5.2.10}$$

as in Eq. (5.2.7), where each one of the $\Phi_{l,i,j}$ *are* $(k+1)$*-times continuously differentiable. Then there exists a constant* C *depending on* f *and its representation, such that we have the following approximation bound in terms of the grid size* G*: there exist* k*-th order B-spline functions* $\Phi_{l,i,j}^G$ *such that for any* $0 \le m \le k$*, we have the bound*

$$\|f - (\mathbf{\Phi}_{L-1}^G \circ \mathbf{\Phi}_{L-2}^G \circ \cdots \circ \mathbf{\Phi}_1^G \circ \mathbf{\Phi}_0^G)\mathbf{x}\|_{C^m} \le C G^{-k-1+m}\,. \tag{5.2.11}$$

Here we adopt the notation of C^m*-norm measuring the magnitude of derivatives up to order* m*:*

$$\|g\|_{C^m} = \max_{|\beta|\le m} \sup_{\mathbf{x}\in[0,1]^n} \left|D^\beta g(\mathbf{x})\right|.$$

We leave the proof and an in-depth discussion on the implications of the theorem in Subsection 5.7.2. Asymptotically, provided that the assumption in Theorem 5.2.1 holds, KANs with finite grid size can approximate the function well with a residue rate independent of the dimension. This comes naturally since we only use splines to approximate 1D functions. In particular, for $m = 0$, we recover the accuracy in L^∞ norm, which in turn provides a bound of RMSE on the finite domain, which gives a scaling exponent $k+1$. Of course, the constant C is dependent on the representation; hence it will depend on the dimension. Notice that if the assumption in the theorem holds for a shallow KAN, it automatically holds for a deeper KAN by setting the remaining layers to identity. We also remark that: since the assumption in the theorem is a strong one, the neural scaling law should not be expected to be universally applicable to all machine learning applications.

KANs take advantage of the intrinsically low-dimensional compositional representation of underlying functions. This result shares an analogy to the rate in generalization error bounds of finite training samples, for a similar space studied for regression problems; see [207, 252], and also specifically for MLPs with ReLU activations [402]. On the other hand, for general Sobolev or Besov spaces, sharp approximation rates have been obtained for ReLU-MLPs (and more generally MLPs with most piecewise polynomial activation functions) [464, 22, 412]. These rates exhibit the curse of dimensionality, which is unavoidable due to the fact that Sobolev and Besov spaces with fixed smoothness are very large in high dimensions. By leveraging the representations of KANs using MLPs, we can provide a more general version of approximation theory in a larger function class, establishing KANs as universal approximators, as stated below.

We establish that MLPs can be represented using KANs of a comparable size. Specifically, we show that any MLP with the ReLU^k activation function can be reparameterized as a KAN with a comparable number of parameters. This shows that the approximation and representation capabilities of KANs are at least as good as MLPs.

Theorem 5.2.2. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Suppose that a function $f : \mathbb{R}^d \to \mathbb{R}$ can be represented by an MLP with width $W \geq 1$, depth $L \geq 1$, and activation function $\sigma_k = \max(0, x)^k$ for $k \geq 1$. Then there exists a KAN g with width W, depth at most $2L$, and grid size $G = 2$ with k-th order B-spline functions such that*

$$g(\mathbf{x}) = f(\mathbf{x}) \tag{5.2.12}$$

for all $\mathbf{x} \in \Omega$.

As a corollary, we can draw conclusions about the approximation capabilities of KANs by leveraging existing results about MLPs (see for instance [21, 267, 251, 413, 414, 204, 295, 409, 465, 412, 463, 462]). For example, we have the following result giving optimal approximation rates for very deep KANs on Sobolev spaces (see for instance [3] for the background on Sobolev spaces).

Corollary 5.2.3. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary, $s > 0$ and $1 \leq p, q \leq \infty$ be such that $1/q - 1/p < s/d$. This guarantees that the compact Sobolev embedding $W^s(L_q(\Omega)) \subset\subset L_p(\Omega)$ holds.*

Let $W_0 := W_0(d)$ be a fixed width (depending upon the input dimension d). Then for any $f \in W^s(L_q(\Omega))$ and any $L \geq 1$, there exists a KAN g with width W_0, depth

L, and grid size $G = 2$ with k-th order B-spline functions such that

$$\|f - g\|_{L_p(\Omega)} \leq CL^{-2s/d}, \tag{5.2.13}$$

where C is a constant independent of L.

This result, which follows immediately from Theorem 5.2.2 and the approximation rates for ReLU (and more generally piecewise polynomial) neural networks derived in [412], shows that very deep KANs attain an exceptionally good approximation rate on Sobolev spaces. In particular, in terms of the number of parameters P they attain an approximation rate of $O(P^{-2s/d})$, while a classical (even non-linear) method of approximation can only attain a rate of $O(P^{-s/d})$ [118]. This phenomenon, which is often called superconvergence [117], also occurs for very deep ReLU^k networks. However, it comes at the cost of parameters which are not encodable using a fixed number of bits and thus is not practically realizable [466, 412].

Now that the basic architecture of KANs is in place, we propose a few techniques to make KANs accurate and interpretable.

5.2.4 Tricks for interpretability: pruning and symbolifying KANs

How do we choose the KAN shape? If we know that the dataset is generated via the symbolic formula $f(x, y) = \exp(\sin(\pi x) + y^2)$, then we know that a $[2, 1, 1]$ KAN is able to express this function. However, in practice we do not know the shape a priori, so it would be nice to have approaches to determine this shape automatically. The idea is to start from a large enough KAN and train it with sparsity regularizations followed by pruning. One may even symbolify activation functions into symbolic functions like exp, sine, etc, to make KANs a useful tool for symbolic regression. The idea is to match learned spline functions with candidates in a symbolic function library specified by human users and replace the spline functions with the best-fitting ones.

5.2.5 Tricks for accuracy: grid update and grid extension

Grid update Since input data and (especially) hidden activations can have time-varying ranges in training, we update grids on the fly based on the statistics of input/activation ranges. The grid is initialized to be in [-1,1] (e.g., when $G = 5$, the grid points are [-1, -0.6, -0.2, 0.2, 0.6, 1.0]), but once it receives input/activations, say, in the range [-3,3] (the maximum and minimum values are 3 and -3, respectively), the grid will be updated to [-3,3] (correspondingly, grid points become [-3,-1.8,-0.6,0.6,1.8,3.0]) to accommodate the whole range.

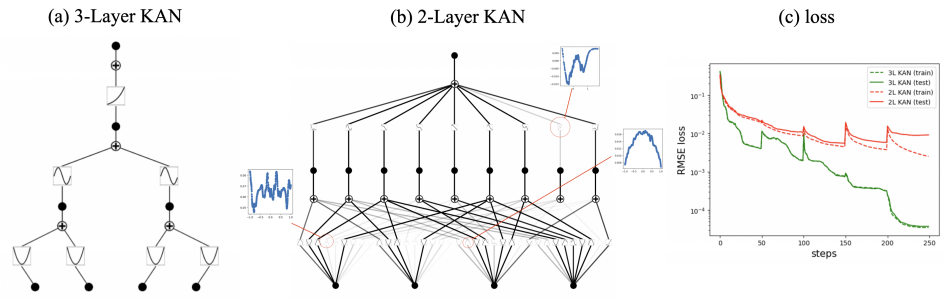


Figure 5.2: Fitting the function $f(x_1, x_2, x_3, x_4) = \exp(\frac{1}{2}(\sin(\pi(x_1^2+x_2^2))+\sin(\pi(x_3^2+x_4^2))))$. (a) 3-Layer KAN admits smooth representations. (b) The 2-Layer KAN learns highly oscillatory representations. (c) The 3-layer KAN achieves lower losses and has a smaller train-test gap than the 2-layer KAN.

Grid extension A spline can be made arbitrarily accurate to a target function as the grid can be made arbitrarily fine-grained. This good feature can be inherited by KANs. By contrast, MLPs do not have the notion of "fine-graining". For KANs, one can first train a KAN with fewer parameters and then extend it to a KAN with more parameters by simply making its spline grids finer, without the need to retrain the larger model from scratch. The main idea of grid extension is: for each 1D function defined on a coarse grid, we determine the coefficient of a finer grid using least squares that minimize the difference between the two curves evaluated on data samples. Details of how to perform grid extension are included in Figure 5.13.

5.2.6 Benefits of deep KANs

It is one of our major contributions to generalize the 2-layer KA representations to multiple layers. Although it is challenging to prove the benefits of deeper KANs theoretically, we want to present a concrete example where 3-layer KANs admit smooth representations while 2-layer KANs do not. We consider fitting a function $f(x_1, x_2, x_3, x_4) = \exp(\frac{1}{2}(\sin(\pi(x_1^2 + x_2^2)) + \sin(\pi(x_3^2 + x_4^2))))$ where we draw samples (3000 training, 1000 training) uniformly from $[-1, 1]^4$. We train a 3L KAN ([4,2,1,1]) and a 2L KAN ([4,9,1]) with the LBFGS optimizer for 250 steps, with increasing $G = 3, 5, 10, 20, 50$ (50 steps for each G). As shown in Figure 5.2, we see that the 3-layer KAN has smooth representations (as expected, since the parse tree of the symbolic formula has depth 3), while the 2-layer KAN learns highly oscillatory functions on some edges. The 3-layer KAN also achieves lower losses than the 2-layer KAN. While the 3-layer KAN has a small train-test gap, the 2-layer KAN starts to overfit at large grid sizes.

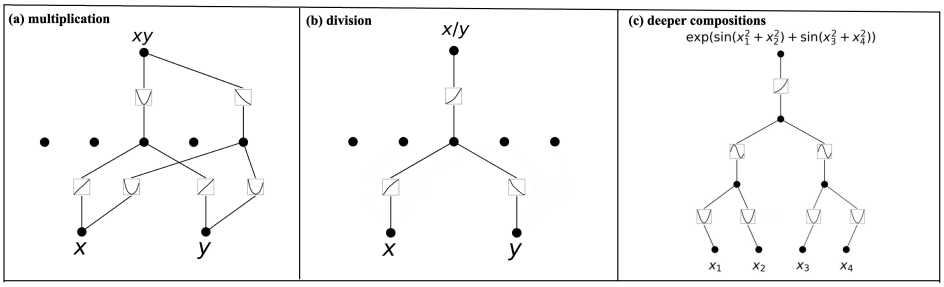


Figure 5.3: KANs are interpretable for simple symbolic tasks.

5.3 KANs Are Interpretable

In this section, we show that KANs can be interpretable on synthetic toy tasks and realistic research questions in math and physics.

Synthetic toy datasets We first examine KANs' ability to reveal the compositional structures in symbolic formulas. Three examples are presented in Figure 5.3. KANs are able to reveal the compositional structures present in these formulas, as well as learn the correct univariate functions. (1) Multiplication $f(x,y) = xy$. KAN computes it via the equation $2xy = (x+y)^2 - (x^2+y^2)$. (2) Division of positive numbers $f(x,y) = x/y$. KAN computes it via $\exp(\log x - \log y)$. (3) Deeper compositions $f(x_1,x_2,x_3,x_4) = \exp(\sin(x_1^2+x_2^2) + \sin(x_3^2+x_4^2))$.

Application to Mathematics: Knot Theory Knot theory is a subject in low-dimensional topology that sheds light on topological aspects of three-manifolds and four-manifolds and has a variety of applications, including in biology and topological quantum computing. In [110], supervised learning and human domain experts were utilized to arrive at a new theorem relating algebraic and geometric knot invariants. They use network attribution methods to find that the signature σ is mostly dependent on meridinal distance μ (real μ_r, imag μ_i) and longitudinal distance λ. We show that KANs can not only identify these important variables with much smaller networks and much more automation, but also present some interesting new results and insights.

We treat 17 knot invariants as inputs and signature as outputs. Similar to the setup in [110], signatures (which are even numbers) are encoded as one-hot vectors and networks are trained with cross-entropy loss. We find that an extremely small $[17,1,14]$ KAN is able to achieve 81.6% test accuracy (while DeepMind's 4-layer width-300 MLP achieves 78% test accuracy). The $[17,1,14]$ KAN ($G=3$, $k=3$) has ≈ 200 parameters, while the MLP has $\approx 3\times10^5$ parameters. It is remarkable that

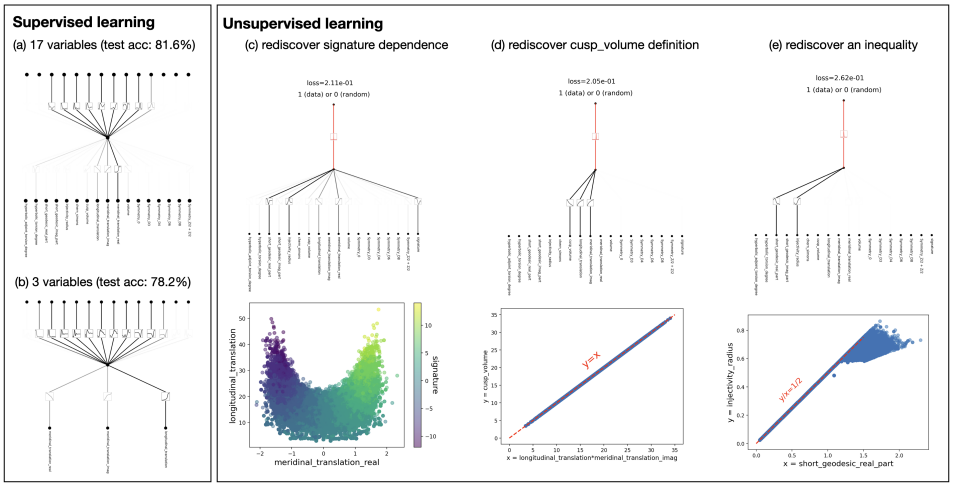


Figure 5.4: Knot dataset. Supervised mode (left): we rediscover DeepMind's three important variables. Unsupervised mode (right): we discover three "new" relations without supervision.

KANs can be both more accurate and much more parameter efficient than MLPs at the same time. In terms of interpretability, we scale the transparency of each activation according to its magnitude, so it becomes immediately clear which input variables are important without the need for feature attribution (see Figure 5.4 left top): signature is mostly dependent on μ_r, and slightly dependent on μ_i and λ, while dependence on other variables is small. We then train a $[3, 1, 14]$ KAN on the three important variables, obtaining test accuracy 78.2% (Figure 5.4 left bottom).

We attempt to make discoveries beyond DeepMind's in the unsupervised learning mode, where we treat all 18 variables (including signature) as inputs. We train 200 networks with different random seeds. They can be grouped into three clusters, with representative KANs displayed in Figure 5.4. These three groups of dependent variables are (1) rediscovering DeepMind's relation in unsupervised learning. (2) cusp volume is by definition of the multiplication of two translations. (3) short geodesic g_r is upper bounded by two times of injecitivy radius [367]. It is interesting that KANs' unsupervised mode can rediscover several known mathematical relations. The good news is that the results discovered by KANs are probably reliable; the bad news is that we have not discovered anything new yet. It is worth noting that we have chosen a shallow KAN for simple visualization, but deeper KANs can probably find more relations if they exist. We would like to investigate how to discover more complicated relations with deeper KANs in future work.

5.4 KANs Are Accurate

In this section, we demonstrate that KANs are more accurate at representing functions than MLPs in various tasks (regression and PDE solving). When comparing two families of models, it is fair to compare both their accuracy (loss) and their complexity (number of parameters). All experiments reported in the work are reproducible on CPUs, usually within minutes, at most in a day. Codes are built based on PyTorch [363].

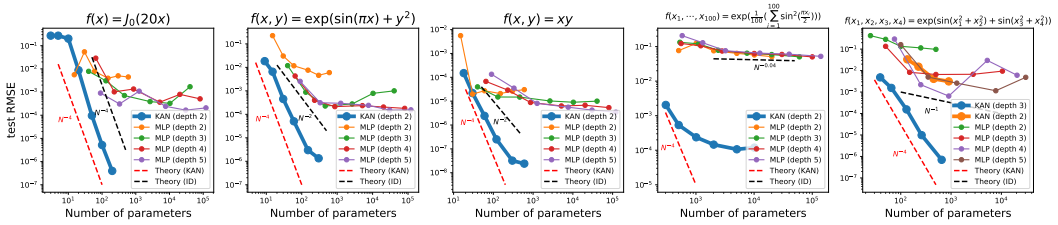


Figure 5.5: Compare KANs to MLPs on five toy examples. KANs can almost saturate the fastest scaling law predicted by our theory ($\alpha = 4$), while MLPs scales slowly and plateau quickly.

Toy datasets In Section 5.2.3, our theory suggested that test RMSE loss ℓ scales as $\ell \propto N^{-(k+1)} = N^{-4}(k = 3)$ with model parameters N. However, this relies on the existence of a smooth Kolmogorov-Arnold representation. As a sanity check, we construct five examples we know have smooth KA representations: (1) $f(x) = J_0(20x)$, which is the Bessel function. Since it is a univariate function, it can be represented by a spline, which is a $[1, 1]$ KAN. (2) $f(x, y) = \exp(\sin(\pi x) + y^2)$. We know that it can be exactly represented by a $[2, 1, 1]$ KAN. (3) $f(x, y) = xy$. We know from Figure 5.3 that it can be exactly represented by a $[2, 2, 1]$ KAN. (4) A high-dimensional example $f(x_1, \cdots, x_{100}) = \exp(\frac{1}{100}\sum_{i=1}^{100}\sin^2(\frac{\pi x_i}{2}))$ which can be represented by a $[100, 1, 1]$ KAN. (5) A four-dimensional example $f(x_1, x_2, x_3, x_4) = \exp(\frac{1}{2}(\sin(\pi(x_1^2 + x_2^2)) + \sin(\pi(x_3^2 + x_4^2))))$ which can be represented by a $[4, 4, 2, 1]$ KAN. The empirical scaling for KANs is quite aligned with theory and outperforms MLPs.

Fitting Images We task KANs with three images: (1) The Cameraman picture is the standard picture for the image fitting task. (2) The turbulence profile is taken from PDEBench [428], demonstrating high-frequency and fractal behavior typical in scientific computing. (3) Van Gogh's *The Starry Night* is quite challenging because it contains fine-grained details as well. In addition to MLPs, We compare KANs with these stronger baselines: (A) MLP with random Fourier features (MLP_RFF).

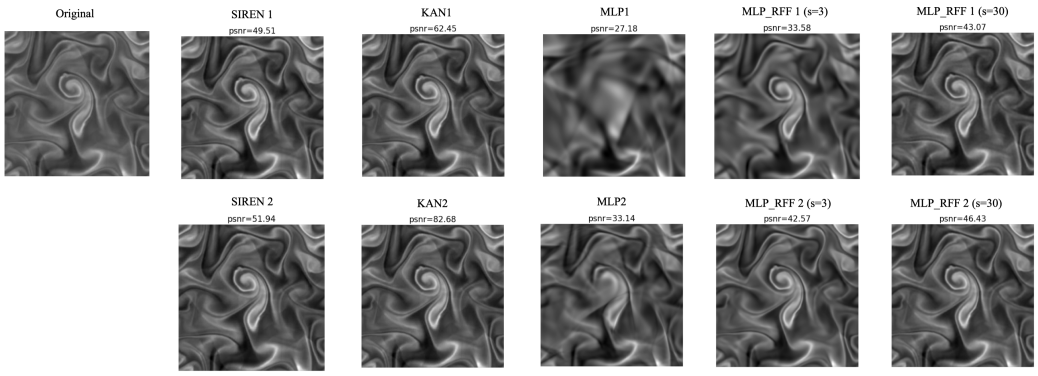


Figure 5.6: Image fitting task (a PDE solution from PDEBench [428]). KAN outperforms baseline methods in terms of PSNR.

Before feeding input coordinates $\mathbf{x} \equiv (x, y)$ to the MLP, we first augment them into a higher-dimensional feature space $\Phi(\mathbf{x}) = (\mathbf{x}, \Phi_1(\mathbf{x}), \cdots, \Phi_{N_f}(\mathbf{x}))$, where $\Phi_i(\mathbf{x}) = (\cos(\mathbf{s}_i \cdot \mathbf{x}), \sin(\mathbf{s}_i \cdot \mathbf{x})), i = 1, \cdots, N_f$, and $\mathbf{s}_i \sim \mathcal{N}(0, s^2)$ (s controls the frequency bias). We choose $N_f = 50$ and $s = 3, 30$. (B) SIREN [415] uses sines as activation functions in MLPs and uses large initialization for the first layer (effectively creating high-frequency features). To compare KANs and baselines as fairly as possible, we try two control strategies (same shape or number of parameters) and report both performance (measured by PSNR) and efficiency (wall time). For all baseline models, 1 means their width is the same as KAN 1, while 2 means their number of parameters is (approximately) the same as KAN 1 ($\sqrt{G}$ times wider, where $G = 10$ is the grid size used in KAN 1). We also explore KAN 2, which uses a finer grid ($G = 100$ instead of $G = 10$) for the first layer only (inspired by the idea of random Fourier features in the input layer). The whole image is treated as the training set and there is no test set. All models are trained with the Adam Optimizer for 15000 steps with learning rate decay (5000 steps for learning rate 10^{-3}, 10^{-4} and 10^{-5}), with batch size 1024, on a V100 GPU.

We have a few observations from the results: (1) KANs are comparable to or even outperform baseline methods (including SIREN) in terms of PSNR, however with more training time. (2) Having random features in the inputs is useful for MLPs, especially high-frequency random features ($s = 30$ outperforms $s = 3$). We may also understand KANs' superior performance as being good at generating random features in early layers. By changing the grid size in the first layer from $G = 10$ to $G = 100$ (KAN 2), PSNR significantly increases with little additional overhead in training time. We show the turbulence profile in Figure 5.6.

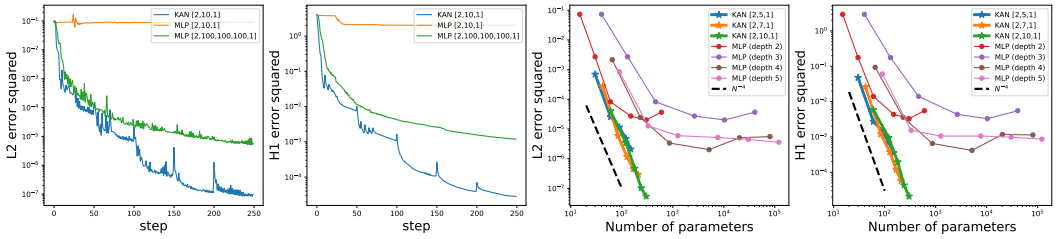


Figure 5.7: The PDE example. We plot L2 squared and H1 squared losses between the predicted solution and ground truth solution. First and second: training dynamics of losses. Third and fourth: scaling laws of losses against the number of parameters. KANs converge faster, achieve lower losses, and have steeper scaling laws than MLPs.

Solving partial differential equations (PDEs We consider a Poisson equation with zero Dirichlet boundary data. For $\Omega = [-1, 1]^2$, consider the PDE $u_{xx}+u_{yy} = f$ with zero boundary condition. We consider the data $f = -\pi^2(1+4y^2)\sin(\pi x)\sin(\pi y^2)+2\pi\sin(\pi x)\cos(\pi y^2)$ for which $u = \sin(\pi x)\sin(\pi y^2)$ is the true solution. We use the framework of physics-informed neural networks (PINNs) [382, 244] to solve this PDE, with the loss function given by $\text{loss}_{\text{pde}} = \alpha\text{loss}_i + \text{loss}_b := \alpha\frac{1}{n_i}\sum_{i=1}^{n_i}|u_{xx}(z_i) + u_{yy}(z_i) - f(z_i)|^2 + \frac{1}{n_b}\sum_{i=1}^{n_b} u^2$, where we use loss_i to denote the interior loss, discretized and evaluated by a uniform sampling of n_i points $z_i = (x_i, y_i)$ inside the domain, and similarly we use loss_b to denote the boundary loss, discretized and evaluated by a uniform sampling of n_b points on the boundary. $\alpha = 0.01$ is the hyperparameter balancing the effect of the two terms. KANs are shown to have Pareto frontiers than MLPs for this simple example.

5.5 KANs Have Less Spectral Bias

In this section, from the perspective of learning and optimization, we study the spectral bias of KANs compared with MLPs. Standard MLPs with ReLU activations (or even tanh) are known to suffer from the spectral bias [381, 460, 459], in the sense that they will fit low-frequency components first. This is in contrast to traditional iterative numerical methods like the Jacobi method that learn high frequencies first [460]. Although the spectral bias acts as a regularizer that improves performance for machine learning applications [381, 472, 376, 476, 152], for scientific computing applications, it may be necessary to learn high-frequencies as well. To alleviate the spectral bias, high-frequency information has to be encoded using methods like Fourier feature mapping [415, 429, 23, 349], or one needs to use nonlinear activation functions more similar to traditional methods; see for example, the hat activation

function [205] which resembles a finite element basis. We demonstrate that KANs are less biased toward low frequencies than MLPs. We highlight that the multi-level learning feature specific to KANs, i.e. grid extension of splines, improves the learning process for high-frequency components. Detailed comparisons with different choices of depth, width, and grid sizes of KANs are made, shedding some light on how to choose the hyperparameters in practice. We remark that although the spectral bias is considered a form of regularization which is desirable for machine learning tasks [381, 472, 376, 476, 152], for scientific computing applications it is typically important to capture all frequencies and so the spectral bias may negatively affect the performance of neural networks for such applications [443, 386, 55, 205].

5.5.1 Spectral bias theory for shallow KANs

We consider the spectral bias properties of KANs with a single layer. This theory is very similar to the theory developed in [205, 475, 394] for the spectral bias of single hidden layer MLPs. The key observation is that a single layer KAN is a linear model. In particular, we see that if $L = 1$ then the KAN applied to an input $\mathbf{x} \in \mathbb{R}^d$ is (for simplicity, we consider the case of a KAN without the SiLu non-linearity)

$$\mathrm{KAN}(\mathbf{x}, \theta)_i = \sum_{j=1}^{d} \sum_{l=1}^{G+k-1} c_{ijl} B_l(x_j), \tag{5.5.1}$$

where $\theta = \{c_{ijl}\}$ are the parameters of the KAN. Here $i = 1, ..., d'$ where d' is the dimension of the output, $j = 1, ..., d$, and $l = 1, ..., G + k - 1$. Note that the only parameter here is the grid size G, since the width is determined by the input and output dimensions.

Based upon this, we can analyze least squares fitting with shallow KANs. In particular, let $\Omega = [-1, 1]^d$ be the (symmetric) unit cube in $\mathbb{R}^d$, let $f : \Omega \to \mathbb{R}^{d'}$ be a target function we are trying to learn, and consider the (continuous) least squares regression loss

$$L(\theta) = \int_\Omega \|f(\mathbf{x}) - \mathrm{KAN}(\mathbf{x}, \theta)\|^2 d\mathbf{x}. \tag{5.5.2}$$

Due to the representation (5.5.1), this loss function is quadratic in the parameters θ. Let M denote the corresponding Hessian matrix, i.e. so that

$$L(\theta) = (1/2)\theta^T M \theta + b^T \theta.$$

This Hessian matrix (indexed by i, j, l) is given by

$$M_{(i,j,l),(i',j',l')} = \begin{cases} \int_\Omega B_l(x_j) B_{l'}(x_{j'}) d\mathbf{x} & i = i' \\ 0 & i \neq i'. \end{cases} \tag{5.5.3}$$

The convergence of gradient descent on the least squares regression is determined by the eigendecomposition of the Hessian matrix M which is estimated in the following theorem. The theorem is essentially a generalization of the fact that the Gram matrix of the B-spline basis is well conditioned (see for instance [119], Theorem 4.2 in Chapter 5).

Theorem 5.5.1. *Given a single hidden layer KAN with grid size G, degree k B-splines, input dimension d and output dimension d', let M denote the Hessian matrix defined in (5.5.3) corresponding to the least squares fitting problem (5.5.2). Then the eigenvalues $0 \leq \lambda_1(M) \leq \cdots \leq \lambda_N(M)$ (here $N = (G + k - 1)dd'$) satisfy*

$$\frac{\lambda_N(M)}{\lambda_{d'(d-1)+1}(M)} \leq Cd \tag{5.5.4}$$

for a constant C depending only on the spline degree k.

Theorem 5.5.1 shows that away from $d'(d-1)$ eigenvectors the matrix M is well conditioned. This means that gradient descent will converge at the same rate in all directions orthogonal to these $d'(d-1)$ eigenvectors. Note that since the number of eigenvectors we must remove is independent of the grid size G, we expect that when G is relatively large most components of the KAN will converge at roughly the same rate toward the solution. Thus the KAN with a large number of grid points will not exhibit the same spectral bias toward low frequencies seen by MLPs. We remark that in constrast the Hessian associated with a two layer ReLU MLP with width n has a condition number which scales like n^4 [205], which explains the strong spectral bias exhibited by ReLU MLPs.

Remark 5.5.2. *We note that the $d'(d-1)$ eigenvectors which must be excluded is not an artifact of the proof. In fact, this is due to the fact that the KAN parameterization is not unique. Indeed, the constant function $f(x) = 1$ can be parameterized in d different ways by using the B-splines in each of the d different directions. This ambiguity gives rise to directions in parameter space where the function parameterized by the KAN doesn't change and this results in degenerate eigenvectors of the matrix M.*

The analysis given is necessarily highly simplified and heuristic. In particular, we only analyze a single layer of the KAN network and consider the continuous least squares loss. Nonetheless, we argue that it gives an explanation for why we would expect KANs to have a significantly different spectral bias than MLPs, and

in particular why we expect that they learn all frequencies roughly similarly. In the remainder of this section, we experimentally test this hypothesis and compare the spectral bias of KANs with MLPs on a variety of simple problems. We implement these numerical experiments using the pykan package version 0.2.5.

5.5.2 1D waves of different frequencies

In the first example, we take the same setting as in [381] and study the regression of a linear combination of waves of different frequencies. Consider the function prescribed as

$$f(x) = \sum A_i \sin(2\pi k_i z + \varphi_i), \quad k = (5, 10, \cdots, 45, 50).$$

The phases φ_i are uniformly sampled from $[0, 2\pi]$ and we consider two cases of amplitudes: one with equal amplitude $A_i = 1$ and another with increasing amplitude $A_i = 0.1i$. We use a neural network, either ReLU MLP or KAN, to regress f sampled at 200 uniformly spaced points in $[0, 1]$, with full batch ADAM iteration as the optimizer with a learning rate of 0.0003. For MLPs, we train with 80000 iterations as in [381]; for KANs, we only train with 8000 iterations. Normalized magnitudes of discrete Fourier transform at frequencies k_i are computed as $|\tilde{f}_{k_i}/A_i|$ and averaged over 10 runs of different phases.

We plot the evolution of $|\tilde{f}_{k_i}/A_i|$ during training across all frequencies; see Figures 5.8, 5.9 for comparisons of MLPs and KANs with different sizes for equal and increasing amplitudes respectively. KANs suffer significantly less than MLPs from spectral biases. Once the size of KANs, especially the grid size and depth is large enough, KANs almost learn all frequencies at the same time, while even very deep and wide MLPs still have difficulties learning higher frequencies, even with 10x epochs!

5.5.3 Gaussian random field

In this example, we consider fitting functions sampled from a Gaussian random field. The target function f is sampled from a d-dimensional Gaussian random field with mean zero and covariance $\exp(-|x - y|^2/(2\sigma^2))$. Here small σ corresponds to rough functions and large σ corresponds to smooth functions.

To approximate the Gaussian random field, we sample f using the KL expansion [243]. We sample $N = 5000$ points uniformly from $[-1, 1]^d$ and calculate the (empirical) covariance matrix K. Then we truncate its first $m < N$ eigenpairs λ_i, ϕ_i,

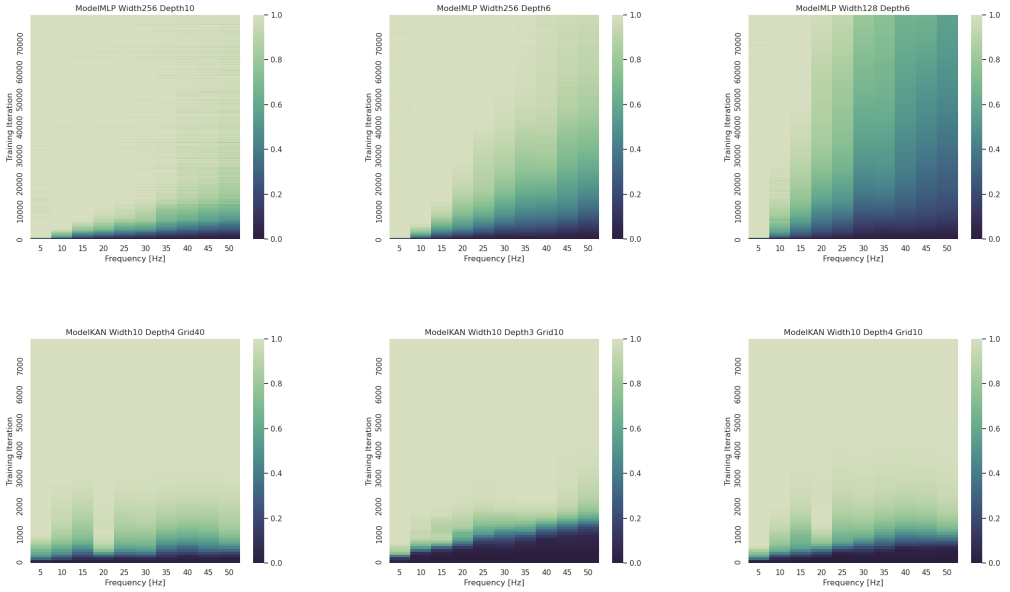


Figure 5.8: 1D wave dataset, where the target function has equal amplitudes of different frequency modes. Under various hyperparameters, MLPs manifest strong spectral biases (top), while KANs do not (bottom). Note that the y axis (training steps) of MLP is 10 times that of KAN.

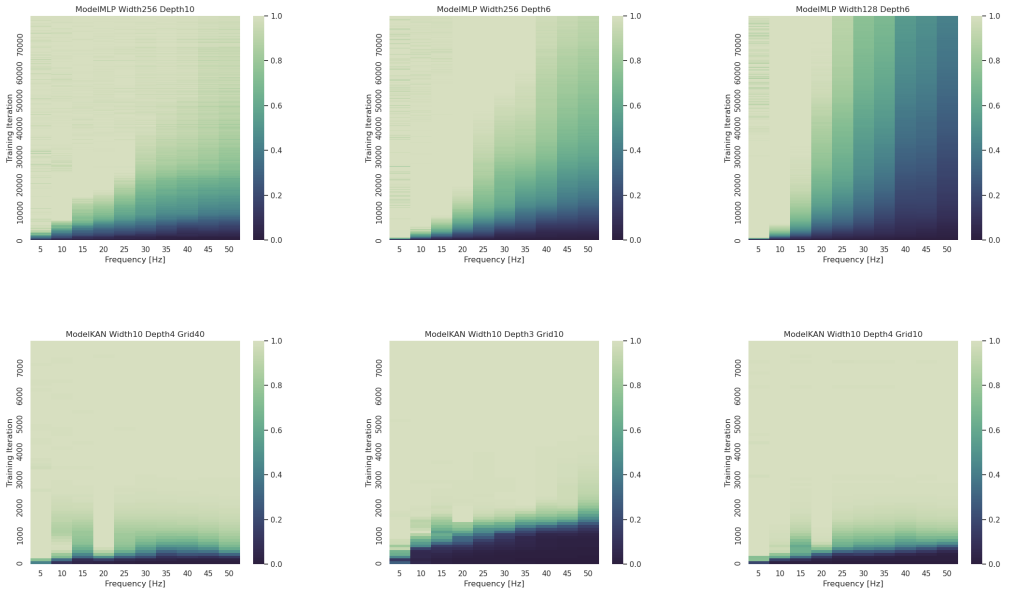


Figure 5.9: 1D wave dataset, where the target function has increasing amplitudes of different frequency modes. Under various hyperparameters, MLPs manifest severe spectral biases (top), while KANs do not (bottom). Note that the y axis (training steps) of MLP is 10 times that of KAN.

with the cutoff threshold $\lambda_{m+1} < 0.1\lambda_1 \leq \lambda_m$ and sample f approximately via

$$f = \sum_{i \leq m} \lambda_i \xi_i \phi_i,$$

where ξ_i are i.i.d standard Gaussians $N(0,1)$. For f with different scales σ and dimensions d, we split the points into 80% training and 20% testing points. We use MLPs and KANs with different sizes to regress on the training set, with the mean squared loss as the loss function. For MLPs, we use 500 iterations of LBFGS iteration, and for KANs, we use the grid extension technique, with grid sizes $(10, 20, 30, 40, 50)$, each trained with 100 iterations of LBFGS.

We plot the loss curves here and compare the losses of different scales σ and dimensions d, using an MLP of 6 layers and 256 neurons in each hidden layer, and KANs with 10 neurons in each hidden layer and $2, 3, 4$ layers; see Figure 5.10 for the regression loss on the training set with dimensions $2, 3, 4$ and scales 2^i, $i = 0, -1, -2, -3$. We see that for larger scale and smoother functions, MLP performs better, while for smaller scale and rough functions, KANs perform better without suffering much from spectral biases, and grid extension is especially helpful. We remark that one can choose smaller grid sizes of KANs for smoother functions and obtain more accurate regressions.

Precisely since KANs are not susceptible to spectral biases, they are likely to overfit to noises. As a consequence, we notice that KANs are more subject to overfitting on the training data regression when the task is very complicated; see the second line of Figure 5.11. On the other hand, we can increase the number of training points to alleviate the overfitting; see the last line of Figure 5.11 where we increased the number of training and test samples by 10x. We remark that the current implementation of grid extension is prone to oscillation after refining grids during the undersampled regime, as observed in [390], and we will improve it in future works.

5.5.4 PDE example

In this example, we solve the 1D Poisson equation with a high-frequency solution, similar to [460]. To be precise, consider the equation with zero Dirichlet boundary condition

$$-u_{xx} = f \quad \text{in}[-1, 1]\,, \quad u(-1) = u(1) = 0\,. \tag{5.5.5}$$

Here for a frequency $k \in \mathbb{N}$, the right-hand-side and the associated true solution are

$$f = \pi^2 \sin(\pi x) + \pi^2 k \sin(k\pi x), \quad u = \sin(\pi x) + \frac{1}{k} \sin(k\pi x).$$

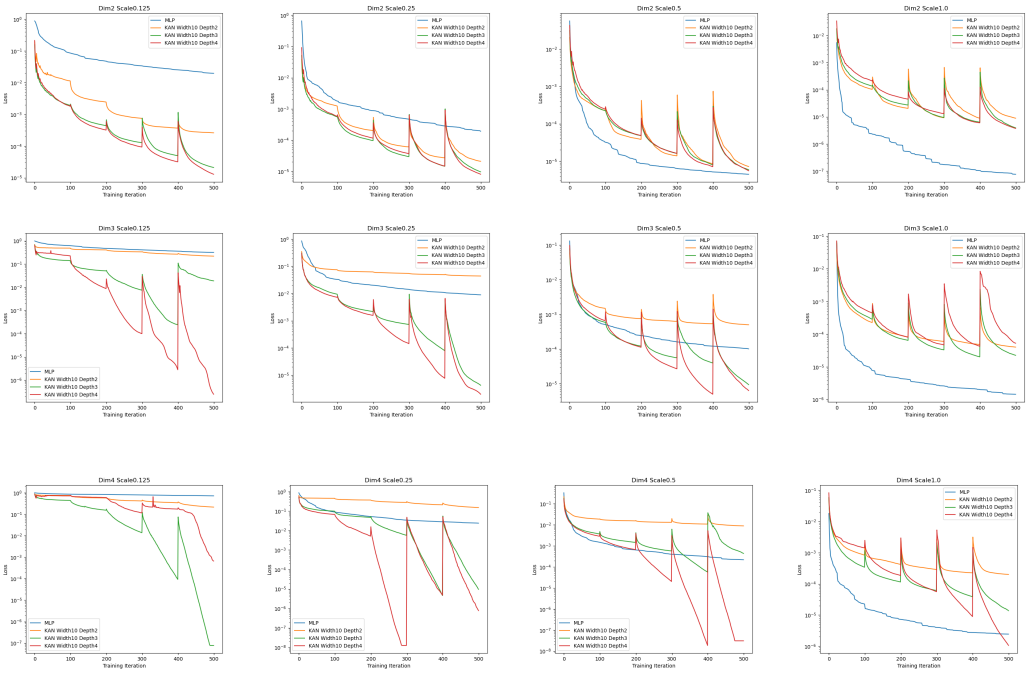


Figure 5.10: The Gaussian random field dataset. Training losses of MLP and KANs, with different scales and dimensions.

The different frequencies are normalized in a way that for $k > 1$, the ground truth has the same energy (H^1) norm. We use the variational form of the elliptic equation and the associated Deep Ritz Method [467]. Parametrizing u by a neural network, we minimize the loss

$$\lambda \int_{-1}^{1} \left(\frac{1}{2} u_x^2 - f u \right) dx + u^2(-1) + u^2(1).$$

For frequencies $k = 2, 4, 8, 16, 32$, we use 2000 uniformly spaced sample points and the neural network using an MLP of 6 layers and 256 neurons in each hidden layer and a KAN of 2 layers with 10 neurons in the hidden layer. We choose the hyperparameter $\lambda = 0.01$ balancing the energy and boundary loss and perform LBFGS iterations. For MLPs, we use 200 iterations, and for KANs, we use grid sizes $(20, 40)$, each trained with 100 iterations. We plot the relative L^2 and H^1 losses compared to the ground truth in Figure 5.12. We can see that KANs perform consistently better, and the residue barely deteriorates when the frequency increases, whereas it becomes extremely hard for MLPs to optimize when $k = 16, 32$.

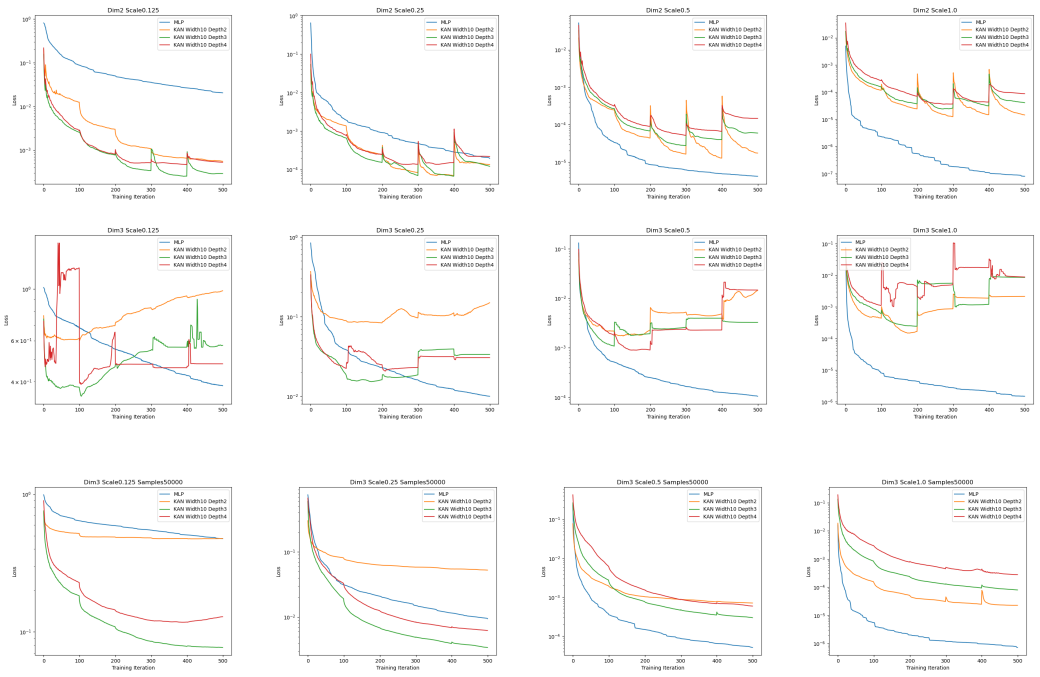


Figure 5.11: The Gaussian random field dataset. Test losses of MLP and KANs, with different scales and dimensions. Increasing the number of samples by 10x helps overfitting.

5.6 Conclusions and Discussions

Inspired by the Kolmogorov-Arnold representation theorem, we propose the Kolmogorov-Arnold Networks (KANs) as promising alternatives to MLPs. Our contributions are three-fold: (1) we put the KA theorem in the perspective of modern machine learning, relating to MLPs, and generalize the representation from two-layer to multiple layers via the KAN layers introduced, greatly enhancing expressive power. (2) we show that KANs are interpretable, serving as a useful tool for scientific discoveries. (3) we show that KANs are accurate and have nice scaling laws via theory and experiments. The major limitation of this work, however, is that our numerical examples focus on various aspects of science and are relatively small-scale. The scalability and extensibility of KANs for large-scale machine-learning tasks are left as future work. Especially we want to highlight the potential of applications of KANs to AI for Science tasks, since KANs can extract interpretable information from data to provide scientific insights, and reversely from scientific prior to build better models [292].

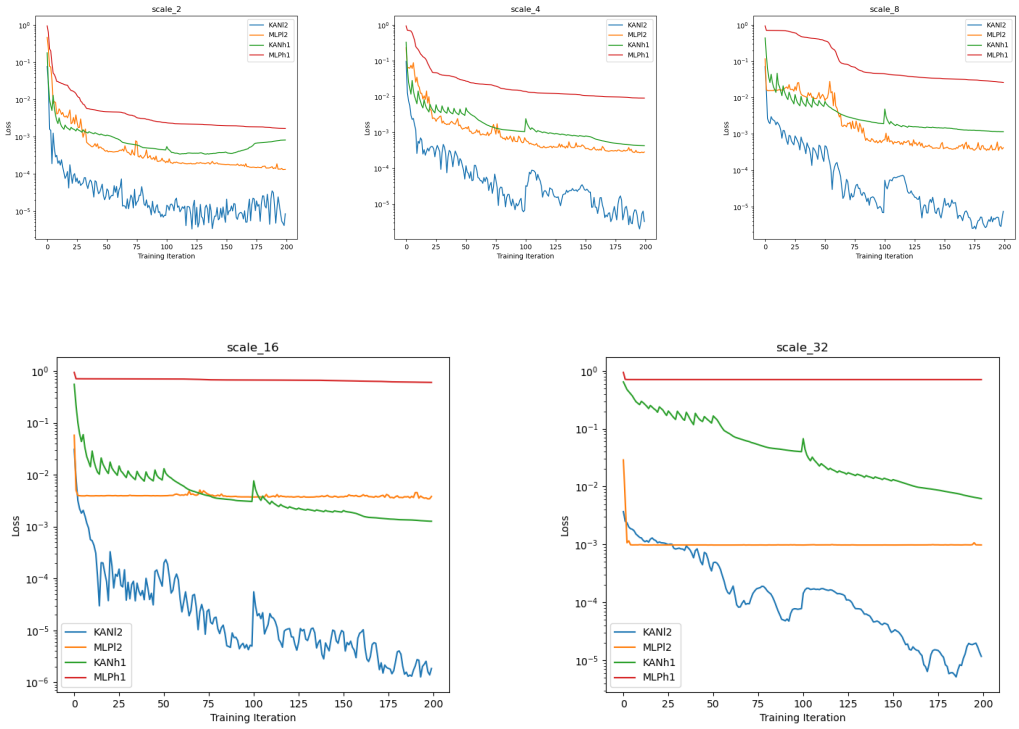


Figure 5.12: Solving PDEs. L^2 and H^1 losses of MLP and KAN with different frequencies of the solution.

5.6.1 Related works

Kolmogorov-Arnold theorem and neural networks. The connection between the Kolmogorov-Arnold theorem (KAT) and neural networks is not new in the literature [374, 403, 422, 255, 283, 261, 266, 145, 230, 377], but the pathological behavior of inner functions makes KAT appear unpromising in practice [374]. Most of these prior works stick to the original 2-layer width-$(2n + 1)$ networks, which were limited in expressive power and many of them are even predating back-propagation. Therefore, most studies were built on theories with rather limited or artificial toy experiments. More broadly speaking, KANs are also somewhat related to generalized additive models (GAMs) [4], graph neural networks [471] and kernel machines [419]. The connections are intriguing and fundamental but might be out of the scope of the current work.

Our contribution lies in generalizing the Kolmogorov network to arbitrary widths and depths, revitalizing and contexualizing them in today's deep learning stream, as well as highlighting its potential role as a foundation model for AI + Science.

There are also subsequent works exploring other parametrizations of activation

functions in the KAN formulation, including special polynomials [5, 406, 423], rational functions [6], radial basis function [276, 426], Fourier series [457], and wavelets [40, 407]. Active follow-up research focuses on applying KANs to various domains, such as partial differential equations [447, 411, 390] and operator learning [2, 411, 342], graphs [44, 112, 247, 473], time series [438, 160, 458, 159], computer vision [86, 10, 269, 85, 407, 35], and various scientific problems [287, 288, 461, 194, 259, 272, 7, 285, 366, 378].

Neural Scaling Laws (NSLs). NSLs are the phenomena where test losses behave as power laws against model size, data, compute etc [241, 192, 176, 200, 408, 19, 327, 420]. The origin of NSLs still remains mysterious, but competitive theories include intrinsic dimensionality [241], quantization of tasks [327], resource theory [420], random features [19], compositional sparsity [374], and maximum arity [326]. This work contributes to this space by showing that a high-dimensional function can surprisingly scale as a 1D function (which is the best possible bound one can hope for) if it has a smooth Kolmogorov-Arnold representation. Our work brings fresh optimism to neural scaling laws. We have shown in our experiments that this fast neural scaling law can be achieved on synthetic datasets, but future research is required to address the question whether this fast scaling is achievable for more complicated tasks (e.g., language modeling): Do KA representations exist for general tasks? If so, does our training find these representations in practice?

Mechanistic Interpretability (MI). MI is an emerging field that aims to mechanistically understand the inner workings of neural networks [354, 318, 441, 141, 339, 477, 290, 140, 107]. MI research can be roughly divided into passive and active MI research. Most MI research is passive in focusing on understanding existing neural networks trained with standard methods. Active MI research attempts to achieve interpretability by designing intrinsically interpretable architectures or developing training methods to explicitly encourage interpretability [290, 140]. Our work lies in the second category, where the model and training method are by design interpretable.

Learnable activations. The idea of learnable activations in neural networks is not new in machine learning. Trainable activations functions are learned in a differentiable way [177, 145, 384, 474] or searched in a discrete way [33]. Activation function are parametrized as polynomials [177], splines [145, 36, 11], sigmoid linear unit [384], or neural networks [474]. KANs use B-splines to parametrize their activation functions.

Symbolic Regression. There are many off-the-shelf symbolic regression methods based on genetic algorithms (Eureka [127], GPLearn [178], PySR [104]), neural-network based methods (EQL [309], OccamNet [128]), physics-inspired method (AI Feynman [435, 436]), and reinforcement learning-based methods [335]. KANs are most similar to neural network-based methods, but differ from previous works in that our activation functions are continuously learned before symbolic snapping rather than manually fixed [127, 128].

Physics-Informed Neural Networks (PINNs) and Physics-Informed Neural Operators (PINOs). In Section 5.4 PDE, we demonstrate that KANs can replace the paradigm of using MLPs for imposing PDE loss when solving PDEs. We refer to Deep Ritz Method [467], PINNs [382, 244] for PDE solving, and Fourier Neural operator [277], PINOs [280, 257, 313], DeepONet [296] for operator learning methods learning the solution map. There is potential to replace MLPs with KANs in all the aforementioned networks.

AI for Mathematics. AI has recently been applied to several problems in Knot theory, including detecting whether a knot is the unknot [181, 245] or a ribbon knot [182], and predicting knot invariants and uncovering relations among them [226, 105, 106, 110]. For a summary of data science applications to datasets in mathematics and theoretical physics see e.g. [396, 191], and for ideas how to obtain rigorous results from ML techniques in these fields, see [180].

5.7 Appendix

5.7.1 Implementation details of KAN

Implementation details. Although a KAN layer Eq. (5.2.5) looks extremely simple, it is non-trivial to make it well optimizable. The key tricks are:

(1) Residual activation functions. We include a basis function $b(x)$ (similar to residual connections) such that the activation function $\phi(x)$ is the sum of the basis function $b(x)$ and the spline function:

$$\phi(x) = w_b b(x) + w_s \text{spline}(x). \tag{5.7.1}$$

We set

$$b(x) = \text{silu}(x) = x/(1 + e^{-x}) \tag{5.7.2}$$

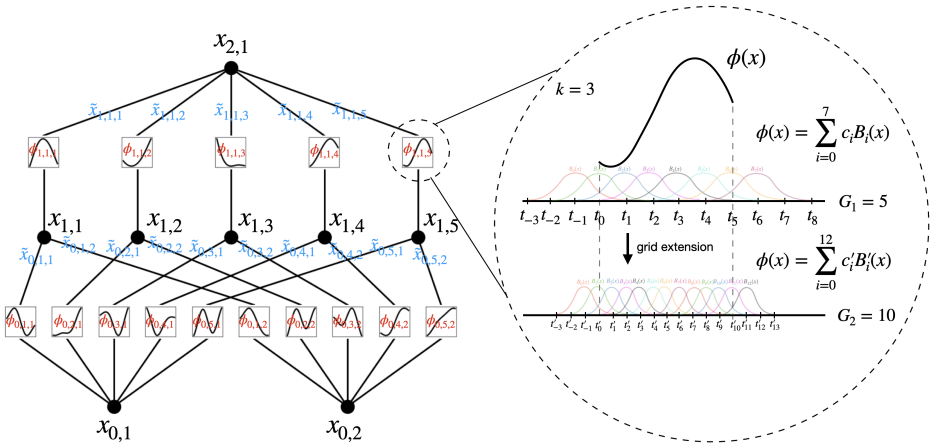


Figure 5.13: Left: Notations of activations that flow through the network. Right: an activation function is parameterized as a B-spline, which allows switching between coarse-grained and fine-grained grids.

in most cases. spline(x) is parametrized as a linear combination of B-splines such that

$$\text{spline}(x) = \sum_i c_i B_i(x) \tag{5.7.3}$$

where c_is are trainable (see Figure 5.13 for an illustration). In principle w_b and w_s are redundant since it can be absorbed into $b(x)$ and spline(x). However, we still include these factors (which are by default trainable) to better control the overall magnitude of the activation function.

(2) Initialization scales. Each activation function is initialized to have $w_s = 1$ and spline$(x) \approx 0$ [2]. w_b is initialized according to the Xavier initialization, which has been used to initialize linear layers in MLPs.

(3) Update of spline grids. We update each grid on the fly according to its input activations, to address the issue that splines are defined on bounded regions but activation values can evolve out of the fixed region during training [3] Grid updates (grid size $G_1 \to G_1$) use the same least square method as grid extensions (grid size $G_1 \to G_2 > G_1$).

[2] This is done by drawing B-spline coefficients $c_i \sim \mathcal{N}(0, \sigma^2)$ with a small σ, typically we set $\sigma = 0.1$.

[3] Other possibilities are: (a) the grid is learnable with gradient descent, e.g., [456]; (b) use normalization such that the input range is fixed. We tried (b) at first but its performance is inferior to our current approach.

Parameter count. For simplicity, let us assume a network

(1) of depth L,

(2) with layers of equal width $n_0 = n_1 = \cdots = n_L = N$,

(3) with each spline of order k (usually $k = 3$) on G intervals (for $G+1$ grid points).

Then there are in total $O(N^2L(G + k)) \sim O(N^2LG)$ parameters. In contrast, an MLP with depth L and width N only needs $O(N^2L)$ parameters, which appears to be more efficient than KAN. Fortunately, KANs usually require much smaller N than MLPs, which not only saves parameters, but also achieves better generalization (see e.g., Figure 5.5 and 5.7) and facilitates interpretability. We remark that for 1D problems, we can take $N = L = 1$ and the KAN network in our implementation is nothing but a spline approximation. For higher dimensions, we characterize the generalization behavior of KANs with a theorem below.

5.7.2 Proofs

Proof of Theorem 5.2.1. By the classical 1D B-spline theory [111] and the fact that $\Phi_{l,i,j}$ as continuous functions can be uniformly bounded on a bounded domain, we know that there exist finite-grid B-spline functions $\Phi^G_{l,i,j}$ such that for any $0 \le m \le k$,

$$\|(\Phi_{l,i,j}\circ\mathbf{\Phi}_{l-1}\circ\mathbf{\Phi}_{l-2}\circ\cdots\circ\mathbf{\Phi}_1\circ\mathbf{\Phi}_0)\mathbf{x}-(\Phi^G_{l,i,j}\circ\mathbf{\Phi}_{l-1}\circ\mathbf{\Phi}_{l-2}\circ\cdots\circ\mathbf{\Phi}_1\circ\mathbf{\Phi}_0)\mathbf{x}\|_{C^m} \le C_0G^{-k-1+m},$$

with a constant C_0 independent of G. We fix those B-spline approximations. Therefore we have that the residue R_l defined via

$$R_l \coloneqq (\mathbf{\Phi}^G_{L-1}\circ\cdots\circ\mathbf{\Phi}^G_{l+1}\circ\mathbf{\Phi}_l\circ\mathbf{\Phi}_{l-1}\circ\cdots\circ\mathbf{\Phi}_0)\mathbf{x}-(\mathbf{\Phi}^G_{L-1}\circ\cdots\circ\mathbf{\Phi}^G_{l+1}\circ\mathbf{\Phi}^G_l\circ\mathbf{\Phi}_{l-1}\circ\cdots\circ\mathbf{\Phi}_0)\mathbf{x}$$

satisfies

$$\|R_l\|_{C^m} \le C_1G^{-k-1+m},$$

with another constant independent of G. Finally notice that

$$f-(\mathbf{\Phi}^G_{L-1}\circ\mathbf{\Phi}^G_{L-2}\circ\cdots\circ\mathbf{\Phi}^G_1\circ\mathbf{\Phi}^G_0)\mathbf{x} = R_{L-1}+R_{L-2}+\cdots+R_1+R_0,$$

we know that (5.2.11) holds for another constant C independent of G. □

Remark: We can be more precise about the dependence of the constant C in the theorem. Define the compositionally smooth function class $C^{n,W,L,k}$ as the class of functions in the form of (5.2.10) such that the input dimension equals n, the width or $\max_{0\le i\le L} n_i$ in the definition (5.2.3) equals $W \ge n$, depth equals L, smoothness equals k. Then C only depends on W, L, k and $\max \|\phi_{l,i,j}\|_{C^m}$.

Proof of Theorem 5.2.2. We will show that each layer of an MLP with the activation function σ_k can be represented by a KAN with two hidden layers, width W and grid size $G = 2$ with degree k B-splines. By composing such layers, we obtain the desired result.

For a single layer of MLP, we consider the linear part and the non-linear activation separately. We first observe that on any compact subset of $\mathbb{R}^W$ the linear function

$$x_i = \sum_{j=0}^{W} a_{ij} x_j^{in} + b_i \tag{5.7.4}$$

can be represented with a single KAN layer of width W by setting ϕ_{ij} to the linear function

$$\phi_{i,j}(x) = a_{ij} x + \frac{b_i}{n}. \tag{5.7.5}$$

We claim that this linear function can be exactly represented on any interval $[-R, R]$ in the form (5.7.1). To do this, we first set $w_b = 0$ and choose the grid points for the B-splines to be

$$\{-(2k-1)R, -(2k-3)R, ..., -R, R, ..., (2k-3)R, (2k-1)R\}.$$

Note that based upon the KAN architecture, this corresponds to the extension of the uniform grid $t_0 = -R, t_1 = R$ which has grid size $G = 1$. It is also easy to verify that there are $(k+1)$ B-splines supported on this grid, whose restriction to $[-R, R]$ span the space of polynomials of degree k. Thus, in particular, any linear function on $[-R, R]$ can be represented as a linear combination of these B-splines.

Next, we consider the non-linear activation, which is given by the coordinatewise application of σ_k, i.e.

$$x_i^{out} = \sigma(x_i). \tag{5.7.6}$$

This can be represented by a single hidden layer KAN by setting

$$\phi_{i,j}(x) = \begin{cases} \sigma_k(x) & i = j \\ 0 & i \neq j. \end{cases} \tag{5.7.7}$$

We claim that the functions σ_k can be represented in the form (5.7.1) on any finite interval $[-R, R]$. To to this, we again set $w_b = 0$ and choose the grid points for the B-splines to be

$$\{-kR, -(k-1)R, ..., -R, 0, R, ..., (k-2)R, (k-1)R\}.$$

This grid is the grid extension of the uniform grid $t_0 = -R, t_1 = 0, t_2 = R$ which has grid size $G = 2$. It is easy also to verify that there are $(k + 2)$ B-splines supported on this grid and that any piecewise polynomial on $[-R, R]$ with a single breakpoint at 0 which is C^{k-1} is a linear combination of these B-splines. Hence the function σ_k can be represented on $[-R, R]$ in the form (5.7.1) using this grid.

The proof is now completed by composing these layers and choosing R sufficiently large so that for any input $x \in \Omega$ (which is bounded) the inputs and outputs of every neuron in the original MLP lie in the interval $[-R, R]$. □

Proof of Theorem 5.5.1. We first observe from (5.5.3) that the matrix M is block diagonal with d' identical blocks. Denoting these $(G+k-1)d \times (G+k-1)d$ blocks by B, it thus suffices to prove that

$$\frac{\lambda_{(G+k-1)d}(B)}{\lambda_d(B)} \leq C. \tag{5.7.8}$$

To do this, we analyze the blocks B and note that they take the form

$$B = \begin{pmatrix} C & D & \cdots & D \\ D & C & \cdots & D \\ \vdots & \vdots & \ddots & \vdots \\ D & D & \cdots & C \end{pmatrix}. \tag{5.7.9}$$

Here the diagonal sub-blocks $C \in \mathbb{R}^{(G+k-1)\times(G+k-1)}$ are the Gram matrix of the one-dimensional B-spline basis, i.e.

$$C_{ij} = \int_0^1 B_i(x) B_j(x) dx, \tag{5.7.10}$$

and the off-diagonal sub-blocks $D \in \mathbb{R}^{(G+k-1)\times(G+k-1)}$ are rank one matrices

$$D = vv^T, \tag{5.7.11}$$

where the vector $v \in \mathbb{R}^{G+k-1}$ is given by

$$v_i = \int_0^1 B_i(x) dx. \tag{5.7.12}$$

It is well-known that the Gram matrix C is well-conditioned uniformly in G for a fixed k, i.e. $\lambda_{G+k-1}(C)/\lambda_1(C) \leq K$ for a fixed constant K depending only upon k. See for instance [119], Theorem 4.2 in Chapter 5, where it is shown that the L_2-norm of a spline and the properly scaled ℓ_2-norm of its B-spline coefficients are equivalent

up to a constant depending only on k. This is equivalent to the well-conditioning of the Gram matrix C.

In addition, we can easily verify using Jensen's inequality (or Cauchy-Schwartz) that $D \preceq C$. Indeed, letting $w \in \mathbb{R}^{G+k-1}$ we see that

$$w^T D w = \left(\int_0^1 f(x) dx \right)^2 \leq \int_0^1 f(x)^2 dx = w^T C w, \tag{5.7.13}$$

where the function $f(x) = \sum_{i=1}^{G+k-1} w_i B_i(x)$.

Let $\mathbf{1} \in \mathbb{R}^d$ be the vector of ones and note that

$$(v \otimes \mathbf{1})(v \otimes \mathbf{1})^T = \begin{pmatrix} D & D & \cdots & D \\ D & D & \cdots & D \\ \vdots & \vdots & \ddots & \vdots \\ D & D & \cdots & D \end{pmatrix} \tag{5.7.14}$$

so that $B - (v \otimes \mathbf{1})(v \otimes \mathbf{1})^T$ is a block diagonal matrix with diagonal blocks $C - D$. We proceed to upper bound the largest eigenvalue of B by

$$\begin{aligned} \lambda_{(G+k-1)d}(B) &= \max_{\|w\|=1} w^T B w \\ &= \max_{\|w\|=1} w^T \begin{pmatrix} C-D & 0 & \cdots & 0 \\ 0 & C-D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C-D \end{pmatrix} w + (w^T (v \otimes \mathbf{1}))^2. \end{aligned} \tag{5.7.15}$$

Writing $w = (w_1, ..., w_d)$ with $w_i \in \mathbb{R}^{G+k-1}$ and $\sum_{i=1}^{G+k-1} \|w_i\|^2 = 1$ and using that $D = vv^T$, we get the bound

$$\begin{aligned} \lambda_{(G+k-1)d}(B) &\leq \max_{\|w_1\|^2+\cdots+\|w_d\|^2=1} \sum_{i=1}^d w_i^T C w_i + \left(\sum_{i=1}^d v^T w_i \right)^2 - \sum_{i=1}^d (v^T w_i)^2 \\ &\leq \max_{\|w_1\|^2+\cdots+\|w_d\|^2=1} \sum_{i=1}^d w_i^T C w_i + (d-1) \sum_{i=1}^d (v^T w_i)^2 \end{aligned} \tag{5.7.16}$$

Since $D \preceq C$ we have $(v^T w_i)^2 \leq w_i^T C w_i$ which gives the bound

$$\lambda_{(G+k-1)d}(B) \leq d \max_{\|w_1\|^2+\cdots+\|w_d\|^2=1} \sum_{i=1}^d w_i^T C w_i = d \lambda_{G+k-1}(C). \tag{5.7.17}$$

Next, we lower bound the d-th eigenvalue of B. For this, we use the Courant-Fisher minimax theorem to see that

$$\lambda_d(B) = \max_{W_d} \min_{w \in W_d,\ \|w\|=1} w^T B w, \tag{5.7.18}$$

where the maximum is taken over all subspaces of W_d of codimension $< d$. We consider the specific subspace

$$W_d = \{(w_1, ..., w_d);\ v^T w_i = 0 \text{ for all } i = 1, ..., d\} \oplus \text{span}(v \otimes \mathbf{1}) \tag{5.7.19}$$

and observe that for any $(w_1, ..., w_d)$ with $v^T w_i = 0$ we have

$$w^T B w = \sum_{i=1}^{d} w_i^T C w_i \geq \lambda_1(C) \sum_{i=1}^{d} \|w_i\|^2 = \lambda_1(C)\|w\|^2, \tag{5.7.20}$$

while for the $w = v \otimes \mathbf{1}$ (which is orthogonal) we have

$$w^T B w = \sum_{i=1}^{d} v^T C v + (d-1) \sum_{i=1}^{d} \|v\|^2 \geq \lambda_1(C) d \|v\|^2 = \lambda_1(C)\|w\|^2. \tag{5.7.21}$$

Thus, $\lambda_d(B) \geq \lambda_1(W)$. Combining these bounds and using the well-conditioning of the Gram matrix C, we get

$$\frac{\lambda_{(G+k-1)d}(B)}{\lambda_d(B)} \leq d \frac{\lambda_{G+k-1}(C)}{\lambda_1(C)} \leq K \tag{5.7.22}$$

for a constant K which only depends upon k. This completes the proof. □

Chapter 6

HIGH PRECISION PINNS IN UNBOUNDED DOMAINS: APPLICATION TO SINGULARITY FORMATION IN PDES

In this chapter, we investigate the high-precision training of Physics-Informed Neural Networks (PINNs) in unbounded domains, with a special focus on applications to singularity formation in PDEs. We propose a modularized approach and study the choices of neural network ansatz, sampling strategy, and optimization algorithm. When combined with rigorous computer-assisted proofs and PDE analysis, the numerical solutions identified by PINNs, provided they are of high precision, can serve as a powerful tool for studying singularities in PDEs. For 1D Burgers equation, our framework can lead to a solution with very high precision, and for the 2D Boussinesq equation, which is directly related to the singularity formation in 3D Euler and Navier-Stokes equations, we obtain a solution whose loss is 4 digits smaller than that obtained in [449] with fewer training steps. We also discuss potential directions for pushing towards machine precision for higher-dimensional problems.

6.1 Introduction

singularity formation is one of the key challenges in the study of partial differential equations (PDEs). Unlike well-posed equations, where one can apply classical existence and uniqueness theorems, singularities often occur in certain solutions of nonlinear PDEs, where we only have guarantees of existence for a short time, but the solution may blow up in finite time. The study of singularities often involves a case-by-case approach and is related to some of the most intriguing mathematics and physical properties, such as the onset of turbulence in the Navier-Stokes equations. The singularity of the Navier-Stokes equations is one of the seven Millennium Prize Problems [146]: widely regarded as the most fundamental and challenging problem in analysis, and is still open. One of the key difficulties in the study of singularities is the lack of understanding of the singularity pattern and its mechanism. The computation of the singularity itself, or infinity, is intractable numerically. A general roadmap is thus to first propose a plausible singularity ansatz that renders the computation feasible, then find candidates of such blowup by numerical simulations, and finally verify the stability of such an ansatz by PDE analysis.

We are often interested in a special structure of singularity: self-similar singularity. Self-similarity relates to the invariance of the solution under scaling transformations and reduces singularity to the existence of a self-similar profile. To be precise, for the quantity of interest $u(x,t)$, we put the ansatz $u(x,t) = (T-t)^{-\alpha}U(x(T-t)^{-\beta})$, where U is the profile function independent of time, T is the blowup time, and $\alpha > 0, \beta$ are the scaling exponents to be determined. Now we can reduce the computation of an infinite u to the computation of a finite, smooth profile U, along with scaling exponents α, β to be inferred. Physics-Informed Neural Networks (PINNs) [382] serve as a powerful tool to find such profiles, with the scaling parameters jointly inferred as inverse problems. It was first introduced in the context of identifying singularity profiles in [449] and has seen success even in problems that are unstable for traditional numerical methods. While PINNs offer a powerful tool to search for candidate profiles, solutions identified by PINNs are often of limited accuracy and far from applicable to rigorous PDE analysis, to the best of the authors' knowledge.

In this work, we aim to systematically study the high-precision training of PINNs, with a special focus on applications to solve profile equations governing singularity formations in PDEs, in fluid dynamics and beyond. We will take a modularized perspective without diving into sophisticated tricks and investigate the following aspects: a good neural network ansatz representing the profile function, a good sampling strategy to tackle the infinite domain with a special focus on imposing boundary conditions, and a good optimization algorithm to train the neural network. We apply our findings to the 1D Burgers equation and the 2D Boussinesq equation, obtaining a precision amenable to rigorous PDE analysis for the 1D Burgers equation and 4 digits better than [449] with fewer training steps for the 2D Boussinesq equation. The 2D Boussinesq equation shares many similarities with the 3D axisymmetric Euler equation for the ideal fluid without viscosity; see the pioneering works of [76, 73, 74] for the connection between the two equations, where the authors used the connection to establish singularity formation for the 3D axisymmetric Euler with boundary.

6.2 Related Works

6.2.1 PINNs

Neural networks have witnessed success in solving PDEs and surrogate modeling in math and science. PINNs in particular have been widely used due to their flexibility and applicability to a wide range of problems [54, 382, 244]. The key idea of PINNs is to enforce the PDE constraints at a set of collocation points and to minimize the

residual of the PDEs as a loss function. By posing the solution of PDEs as an optimization problem, PINNs are especially suited to solve inverse problems [383, 469, 297, 468], where the solution of the PDE and the underlying parameters can be jointly inferred. In [449], the authors used PINNs to study the blowup of the 1D Burgers equation, the 1D family of generalized Constantin-Lax Majda equations, and the 2D Boussinesq equation.

Another line of work, operator learning [257], focuses on learning the solution operator instead of learning a single instance of solution, where Fourier Neural Operators (FNOs) [277, 278, 279, 281] and DeepONets [296] are two families of representative works in this direction. Once a solution operator is learned, it can be evaluated in a resolution-free manner at any point in the domain. Data are often augmented to enhance the solution accuracy, while the loss function can also incorporate the PDE constraints, termed Physics Informed Neural Operator (PINO) [280]. In [313], the authors used PINO with Fourier continuation to study the blowup of the 1D Burgers equation.

6.2.2 Self-similar singularity and computer assisted proofs

Self-similar singularity of the ansatz $u(x,t) = (T-t)^{-\alpha}U(x(T-t)^{-\beta})$ is generic in the study of singularity formation in PDEs, where one uses the scaling invariance of the PDE and can reduce the computation of an infinite u to the computation of a finite, smooth (approximate) profile U. Such structures exist even in the simple Riccati ODE $u_t = u^2$ with an exact solution $u = (T-t)^{-1}$.

The approximate profile can be identified via explicit construction or numerical computation. Working in the rescaled, self-similar variables and performing stability analysis around the profile U provides a powerful tool to establish the singularity formation, for nonlinear Schrodinger equations [315, 321], incompressible fluids [136, 72, 76, 73, 219], compressible fluids [320, 319], and beyond. Until recently, most of the works relied on an explicit profile and spectral information of the associated linearized operator to establish linear and nonlinear stability. In [73, 74], the authors used computer-assisted proofs with a sophisticated numerical profile obtained by solving the dynamic rescaling equations in time to obtain an approximate steady state. By analyzing the stability of the approximate profile, they established the singularity formation for the 2D Boussinesq equation and the 3D axisymmetric Euler equation with boundary. And in [218, 77], the authors provided a framework using only local information for stability analysis, bypassing spectral information

and allowing for numerical profiles with computer-assisted proofs, for problems beyond self-similarity.

6.2.3 Towards high precision training

Various methods have been proposed in the literature to improve the accuracy of PINNs. One line of work focuses on a better representation of the solution. In [327, 448], the boosting technique was proposed, where a sum of a sequence of neural networks with decreasing magnitude was used to learn the solution; at each stage, a new neural network is trained to learn the residual. To overcome the spectral bias [381] of multilayer perceptrons (MLPs), or the favor of learning low-frequency modes [459, 460] in the solution, one can use Fourier feature encoding [415, 429, 345], or different activation functions [233, 232, 205, 475, 446]. In particular, Kolmogorov-Arnold Networks (KANs) [292, 293] that leverage nonlinear learnable activation functions and the Kolmogorov-Arnold representation theorem were proposed and further investigated in the PINN setting [447, 411, 432]. Another line of work improves the optimization landscape during the training of PINNs. Various optimizers, which we will detail in Subsection 6.3.3, have been proposed to improve the convergence rate. Adaptive design of points sampling [8, 454, 390] and adaptive weighting of different terms [442, 455, 314] in the loss function were also proposed to improve the accuracy of PINNs.

We will only focus on applying hard constraints and choosing a good optimizer in this work and leave the exploration of more sophisticated tricks for future work.

6.3 Methodology

We outline our methodology of high-precision training for PINNs on the whole space in this section. We work under the general formulation of the profile equation

$$L(U, \lambda) = 0,$$

where $U(y)$ is the profile function, λ is a set of scaling parameters to be determined, and L is the nonlinear differential operator. For our problems of interest, U=0 will be a trivial solution satisfying the equation.

6.3.1 Infinite domain

The key challenges we are facing here are sampling and learning on an infinite domain. For a given budget of a finite number of sampling points, we need to sample the domain in a way that the resulting solution is accurate and generalizes

well throughout the domain. In the meantime, We want the neural network to be able to represent the profile function and the initialization of parameters to favor learning of such representations. To this end, we adopt an exponential "mesh" in our sampling strategy: Consider an auxiliary variable z such that $y = \sinh(z) = \frac{e^z - e^{-z}}{2}$, and sample z uniformly in a finite region. Here we choose the sinh transformation as in [449] to respect the parity of the functions, detailed in the subsequent subsection. Such a transformation maps roughly $z \in [-30, 30]$ to $y \in [-5 \times 10^{12}, 5 \times 10^{12}]$.

Boundary conditions are another important aspect when learning on the whole space. For our problems of interest, U by itself will not have sufficient decay at infinity, and one approach adopted in [449] is to impose Neumann boundary conditions at infinity, or numerically on the boundary of the domain of the z variables. To rule out the trivial solution $U = 0$, we need to enforce a nondegeneracy condition, often posed at the origin. We will discuss the enforcement of these conditions in the following subsection. We refer to this formulation as boundary conditions using **weak asymptotics**.

On the other hand, we can enforce stronger information on the boundary. If we know the exact asymptotic behavior of the solution at infinity as g, for example a power law, we can introduce a smooth cutoff function χ with $\chi(0) = 0, \chi(\infty) = 1$ and the ansatz $U = \tilde{U} + \chi g$. We can then enforce Dirchlet boundary conditions at infinity for $\tilde{U}$ represented by the neural network. We refer to this formulation as boundary conditions using **exact asymptotics**. We will demonstrate for the 1D example that PINNs using exact asymptotics will outperform those using weak asymptotics by a large margin.

A priori, the exact asymptotics information is not available, and one can first train a neural network U_w with boundary conditions using weak information, and distill the information of asymptotics g from U_w. We refer to this formulation as boundary conditions using **hybrid asymptotics**. For example, for the 2D Boussinesq equation, borrowing ideas from [73], one can use function fitting and symbolic regression to extract asymptotics g from U_w, filtering out the noisy residues, such that g is a symbolic function approximating U_w at infinity. We will leave this approach to future work.

6.3.2 Hard constraint

Hard constraints are important concepts in the parametrization of the solution space for PINNs. When enforced properly, they will guarantee physical properties of the

solution [297, 389, 332, 132], and can impose the solution to be in the correct manifold. While most of the previous works focus on hard constraints of boundary conditions, we emphasize the enforcement of hard constraints in the following senses: the parity of the learned function and the nondegeneracy conditions. Empirically we observe a better convergence rate and a more stable solution when enforcing hard constraints.

Parity. For a function $f(y_i, \hat{y}_i)$ even/odd in the variable y_i, we train a neural network with the following ansatz $f = (f_{nn}(y_i, \hat{y}_i) \pm f_{nn}(-y_i, \hat{y}_i))/2$.

Nondegeneracy conditions. As discussed in the previous subsection, we need to enforce nondegeneracy conditions to rule out the trivial solution $U = 0$ when using weak asymptotics. For example, for the 1D Burgers equation, we know that U is odd and necessarily $U'(0) = -1$; we can enforce $U'''(0) = 6$. We will enforce a hard constraint via Taylor expansion at the origin as $U = -z + z^3 + z^4 U_1$, for an odd function U_1. Similarly for the 2D Boussinesq equation, we enforce $\partial_1 \Omega(0, 0) = -1$ and Ω is odd in z_1 via a Taylor expansion as $\Omega = -z_1 + z_1 z_2 \Omega_1 + z_1^2 \Omega_2$, where Ω_1, Ω_2 are even and odd functions in z_1 respectively.

6.3.3 Optimizer: Self-Scaled BFGS methods

A common practice of training PINNs is to use the Adam optimizer. As a stochastic first-order method, Adam is known to be robust and efficient in training deep neural networks and can empirically escape local minima. To further improve convergence to the minimizer, one can apply second-order methods with a higher convergence rate, like L-BFGS, after training with Adam for a few epochs. While this seems to be a gold standard in the training of PINNs [387], various optimizers have been investigated, including variants of second-order quasi-Newton methods [387, 444], and optimizers using natural gradients [334, 238, 84]. We highlight and use the self-scaled BFGS methods proposed in [12, 13] and introduced to the PINNs context in [437, 250]. BFGS methods use an approximation of the inverse of the Hessian matrix to precondition the gradient for the update direction. To be precise, consider the parameters Θ_k and learning rate α_k at step k, with loss function $\mathcal{J}(\Theta)$, then the update rule for Θ is

$$\Theta_{k+1} = \Theta_k - \alpha_k H_k \nabla \mathcal{J}(\Theta_k).$$

Different choices of updating the approximate inverse Hessian H_k lead to different optimizers, and L-BFGS in particular is a memory-efficient way for the updates by

storing only vectors instead of the whole matrix. The self-scaled BFGS methods use a scaling compared to the standard BFGS update of the inverse Hessian. More precisely, for the auxiliary variables

$$s_k = \Theta_{k+1} - \Theta_k, \quad y_k = \mathcal{J}(\Theta_{k+1}) - \mathcal{J}(\Theta_k),$$
$$v_k = \sqrt{y_k \cdot H_k y_k} \left[\frac{s_k}{y_k \cdot s_k} - \frac{H_k y_k}{y_k \cdot H_k y_k} \right],$$

we have for the scalers τ_k and ϕ_k:

$$H_{k+1} = \frac{1}{\tau_k} \left[H_k - \frac{H_k y_k \otimes H_k y_k}{y_k \cdot H_k y_k} + \phi_k v_k \otimes v_k \right] + \frac{s_k \otimes s_k}{y_k \cdot s_k},$$

where the original BFGS corresponds to the choices $\tau_k = \phi_k = 1$. While this is only a simple modification of the original BFGS, the authors in [437] demonstrated a much improved convergence rate across a variety of benchmarks, including the Helmholtz equation, the nonlinear Poisson equation, the nonlinear Schrödinger equation, the Korteweg-De Vries equation, the viscous Burgers equation, the Allen-Cahn equation, 3D Navier-Stokes: Beltrami flow, and the lid-driven cavity. We use the self-scaled Broyden methods proposed in [437]; see equations (13)-(23) therein for details on the choices of τ_k and ϕ_k.

On the role of minibatch training or random resampling. One of the common practices when training PINNs is to use random resampling of the collocation points. This can enhance the performance of SGD-based methods like Adam empirically. However, full-batch second-order methods like BFGS with supposedly higher-order accuracy do not adapt well to random resampling since they rely on past trajectories for Hessian updates. One empirical observation, as proposed in [448], is that when one uses an optimizer with fixed resolution like BFGS, it will be able to generalize in the regions where sampling points are sufficient. However, in the undersampled regions, the learned solution generalizes poorly. In an abstract form, there exists a critical batchsize N_c, such that when $N > N_c$, fixed sampling will be preferred, while for $N < N_c$, fixed sampling will have very bad generalization. N_c would depend on both the equation and the scale of the neural network. Empirically, we observe that roughly 10k points are sufficient for generalization with fixed training points. We resample every 1000 epochs to further reduce overfitting. See details on the choices of the batch size in the experiments section.

6.4 Experiments

In this section, we describe our numerical experiments on the blowup profiles for 1D Burgers equation and 2D Boussinesq Equation. The codes are available at https://github.com/RoyWangyx/High-precision-PINNs-unbounded-domains-/tree/main. When training both equations, we denote the PDE by $L(U(y)) = 0$ and the boundary condition by $B(U) = 0$. We use auxiliary variables $z = \sinh^{-1} y$ as in Subsection 6.3.1 and consider the following combination of interior, boundary, and smoothness losses as in [449]

$$\begin{aligned} loss &= 0.1(L_i + L_s) + L_b \\ &= 0.1\Big(\frac{1}{N_i}\sum_{j=1}^{N_i}[\hat{L}(U_{nn}(z_j))]^2 + \frac{1}{N_s}\sum_{j=1}^{N_s}|\nabla_{z_j}\hat{L}(U_{nn}(z_j))|^2\Big) + \frac{1}{N_b}\sum_{j=1}^{N_b}[\hat{B}(U_{nn})]^2, \end{aligned} \tag{6.4.1}$$

where $U_{nn}(z)$ is supposed to approximate $\hat{U}(z) = U(y)$ in the z-variables, and $\hat{L}$, $\hat{B}$ denotes the PDE and the boundary condition transformed in the z-variables; see [449] for a concrete formula for the 2D Boussinesq equation.

6.4.1 Burgers equation

For the 1D Burgers equation

$$u_t + uu_x = 0, \tag{6.4.2}$$

consider the self-similar ansatz that respects the scaling symmetry

$$u(x,t) = (1-t)^{\lambda}U(y), \quad y = x(1-t)^{-1-\lambda}. \tag{6.4.3}$$

The profile equation for U used for the PDE loss in (6.4.1) is

$$-\lambda U + ((1+\lambda)y + U)U_y = 0. \tag{6.4.4}$$

We impose an odd symmetry on U, and the profile equation has implicit solutions

$$y + U + CU^{1+1/\lambda} = 0. \tag{6.4.5}$$

for any constant C, as in the setting of [449], where we know that the most stable solutions correspond to $\lambda = 0.5$ and there are nonsmooth solutions at e.g. $\lambda = 0.4$.

In this example, we assume that we first train the neural network on a bounded domain and infer the correct λ already, for example via the method in [449]. Now we focus on fixing λ and learn U on the unbounded domain. Using an MLP with activation function tanh, 4 layers and 20 neurons per layer and a hard constraint

on parity, we use the optimizer SSBroyden1 as in [437] with 20000 epochs and resampling every 1000 epochs. z is sampled uniformly on $[0, 30]$ with a batchsize 10000 for both the interior and smoothness losses, corresponding to a domain $[0, 5 \times 10^{12}]$ in the y variables.

For the formulation using weak asymptotics as in Subsection 6.3.1, we use the Neumann boundary condition $U_y = 0$ and enforce hard constraint of nondegeneracy conditions as in Subsection 6.3.2. For the formulation using exact asymptotics as in Subsection 6.3.1, we use Dirichlet boundary condition $\tilde{U} = 0$ and the cutoff function $\chi = (\frac{y}{1+y})^{15}$, since the far field is captured by the exact asymptotics $g = -y^{\frac{\lambda}{1+\lambda}}$.

We present the following results of $\lambda = 0.4, 0.5$ using weak and exact asymptotics: see Figure 6.1 for the equation residue of the solution at the final stage and Figure 6.2 for the evolution of the losses. We are able to achieve high accuracy over a large domain, but using exact asymptotics is preferred for both the smooth and nonsmooth case of λ.

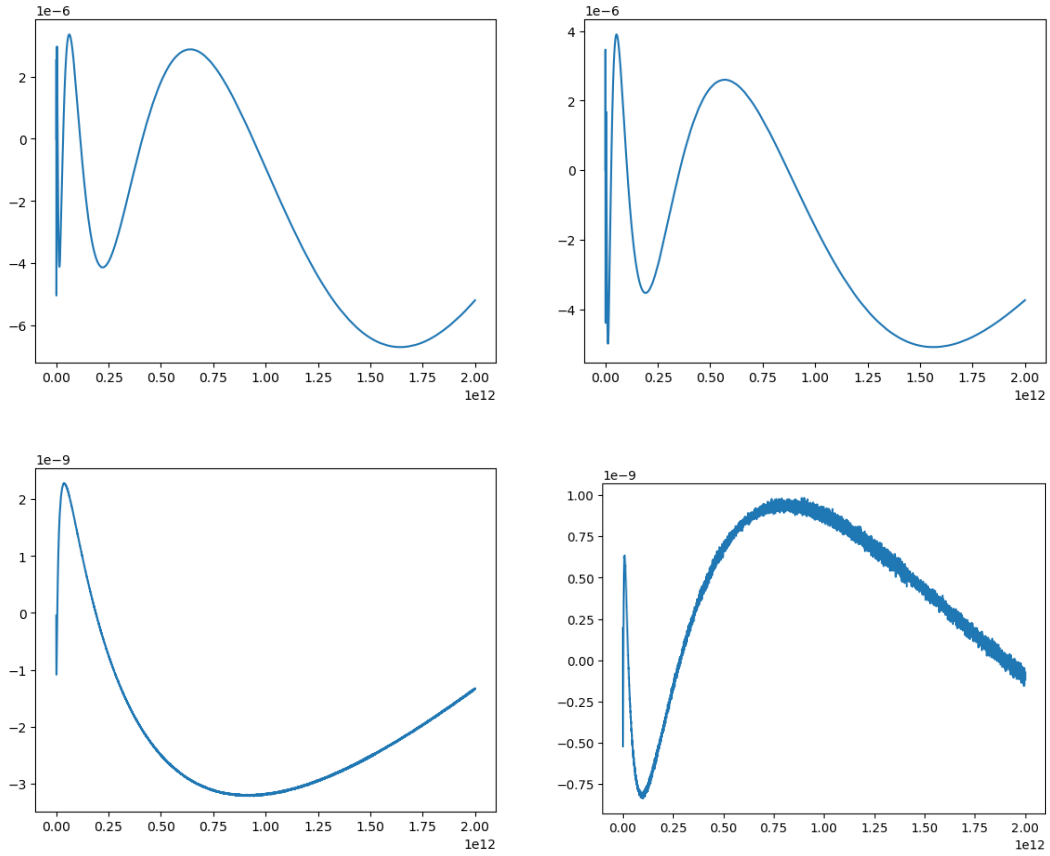


Figure 6.1: Final residue in a large domain for 1D Burgers. Upper: weak asymptotics; down: exact asymptotics; left: $\lambda = 0.4$; right: $\lambda = 0.5$.

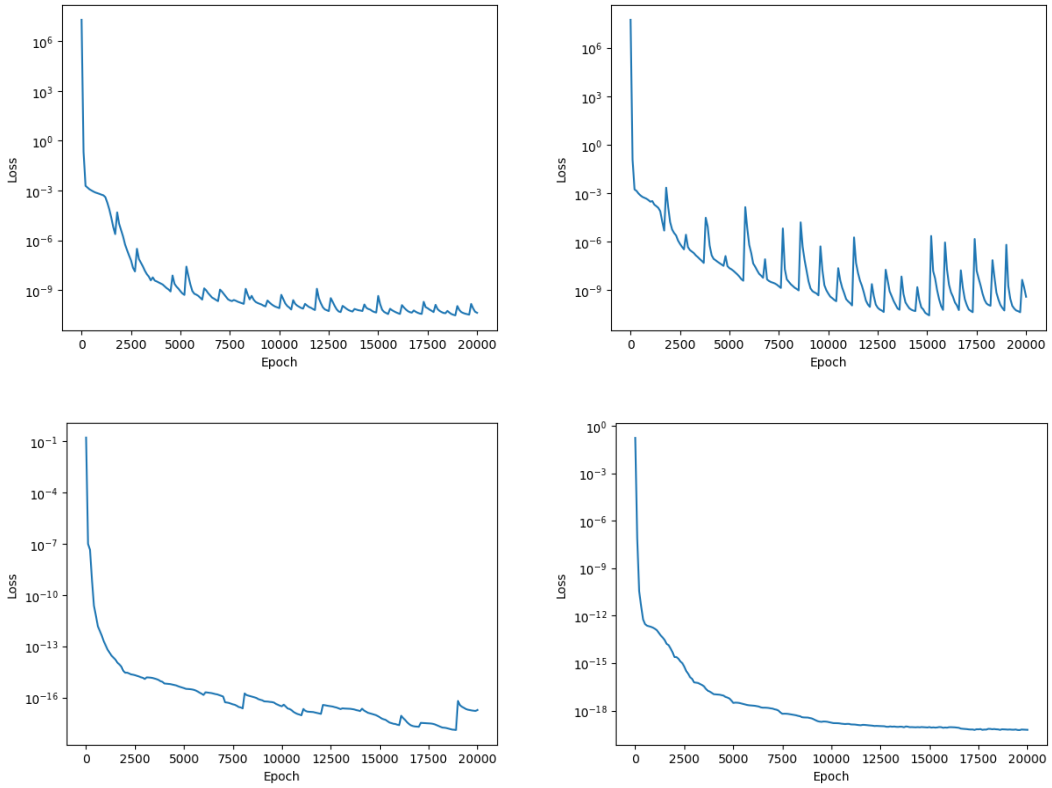


Figure 6.2: Trajectory of losses for 1D Burgers. Upper: weak asymptotics; down: exact asymptotics; left: $\lambda = 0.4$; right: $\lambda = 0.5$.

6.4.2 Boussinesq equation

For the 2D Boussinesq equation on the half plane, in vorticity form with the self-similar ansatz, we get the following profile equations for $(\Omega, U_1, U_2, \Phi, \Psi)$ as in [449]:

$$\begin{aligned}
\Omega + ((1+\lambda)(y_1, y_2)^T + (U_1, U_2)^T) \cdot \nabla\Omega &= \Phi, \\
(2 + \partial_{y_1} U_1)\Phi + ((1+\lambda)(y_1, y_2)^T + (U_1, U_2)^T) \cdot \nabla\Phi &= -\partial_{y_1} U_2 \Psi, \\
(2 + \partial_{y_2} U_2)\Psi + ((1+\lambda)(y_1, y_2)^T + (U_1, U_2)^T) \cdot \nabla\Psi &= -\partial_{y_2} U_1 \Phi, \\
\partial_{y_1} U_1 + \partial_{y_2} U_2 = 0, \quad \Omega = \partial_{y_1} U_2 - \partial_{y_2} U_1, \quad \partial_{y_1} \Psi &= \partial_{y_2} \Phi,
\end{aligned}$$

where (Ω, U_1, Φ) are odd and (U_2, Ψ) are even in y_1 and we are in the half plane $y_2 \geq 0$.

For the boundary conditions, we impose a non-penetration boundary condition $U_2(y_1, 0) = 0$ along with decaying weak asymptotics at the far field, with Dirichlet boundary conditions $\Phi = \Psi = 0$ and Neumann boundary conditions for the velocity field $\nabla(U_1, U_2)^T = 0$. For the nondegeneracy condition, we impose $\partial_{y_1}\Omega(0,0) = -1$

and use Taylor expansion to enforce a hard constraint as in Subsection 6.3.2. We find that enforcing a hard constraint is much more effective to avoid converging to a trivial solution than enforcing soft constraints.

For each function, we use a 7-layer MLP with width 30, hard constraints on parity, and activation function SiLU = $\frac{x}{1+e^{-x}}$ to better model the growth at the far field. For sampling, we sample 1000 points on each boundary of the square $(z_1, z_2) \in [0, 30]^2$, and 5000 points each for the interior and smoothness losses, where we sample (z_1, z_2) with equal probability uniformly on $[0, 30]^2$ and $[0, 5]^2$ for the interior loss and with equal probability uniformly on $[0, 3]^2$ and $[0, 0.5]^2$ for the smoothness loss, ensuring smoothness near the origin. Again, we are computing effectively in a large domain $[0, 5 \times 10^{12}]^2$ in the y variables.

For optimization, we use Adam for 10000 epochs with resampling, followed by the optimizer SSBroyden 1 as in [437] with 40000 epochs and resampling every 1000 epochs. The learning rate of Adam is set to be 0.001 for the functions and 0.1 with $\beta = (0.9, 0.9)$ for λ, with a decay of 0.9 after 5000 epochs.

We present the final profiles and the residue of the PDEs near the origin respectively in Figure 6.3 and 6.4, and the evolution of losses in Figure 6.5. Compared to [449], we achieve a training loss of 4 digits smaller and equation residues of 2 digits smaller. We remark that due to computational constraints and the cost of a full batch optimizer involving the approximation of the Hessian matrix, this is the largest neural network affordable. We use 10 days of CPU time on a MacPro 2019 with 2.5GHz 28-core Intel Xeon W processor.

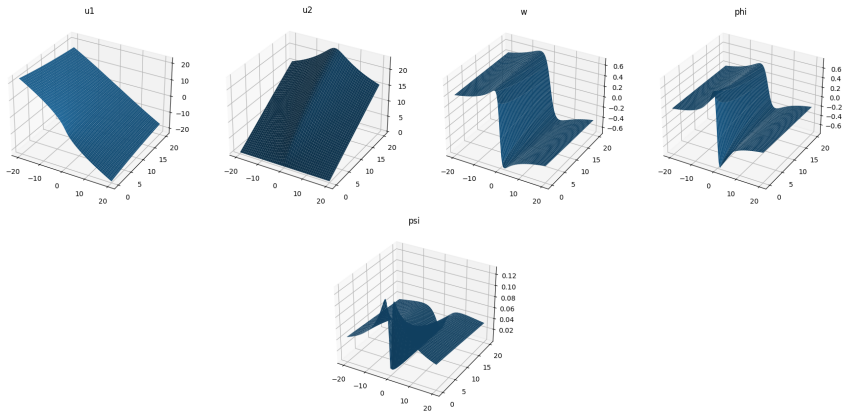


Figure 6.3: Final profiles for 2D Boussinesq

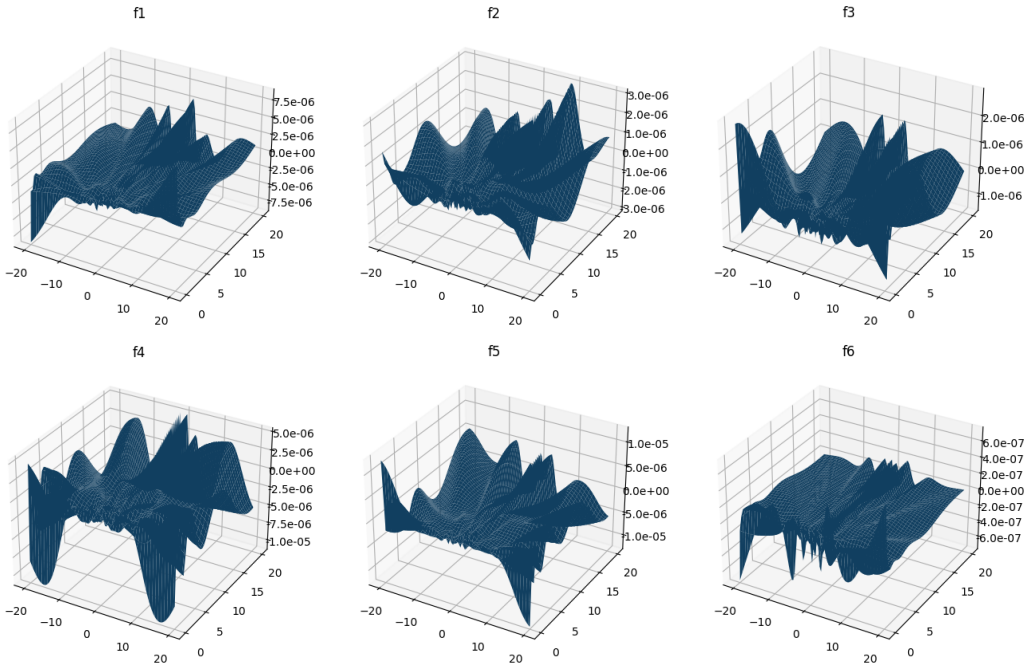


Figure 6.4: Final equation residues for 2D Boussinesq

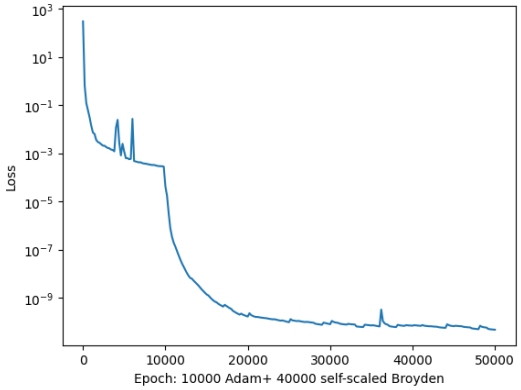


Figure 6.5: Trajectory of losses for 2D Boussinesq: 10000 Adam iterations followed by 40000 self-scaled Broyden iterations.

6.5 Conclusions and Future Work

We demonstrate the importance of enforcing appropriate asymptotics, enforcing hard constraints, and adopting a better optimizer for solving PDEs using neural networks on an infinite domain. We achieve better accuracy for problems crucial to the study of singularity formations. As a future direction, we believe a better enforcement of far-field asymptotics, formulated as hybrid asymptotics, might have the potential of driving PDE residues to machine precision, potentially amenable to rigorous computer-assisted proofs using the profiles identified by the neural networks. Another direction is to use PINO, the idea of operator learning on a range of scaling parameters, to learn a collection of profiles with different scalings.

6.6 On Weak Convection Model to 3D Euler and Numerical Stability Analysis

A natural next step of the present program is to study the weak-convection regime, namely the dynamic-rescaling system in which the transport is weakened to encourage blowup; as in [289].

$$\partial_t u_1 + \varepsilon\, u^r\, \partial_r u_1 + \varepsilon\, u^z\, \partial_z u_1 = 2u_1\, \partial_z \psi_1 \tag{6.6.1}$$

$$\partial_t \omega_1 + \varepsilon\, u^r\, \partial_r \omega_1 + \varepsilon\, u^z\, \partial_z \omega_1 = 2u_1\, \partial_z (u_1) \tag{6.6.2}$$

$$-\left[\partial_r^2 + \frac{3}{r}\partial_r + \partial_z^2\right]\psi_1 = \omega_1. \tag{6.6.3}$$

In this setting, the profile equations retain the main hyperbolic scaling structure while the nonlinear convection terms enter with a small prefactor ε, which makes the model both analytically tractable and still rich enough to capture singular behavior. Our computations and formal asymptotics indicate that this weak-convection model admits a nontrivial singular profile: after rescaling, one finds a coherent stationary state (U, Ω, Ψ) together with scaling parameters $(\lambda, C, c_u, c_\omega)$ for which the original solution concentrates and the relevant amplitude grows according to the self-similar law. In this sense, the model provides clear evidence of singularity formation, while also preserving enough of the geometric structure of the full equation to serve as an informative intermediate problem between purely leading-order toy models and the fully coupled dynamics. We are able to push to a higher $\varepsilon = 0.3$ compared to the reported in [289], using neural networks.

An important direction for future work, in collaboration with Prof. Anima's group, is to complement the existence and numerical construction of such singular profiles with a stability theory. The main difficulty is that the correct coercive functional is not known a priori: because the singular profile is strongly anisotropic and exhibits

nontrivial growth/decay across spatial regions, standard polynomial or isotropic Sobolev weights are unlikely to capture the true dissipative mechanism of the linearized operator. To address this, we propose to parametrize a family of singular weights by a neural network and to learn them through a min–max formulation, maximizing the coercivity while minimizing over all admissible functions. Such a neural min–max procedure is attractive because it allows the weight to adapt to the hidden geometry of the singularity, rather than imposing an ansatz by hand. If successful, it would provide a data-informed route toward identifying the correct singular weights, proving linear stability in a sharp weighted space, and ultimately opening the door to a nonlinear stability argument for the weak-convection singularity. We also study the linearized spectrum numerically to infer empirically the stability and instability of the profiles.

Chapter 7

EXPONENTIALLY CONVERGENT MULTISCALE FINITE ELEMENT METHOD

In this chapter, we present the exponentially convergent multiscale finite element method (ExpMsFEM), developed for efficient model reduction of PDEs in heterogeneous media without scale separation and for high-frequency wave propagation problems. This method, based on a series of articles [81, 83, 82], systematically enriches the approximation space within a non-overlapping domain decomposition framework to achieve nearly exponential convergence with respect to the number of basis functions. We provide a concise overview based on our review paper [82], with technical details and full theoretical analysis found in the cited works.

A central challenge in achieving exponential convergence is overcoming the algebraic Kolmogorov n-width barrier, which requires basis functions that depend on the right-hand side of the equation. ExpMsFEM addresses this by introducing a two-part function representation: an offline stage, where basis functions independent of the right-hand side are constructed and used to assemble the Galerkin system; and an online stage, where problem-dependent correctors are computed efficiently and in parallel. This decomposition enables high-accuracy solutions and computational reuse in multi-query scenarios.

7.1 Introduction

Multiscale methods provide an efficient way to solve challenging PDEs. A few local basis functions adapted to the problem are constructed offline to provide an effective model reduction of the equation. One can then use the reduced model to compute the solution online, possibly with different right-hand sides and in a way much faster than solving the original equation. This property is beneficial in multi-query scenarios such as optimal design and inverse problems. Moreover, multiscale methods are inevitable for challenging problems in rough media and high-frequency wave propagation since standard numerical methods suffer from a vast number of degrees of freedom. See examples of the failure of finite element methods (FEMs) in elliptic equations with rough coefficients [15] and the pollution effect in the Helmholtz equation [17].

In this chapter, we present the framework of ExpMsFEM, the exponentially convergent multiscale finite element method. It is a generalization of the classical MsFEM [220]. The main contribution of ExpMsFEM is the systematic improvement over MsFEM to achieve exponentially convergent accuracy regarding the number of basis functions. Also, unlike most generalizations of MsFEM in the literature, ExpMsFEM does not rely on the partition of unity functions to connect local and global approximation spaces. Instead, ExpMsFEM uses edge localization and coupling intrinsic to the non-overlapped domain decomposition to communicate the local and global approximations.

In the literature, exponentially convergent multiscale methods have been pioneered in the work of optimal basis [14] based on the partition of unity functions; see also the developments in [416, 51, 79, 18, 401, 303, 304]. The work demonstrates the importance of Caccioppoli's inequality in establishing exponential convergence; more precisely, the inequality implies the *low approximation complexity* of the restriction operator acting on harmonic-type functions. The theory of ExpMsFEM is also based on some arguments using Caccioppoli's inequality. Additionally, since no partition of unity functions is used, technical tools such as C^α estimates and trace theorems are needed to analyze ExpMsFEM. ExpMsFEM was the first method to achieve exponential convergence on Helmholtz equations. We will comment on the similarities and differences between the optimal basis work and ExpMsFEM at the end of the chapter. We focus on articulating the main ideas and the computational framework in the case of 2D stationary problems with homogeneous boundary data. We provide references for the detailed analysis in the corresponding papers [81, 83].

7.1.1 Organization

In Section 7.2, we present the model problem that is the focus of this chapter. In Section 7.3, we present the motivation and framework of the ExpMsFEM. We provide numerical experiments to demonstrate the effectiveness of the ExpMsFEM framework in Section 7.4. In Section 7.5, we discuss related literature, future possibilities, and open questions.

7.2 Model Problem

Consider the model problem in a bounded domain $\Omega \subset \mathbb{R}^d$ with a Lipschitz boundary Γ. Here, $d = 2$. For generality, the boundary can contain disjoint parts $\Gamma = \Gamma_1 \cup \Gamma_2$ where Γ_1 corresponds to the Dirichlet boundary conditions and Γ_2 corresponds to the Neumann and Robin type boundary conditions.

The model equation is:

$$\begin{cases} -\nabla \cdot (A\nabla u) + Vu = f, \text{ in } \Omega \\ u = 0, \text{ on } \Gamma_1 \\ A\nabla u \cdot \nu = \beta u, \text{ on } \Gamma_2 . \end{cases} \tag{7.2.1}$$

Here, A, V, β are functions in $L^\infty(\Omega)$ and can be rough, which makes the solution oscillating and difficult to solve. The vector ν is the outer normal to the boundary.

In particular, when $V = 0$, the equation is the standard elliptic equation [81]. If $Vu = -k^2 u$ and u is a complex-valued function, one obtains the Helmholtz equation [83] with wavenumber k.

The weak formulation of (7.2.1) is given by

$$a(u, v) := (A\nabla u, \nabla v)_\Omega + (Vu, v)_\Omega - (\beta u, v)_{\Gamma_2} = (f, v)_\Omega, \quad \forall v \in \mathcal{H}(\Omega) , \tag{7.2.2}$$

where $(\cdot, \cdot)_X$ is the standard L^2 inner product on the set X. The space for v is $\mathcal{H}(\Omega) := \{w \in H^1(\Omega) : w|_{\Gamma_1} = 0\}$ and the solution $u \in \mathcal{H}(\Omega)$. The energy norm $\|\cdot\|_{\mathcal{H}(\Omega)}$ is defined as

$$\|w\|^2_{\mathcal{H}(\Omega)} := (A\nabla w, \nabla w)_\Omega + (|V|w, w)_\Omega .$$

Here, we adopt an abuse of notation that the space can be real-valued or complex-valued, depending on the context.

A generic assumption for A is $0 < A_{\min} \leq A(x) \leq A_{\max} < \infty$. We will present more detailed assumptions on V, β later in specific problems that our theory in [81, 83] covers. Indeed, the theory can encompass the case for very general V, provided that $|(Vu, u)|_\Omega \leq V_0(u, u)_\Omega$ for some constant V_0 and the PDE satisfies good stability estimates; see for example the rough Helmholtz example in [83]. In this review, we mainly focus on the *conceptual algorithmic framework* of solving the equation (7.2.1) via ExpMsFEM rather than a detailed analysis of the equation and the method.

7.3 The ExpMsFEM Framework

In subsection 7.3.1, we discuss the general recipe for solving PDEs as a function approximation problem. This motivates us to find accurate function representations to be used in the Galerkin method. We explain how ExpMsFEM manages to get exponentially convergent representations in subsections 7.3.2, 7.3.3, 7.3.4 and 7.3.5.

7.3.1 Solving PDEs as function approximation

By the standard finite element theory (e.g., [43]), when using the Galerkin method to solve (7.2.2), a key step is to find a function representation, or a space of basis functions that can approximate the solution accurately. More precisely, suppose the space is S, then, one usually wants

$$\eta(S) := \sup_{f\in L^2(\Omega)\setminus\{0\}} \inf_{v\in S} \frac{\|N(f)-v\|_{\mathcal{H}(\Omega)}}{\|f\|_{L^2(\Omega)}} \tag{7.3.1}$$

to be small. Here, $N : f \to u$ is the solution operator[1] of (7.2.1).

For example, consider the elliptic equation with $V = 0$ and $\Gamma_2 = \emptyset$. In such cases, the Galerkin method provides an optimal approximation of the solution in the space of basis functions with respect to the energy norm [43, 81], due to the Galerkin orthogonality. Therefore, a small $\eta(S)$ directly implies a small error in the solution. For the Helmholtz equation, similar arguments hold based on the Gårding-type inequality, which leads to the quasi-optimality of the solution; see, for example, [317, 83]. The failure of many finite element methods in elliptic equations with rough coefficients [15] and Helmholtz's equations [17] is due to the poor approximation property. $\eta(S)$ is typically not small if S is the standard finite element space, such as the space of tent functions.

Conceptually, ExpMsFEM finds an exponentially convergent function representation of the solution through the following three steps: (1) harmonic-bubble splitting, (2) edge localization, (3) oversampling and exponentially convergent singular value decomposition (SVD). We will detail the three steps and discuss relevant rigorous results at the end of subsections 7.3.2, 7.3.3, and 7.3.4. Then, we summarize the algorithm in subsection 7.3.5.

7.3.2 Harmonic-bubble splitting

Consider a shape regular and uniform partition of the domain Ω into finite elements with a mesh size H. The collection of elements is denoted by $\mathcal{T}_H = \{T_1, T_2, ..., T_r\}$. Let $\mathcal{E}_H = \{e_1, e_2, ..., e_q\}$ be the collection of edges in the interior of Ω. We use $\mathcal{N}_H = \{x_1, x_2, ..., x_p\}$ to denote the collection of interior nodes. We also use E_H to denote the collection of interior edges as a set, i.e., $E_H = \bigcup_{e\in\mathcal{E}_H} e \subset \Omega$. A more detailed explanation of the mesh structure can be found in [81, 83].

[1] Sometimes, N is chosen to be the solution operator of the adjoint equation; for example see [317].

In each element $T \in \mathcal{T}_H$, we decompose the solution u into $u = u_T^{\mathsf{h}} + u_T^{\mathsf{b}}$ such that

$$\begin{cases} -\nabla \cdot (A\nabla u_T^{\mathsf{h}}) + V u_T^{\mathsf{h}} = 0, \text{ in } T \\ \qquad\qquad u_T^{\mathsf{h}} = u, \text{ on } \partial T \setminus (\Gamma_1 \cup \Gamma_2) \\ \qquad\qquad u_T^{\mathsf{h}} = 0, \text{ on } \partial T \cap \Gamma_1 \\ \qquad A\nabla u_T^{\mathsf{h}} \cdot \nu = \beta u_T^{\mathsf{h}}, \text{ on } \partial T \cap \Gamma_2 , \\ -\nabla \cdot (A\nabla u_T^{\mathsf{b}}) + V u_T^{\mathsf{b}} = f, \text{ in } T \\ \qquad\qquad u_T^{\mathsf{b}} = 0, \text{ on } \partial T \setminus (\Gamma_1 \cup \Gamma_2) \\ \qquad\qquad u_T^{\mathsf{b}} = 0, \text{ on } \partial T \cap \Gamma_1 \\ \qquad A\nabla u_T^{\mathsf{b}} \cdot \nu = \beta u_T^{\mathsf{b}}, \text{ on } \partial T \cap \Gamma_2 . \end{cases} \tag{7.3.2}$$

In short, u_T^{h} incorporates the interior boundary value of u on the element, while u_T^{b} contains information of the right-hand side. All equations in (7.3.2) should be understood in the standard weak sense as in (7.2.2).

We can further define a global decomposition $u = u^{\mathsf{h}} + u^{\mathsf{b}}$, such that for each T, it holds that $u^{\mathsf{h}}(x) = u_T^{\mathsf{h}}(x)$, $u^{\mathsf{b}}(x) = u_T^{\mathsf{b}}(x)$ when $x \in T$. Here, the component u_T^{h} (resp. u^{h}) is called the local (resp. global) *harmonic part*, u_T^{b} (resp. u^{b}) is the local (resp. global) *bubble part*, of the solution u. Here, the harmonic part u^{h} is not necessarily a harmonic function due to the existence of A and V, but it has a similar low complexity property that a harmonic function has, due to the iterative argument of Caccioppoli's inequality first proposed in [14]. We will discuss this low complexity property in subsection 7.3.4.

Now, in the representation $u = u^{\mathsf{h}} + u^{\mathsf{b}}$, the part u^{b} can be directly computed by solving local problems in parallel since the local boundary conditions are all known. We are left to deal with the part u^{h}.

Remark 7.3.1. *We discuss several theoretical concerns and possible generalizations below:*

- *A sufficient condition for the local components in (7.3.2) to be well-defined is that the operator $u \longrightarrow -\nabla \cdot (A\nabla u) + Vu$ (as well as the corresponding boundary conditions) is elliptic in each local element, implied by the Poincaré inequality. In [81], we consider elliptic equations with $V = 0$ and $\Gamma_2 = \emptyset$, so this condition is satisfied. In [83], we consider the Helmholtz equation where $V < 0, |V| = O(k^2)$ and $\operatorname{Re}\beta = 0, \operatorname{Im}\beta = O(k)$. For such a case, the elliptic property is guaranteed when $H = O(1/k)$.*

- *For the global components u^{h}, u^{b} to be well-defined, we need the condition that the solution u is continuous. This can be guaranteed by the C^{α} estimates of the equation (7.2.1) under the assumptions mentioned earlier; see discussions in [81, 83].*

- *We can generalize the above decomposition to PDEs with inhomogeneous boundary conditions. To achieve so, we incorporate these boundary data into the equation for u^{b}; see also Section 5.3 in [83] for a concrete example of problems with inhomogeneous boundary data.*

7.3.3 Edge localization

The next step is to find some local basis functions that accurately approximate u^{h}. ExpMsFEM uses the idea of *edge localization* to localize this approximation task.

First, we define the "harmonic extension" operator Q_{E_H} that maps the edge values $\tilde{u}^{\mathsf{h}} = u^{\mathsf{h}}|_{E_H} \in H^{1/2}(E_H)$ to $u^{\mathsf{h}} \in H^1(\Omega)$, through the relation in the first set of equation in (7.3.2). Here, we adopt the convention that if we write a tilde on the top of a function, it is the restriction of this function on the edge set. We have that $u^{\mathsf{h}} = Q_{E_H}\tilde{u}^{\mathsf{h}} = Q_{E_H}\tilde{u}$, since u^{h} and u have the same edge values.

Then, let $C(E_H)$ be the space of continuous functions on E_H. We consider the edge interpolation operator $I_H : H^{1/2}(E_H) \cap C(E_H) \to H^{1/2}(E_H) \cap C(E_H)$ such that

$$I_H\tilde{u} = \sum_{x_i \in \mathcal{N}_H} \tilde{u}(x_i)\tilde{\psi}_i$$

where the edge function $\tilde{\psi}_i$ is linear on E_H and satisfies $\tilde{\psi}_i(x_j) = \delta_{ij}$. Note that by the convention of our notation we have $\psi_i = Q_{E_H}\tilde{\psi}_i \in H^1(\Omega)$. It is worth noting that $\psi_i's$ are the basis functions used in the vanilla MsFEM.

With the interpolation operator, we can write

$$Q_{E_H}\tilde{u} = Q_{E_H}(\tilde{u} - I_H\tilde{u}) + \sum_{x_i \in \mathcal{N}_H} u(x_i)\psi_i \,.$$

Now, the residue $\tilde{u} - I_H\tilde{u}$ is zero at each interior node. This property allows us to localize the residue to each edge. Indeed, by an abuse of notation, we can write

$$Q_{E_H}(\tilde{u} - I_H\tilde{u}) = \sum_{e \in \mathcal{E}_H} Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e \,, \tag{7.3.3}$$

where we equate the function $(\tilde{u} - I_H\tilde{u})|_e$ that is defined on e to its zero extension to E_H, so that $(\tilde{u} - I_H\tilde{u})|_e \in H^{1/2}(E_H)$ and thus $Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e$ makes sense.

Therefore, we localize the approximation task of u^{h} to $Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e$, which is defined for each edge e.

Remark 7.3.2. *Again, we discuss several theoretical concerns below:*

- *Once the condition in Remark 7.3.1 is satisfied, the extension operator Q_{E_H} is well-defined because the local equation is elliptic.*
- *According to the comment in Remark 7.3.1, the solution u is continuous, so the nodal interpolation $I_H\tilde{u}$ is well-defined.*
- *One can rigorously show that if we can approximate each local term with*

$$\|Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e - w_e\|_{\mathcal{H}(\Omega)} \leq \epsilon_e\,,$$

then the global approximation error satisfies

$$\|Q_{E_H}(\tilde{u} - I_H\tilde{u}) - \sum_{e\in\mathcal{E}_H} w_e\|^2_{\mathcal{H}(\Omega)} \leq C_{\mathrm{mesh}} \sum_{e\in\mathcal{E}_H} \epsilon_e^2\,,$$

where C_{mesh} is a constant dependent on the mesh structure only. In our previous work [81, 83], we formalize the approximation in the edge space via the $H^{1/2}_{00}(e)$ norm, which is equivalent to the $\mathcal{H}(\Omega)$ norm here after the extension by Q_{E_H}; see Proposition 2.5 and Theorem 2.6 in [81]. In this review chapter, we explain the ideas using Q_{E_H} rather than $H^{1/2}_{00}(e)$, since the former is more concise in an algorithm-focused exposition.

We call the step from local approximation to global approximation edge coupling.

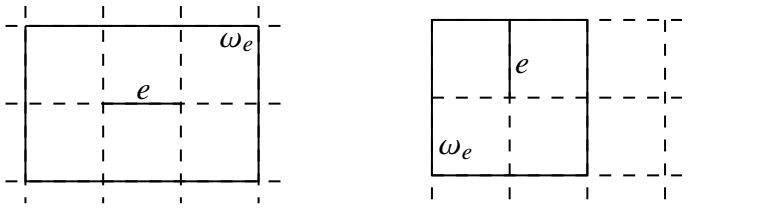
7.3.4 Exponentially convergent SVD

Recall that by using the harmonic-bubble splitting and edge localization, we get the representation

$$u = u^{\mathsf{h}} + u^{\mathsf{b}} = \sum_{e\in\mathcal{E}_H} Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e + \sum_{x_i\in\mathcal{N}_H} u(x_i)\psi_i + u^{\mathsf{b}}\,. \tag{7.3.4}$$

ExpMsFEM then relies on oversampling and local SVD to get an exponentially convergent approximation of each $Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e$. For each e, consider an oversampling domain $w_e \supset e$. Any domain containing e in the interior may be used, and as an illustrative example, we set

$$\omega_e = \overline{\bigcup\{T \in \mathcal{T}_H : \overline{T} \cap e \neq \emptyset\}}\,.$$



interior edge edge connected to boundary

Figure 7.1: Illustration of oversampling domains. On the right, we use an edge connected to the upper boundary as an illustrating example.

An illustration of this choice for a quadrilateral mesh is given in Figure 7.1.

We can view $(\tilde{u} - I_H\tilde{u})|_e$ as the image of an operator acting on $u|_{\omega_e} \in H^1(\omega_e)$. We denote this operator by R_e such that $Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e = Q_{E_H}R_e(u|_{\omega_e})$. Now, we apply the harmonic-bubble splitting in subsection 7.3.2 to the domain ω_e, which leads to $u|_{\omega_e} = u^{\mathsf{h}}_{\omega_e} + u^{\mathsf{b}}_{\omega_e}$. It follows that

$$Q_{E_H}(\tilde{u} - I_H\tilde{u})|_e = Q_{E_H}R_e u^{\mathsf{h}}_{\omega_e} + Q_{E_H}R_e u^{\mathsf{b}}_{\omega_e}\,. \tag{7.3.5}$$

The term $R_e u^{\mathsf{h}}_{\omega_e}$ is a restriction of a harmonic part. As we mentioned at the beginning of this chapter, one can prove that the restriction operator acting on harmonic-type functions is of *low approximation complexity*. More precisely, consider the space of harmonic parts in ω_e, defined via

$$\begin{aligned} U(\omega_e) := \{v \in \mathcal{H}(\omega_e) : &-\nabla\cdot(A\nabla v) + Vv = 0, \text{ in } \omega_e \\ &A\nabla v\cdot\nu = \beta v, \text{ on } \Gamma_1 \cap \partial\omega_e\}\,. \end{aligned} \tag{7.3.6}$$

The space is equipped with the norm $\|\cdot\|_{\mathcal{H}(\omega_e)}$. Then, one can show that the left singular values (in descending order) of the local operator

$$Q_{E_H}R_e : (U(\omega_e), \|\cdot\|_{\mathcal{H}(\omega_e)}) \to (\mathcal{H}(\Omega), \|\cdot\|_{\mathcal{H}(\Omega)})$$

decays as $\lambda_{e,m} \leq C\exp(-bm^{\frac{1}{d+1}})$ in dimension d, for some generic constant C, b independent of m and H. Equivalently, if we write the left singular vectors as $v_{e,m} \in H^1(\Omega)$, which is local and supported in the neighboring elements of the edge e, then there exists some coefficient $b_{e,j}$ such that

$$\|Q_{E_H}R_e u^{\mathsf{h}}_{\omega_e} - \sum_{1\leq j\leq m} b_{e,j}v_{e,j}\|_{\mathcal{H}(\Omega)} \leq C\exp(-bm^{\frac{1}{d+1}})\|u^{\mathsf{h}}_{\omega_e}\|_{\mathcal{H}(\omega_e)}\,. \tag{7.3.7}$$

For more details, see Theorem 3.10 in [83]. Then, summing these local errors up, we get

$$\begin{aligned}\sum_{e\in\mathcal{E}_H}\|u^{\mathsf{h}}_{\omega_e}\|^2_{\mathcal{H}(\omega_e)} &\leq 2\sum_{e\in\mathcal{E}_H}(\|u|_{\omega_e}\|^2_{\mathcal{H}(\omega_e)}+\|u^{\mathsf{b}}_{\omega_e}\|^2_{\mathcal{H}(\omega_e)})\\ &= O(\|u\|^2_{\mathcal{H}(\Omega)}+\|f\|^2_{L^2(\Omega)}),\end{aligned} \tag{7.3.8}$$

where we used the fact that $\|u^{\mathsf{b}}_{\omega_e}\|_{\mathcal{H}(\omega_e)} = O(\|f\|_{L^2(\omega_e)})$ by the elliptic estimate.

Combining the above estimates with edge coupling in Remark 7.3.2, we get the representation

$$\begin{aligned}u = u^{\mathsf{h}}+u^{\mathsf{b}} = &\sum_{e\in\mathcal{E}_H}\sum_{1\leq j\leq m} b_{e,j}v_{e,j}+\sum_{x_i\in\mathcal{N}_H}u(x_i)\psi_i+u^{\mathsf{n}}\\ &+O\left(\exp(-bm^{\frac{1}{d+1}})(\|u\|_{\mathcal{H}(\Omega)}+\|f\|_{L^2(\Omega)})\right),\end{aligned} \tag{7.3.9}$$

where $u^{\mathsf{n}} := u^{\mathsf{b}} + \sum_{e\in\mathcal{E}_H} Q_{E_H}R_e u^{\mathsf{b}}_{\omega_e}$ is a part that depends on f locally.

Remark 7.3.3. *We discuss several theoretical aspects and the implications of the above representation.*

- *The proof of the exponentially decaying singular values of $Q_{E_H}R_e$ is based on two steps. The first step is the iterative argument of Caccioppoli's inequality, first proposed in [14] and then refined in [304]. It shows that the singular values of the restriction operator on $U(\omega_e)$, which restricts a function from the original domain ω_e to a subdomain $\omega^* \supset e$, decay nearly exponentially fast. The second step is based on a stability estimate of the operator $Q_{E_H}R_e$ acting on $U(\omega^*)$; see Lemma 3.10 in [81] or Lemma 6.1, 6.2 in [83].*

- *We can understand that the oversampling technique is used to take advantage of the low complexity property of the restriction operator. Historically, the idea of oversampling was proposed in [220] to reduce the resonance error in MsFEM.*

- *The remarkable thing about the representation in (7.3.9) is the exponentially decaying error bound.*

  First, for elliptic equations with rough coefficients, the error bound implies that these basis functions can capture the behavior of the solution, which is a hard task for FEMs. Therefore, ExpMsFEM overcomes the difficulty of rough coefficients.

Second, for the Helmholtz equation, the stability constant of the solution operator can depend on k; indeed, this is the main cause of the pollution effect [17]. Denote the stability constant by $C_{stab}(k)$ such that $\|u\|_{\mathcal{H}(\Omega)} \leq C_{stab}(k)\|f\|_{L^2(\Omega)}$. A prevalent and reasonable assumption on the constant is that of polynomial growth, namely $C_{\text{stab}}(k) \leq C(1+k^\gamma)$ for some constants γ and C; see, for example, [260]. In such case, we can further bound the error by

$$\exp(-bm^{\frac{1}{d+1}})(\|u\|_{\mathcal{H}(\Omega)}+\|f\|_{L^2(\Omega)}) \leq \exp(-bm^{\frac{1}{d+1}})(C(1+k^\gamma)+1)\|f\|_{L^2(\Omega)}\,.$$

Therefore, once the number of basis functions per edge $m \sim \log^{d+1}(k)$ (logarithmically on k only), the approximation error can be uniformly small for all k. It implies that the quantity $\eta(S)$ in (7.3.1) is small, which is important in determining the error of Galerkin's methods. In this sense, ExpMsFEM overcomes the difficulty of the pollution effect by using basis functions whose number scales at most $\log^{d+1}(k)$.

- *The exponentially accurate representation in (7.3.9) will not be possible if we do not use terms dependent on the right-hand side. Indeed, using basis functions independent of f, the optimal approximation error rate will be algebraic if the right-hand side is in $L^2(\Omega)$ only, due to well-known results in approximation theory (the Kolmogorov n-width [368, 316]); see also the complexity analysis of the Green function of Helmholtz's equation [143]. From this perspective, we can understand that ExpMsFEM breaks the Kolmogorov barrier by using nonlinear model reduction [365], i.e., the basis functions can depend on the input of the model, here the right-hand side.*

7.3.5 The solver based on ExpMsFEM

Now, we can use the representation in (7.3.9) to solve the equation efficiently. First, we form $\psi_i, v_{e,j}$ by computing the local extension $Q_{E_H}\tilde{\psi}_i$ for each node and the top-m left singular vectors $v_{e,j}, 1 \leq j \leq m$ of the local operator $Q_{E_H}R_e$ for each e; problems on different nodes and edges are independent and parallelizable. These become our offline basis functions.

For any right-hand side f, we compute the online part u^{n} by solving local linear equations involving f. This step can be parallelized.

Then, we form an effective equation for $u - u^{\mathsf{n}}$ as

$$a(u-u^{\mathsf{n}}, v) = (f, v)_\Omega - a(u^{\mathsf{n}}, v)\,, \tag{7.3.10}$$

for any $v \in \mathcal{H}(\Omega)$. We solve the equation for $u - u^{\mathsf{n}}$ using a Galerkin method. As an example, using the Ritz-Galerkin method, we choose

$$S = \text{span}\left\{\psi_i \text{ for } x_i \in \mathcal{N}_H,\ v_{e,j} \text{ for } 1 \le j \le m, e \in \mathcal{E}_H\right\},$$

and find a numerical solution $u_S \in S$ that satisfies

$$a(u_S, v) = (f, v)_\Omega - a(u^{\mathsf{n}}, v), \tag{7.3.11}$$

for any $v \in S$. The final numerical solution is given by $u_S + u^{\mathsf{n}}$. We call u^{n} the online part and u_S the offline part since u_S lies in a space that is independent of f.

Note that in the Galerkin method for solving u_S, the stiffness matrix only needs to be assembled once and can be used for different f afterward. We can understand (7.3.10) as a reduced model of the original equation.

Remark 7.3.4. *We discuss several theoretical aspects regarding the effectiveness of the above method.*

- *The accuracy of the numerical solution is due to the quasi-optimality property mentioned earlier in subsection 7.3.1: once $\eta(S)$ is small, the solution error is of the same order compared to the optimal approximation using the basis functions, which is exponentially small according to the representation* (7.3.9).
- *When the solution is complex-valued, such as in the Helmholtz equations, we can use both the Ritz and Petrov versions of the Galerkin methods; for the former, if $\overline{S} \neq S$, we need to replace S by $S + \overline{S}$; see discussions in [83].*
- *One thing worth noting is that $\|u^{\mathsf{n}}\|_{\mathcal{H}(\Omega)}$ is of order $O(H)$, due to the standard elliptic estimate [81, 83]. Therefore, if we aim for $O(H)$ accuracy only, we can ignore this part, and simply setting $u^{\mathsf{n}} = 0$ in the above algorithm will lead to a solution accurate up to $O(H)$.*

7.4 Numerical Experiments

In this section, we present some numerical experiments to demonstrate the effectiveness of ExpMsFEM. For all the experiments, we consider the domain $\Omega = [0, 1] \times [0, 1]$ and discretize it by a uniform two-level quadrilateral mesh; see a fraction of this mesh in Figure 7.2, where we also show an edge e and its oversampling domain ω_e in solid lines. The coarse and fine mesh sizes are denoted by H and h, respectively.

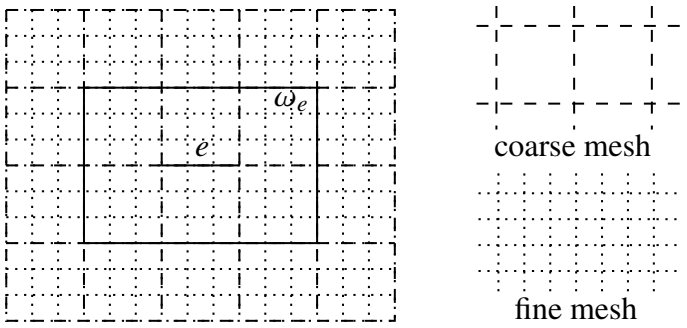


Figure 7.2: Two level mesh: a fraction

For a given equation, we compute the reference solution u_{ref} using the classical FEM on the fine mesh with a sufficiently small h, which we choose to be $h = 1/1024$. By *a posteriori* estimates, we can check that the fine mesh indeed resolves the corresponding problems; thus, the associated fine mesh solutions could serve as accurate reference solutions for all of our numerical examples. In our numerical computation, we solve local problems that are required in the ExpMsFEM framework using the fine mesh. For detailed implementation, we refer to [81, 83].

Remark 7.4.1 (Accuracy on the discrete level)**.** *For simplicity of presentation, we do not provide error analysis of ExpMsFEM on the fully discrete level, where the accuracy of the local problems can depend on the resolution of the fine grid. For a detailed error estimate on the fully discrete level in the context of partition of unity methods, see, for example, [303, 302].*

The accuracy of a numerical solution u_{sol} is computed by comparing it with the reference solution u_{ref} on the fine mesh. The accuracy will be measured both in the L^2 norm and energy norm:

$$\begin{aligned} e_{L^2} &= \frac{\|u_{\text{ref}} - u_{\text{sol}}\|_{L^2(\Omega)}}{\|u_{\text{ref}}\|_{L^2(\Omega)}}, \\ e_{\mathcal{H}} &= \frac{\|u_{\text{ref}} - u_{\text{sol}}\|_{\mathcal{H}(\Omega)}}{\|u_{\text{ref}}\|_{\mathcal{H}(\Omega)}}. \end{aligned} \tag{7.4.1}$$

In subsection 7.4.1, we consider an elliptic equation where the coefficient $A(x)$ is periodic but contains multiple scales. This example demonstrates the exponential accuracy of ExpMsFEM. In subsection 7.4.2, we consider an elliptic equation where $A(x)$ is of high contrast. This example shows the robustness of ExpMsFEM regarding the high contrast. In subsection 7.4.3, an instance of Helmholtz's equation with

rough media and mixed boundary conditions is presented. This example illustrates the effectiveness of ExpMsFEM in solving general indefinite Helmholtz equations.

7.4.1 A periodic example with multiple spatial scales

In the first example, we consider an elliptic problem ($V = 0$) with multiple spatial scales. We choose the coefficient A with five scales as follows:

$$A(x) = \frac{1}{6}\left(\frac{1.1+\sin(2\pi x_1/\epsilon_1)}{1.1+\sin(2\pi x_2/\epsilon_1)} + \frac{1.1+\sin(2\pi x_2/\epsilon_2)}{1.1+\cos(2\pi x_1/\epsilon_2)} + \frac{1.1+\cos(2\pi x_1/\epsilon_3)}{1.1+\sin(2\pi x_2/\epsilon_3)} + \frac{1.1+\sin(2\pi x_2/\epsilon_4)}{1.1+\cos(2\pi x_1/\epsilon_4)} + \frac{1.1+\cos(2\pi x_1/\epsilon_5)}{1.1+\sin(2\pi x_2/\epsilon_5)} + \sin\left(4x_1^2x_2^2\right) + 1\right), \tag{7.4.2}$$

where $x = (x_1, x_2)$, $\epsilon_1 = 1/5$, $\epsilon_2 = 1/13$, $\epsilon_3 = 1/17$, $\epsilon_4 = 1/31$, $\epsilon_5 = 1/65$. We choose homogeneous Dirichlet boundary conditions, i.e., $\Gamma_2 = \emptyset$. We set $f = -1$.

In this example, we illustrate the exponential accuracy and the convergence rate with respect to the coarse mesh size H. We take $H = 2^{-i}$, $i = 3, 4, ..., 7$ and take $m = 1, 2, ..., 6$ for each H. The numerical results are shown in Figure 7.3, where $N_c = 1/H$. We can see an exponential decay of errors for every coarse mesh size

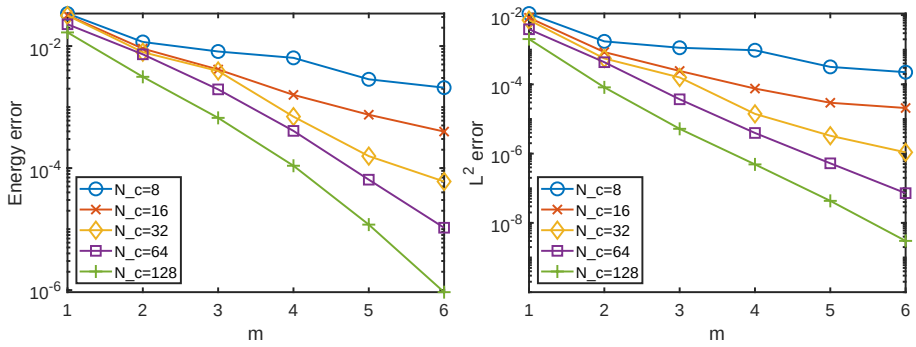


Figure 7.3: Numerical results for the periodic example. Left: $e_{\mathcal{H}}$ versus m; right: e_{L^2} versus m.

H. For smaller H, the convergence is faster. This can be understood as a finite-resolution effect. For example, when $H = 1/128$, there are only $H/h - 1 = 7$ total degrees of freedom on each edge, so of course, $m = 6$ basis per edge would result in a very accurate solution.

7.4.2 An example with high contrast channels

In the second example, we consider an elliptic problem ($V = 0$) with high contrast channels. Let

$$X := \{(x_1, x_2) \in [0,1]^2, x_1, x_2 \in \{0.2, 0.3, ..., 0.8\}\} \subset [0,1]^2,$$

and the coefficient is defined as

$$A(x) = \begin{cases} 1, & \text{if } \operatorname{dist}(x, X) \geq 0.015 \\ M, & \text{else}\,. \end{cases}$$

Here, M is a parameter controlling the contrast. We visualize $\log_{10} A$ in the left plot of Figure 7.4 for $M = 10^6$. Again, we choose homogeneous Dirichlet boundary

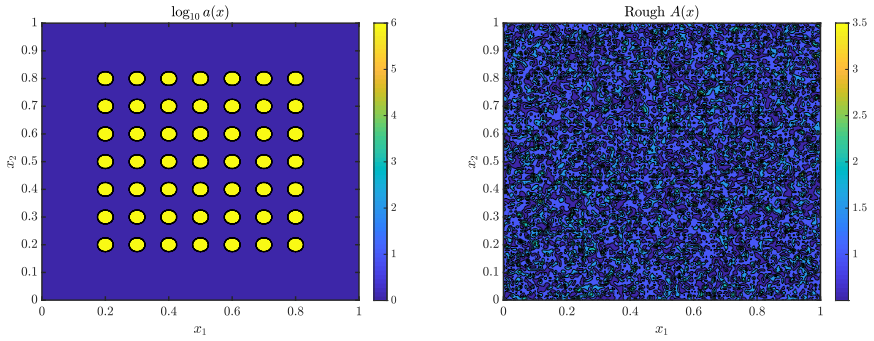


Figure 7.4: Left: the contour of $\log_{10} A$ for the high contrast example; right: the contour of A for the rough media example.

conditions, i.e., $\Gamma_2 = \emptyset$, with a non-constant right-hand side $f(x) = x_1^4 - x_2^3 + 1$.

In this example, we illustrate the convergence rate w.r.t the contrast M. We take different M using the coarse mesh size $H = 2^{-5}$ and $m = 1, 2, ..., 7$. The numerical results are shown in Figure 7.5. We observe a consistently exponential error decay

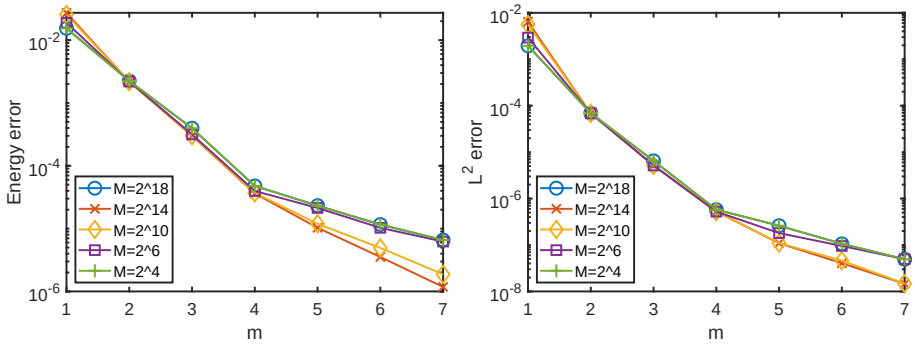


Figure 7.5: Numerical results for the high contrast example. Left: $e_{\mathcal{H}}$ versus m; right: e_{L^2} versus m.

independent of the contrast. Thus, our method demonstrates robustness with respect to the contrast $A(x)$. An intuitive explanation for this robustness could be that every step in ExpMsFEM is adaptive to $A(x)$. For example, the singular value decay of the operator $Q_{E_H} R_e$ would have some robustness regarding high contrasts in

$A(x)$ because both of the norms in the domain and image of the operator are $A(x)$-weighted. We leave the theoretical analysis of deriving an $A(x)$-adapted estimate for future study.

Also, we would like to mention that the size $h = 1/1024$ of the fine mesh can actually resolve contrasts $M = 2^4$ and 2^6 only; for higher contrasts, a posterior error analysis shows the reference solution on the fine mesh is not very accurate. However, we consistently observe a small error in our solution compared to the fine mesh solution, even in the regime where the fine mesh solution itself is not accurate. This implies that ExpMsFEM admits a very accurate dimension reduction of the equation on the fine mesh.

7.4.3 An example of Helmholtz equation with rough field and mixed boundary

In the last example, we consider the Helmholtz equation. This example is the same as Example 3 in [83]. We present it here to demonstrate that our methods are effective for complicated coefficients and mixed boundary conditions.

We impose the homogeneous Dirichlet boundary condition on $(x_1, 0), x_1 \in [0, 1]$, the homogeneous Neumann boundary condition on $(x_1, 1), x_1 \in [0, 1]$, and the homogeneous Robin boundary condition on the other two parts of $\partial\Omega$. We choose $A(x)$ to be a realization of some random field; more precisely, we set

$$A(x) = |\xi(x)| + 0.5\,, \tag{7.4.3}$$

where the field $\xi(x)$ satisfies

$$\xi(x) = a_{11}\xi_{i,j} + a_{21}\xi_{i+1,j} + a_{12}\xi_{i,j+1} + a_{22}\xi_{i+1,j+1}, \text{ if } x \in [\frac{i}{2^7}, \frac{i+1}{2^7}) \times [\frac{j}{2^7}, \frac{j+1}{2^7})\,.$$

Here, $\{\xi_{i,j}, 0 \le i, j \le 2^7\}$ are i.i.d. standard Gaussian random variables. In addition, $a_{11} = (i + 1 - 2^7x_1)(j + 1 - 2^7x_2)$, $a_{21} = (2^7x_1 - i)(j + 1 - 2^7x_2)$, $a_{12} = (i+1-2^7x_1)(2^7x_2 - j)$, $a_{22} = (2^7x_1 - i)(2^7x_2 - j)$ are interpolating coefficients to make $\xi(x)$ piecewise linear. A sample from this field is displayed in the right plot of Figure 7.4.

Moreover, we also take V/k^2 and β/ik as independent samples drawn from this random field. We choose the wavenumber $k = 2^5$, the right-hand side $f(x_1, x_2) = x_1^4 - x_2^3 + 1$, and the coarse mesh $H = 2^{-5}$. Again, we take $m = 1, 2, ..., 7$ and present the numerical results in Figure 7.6. Clearly, a nearly exponential rate of convergence is still observed for this challenging example.

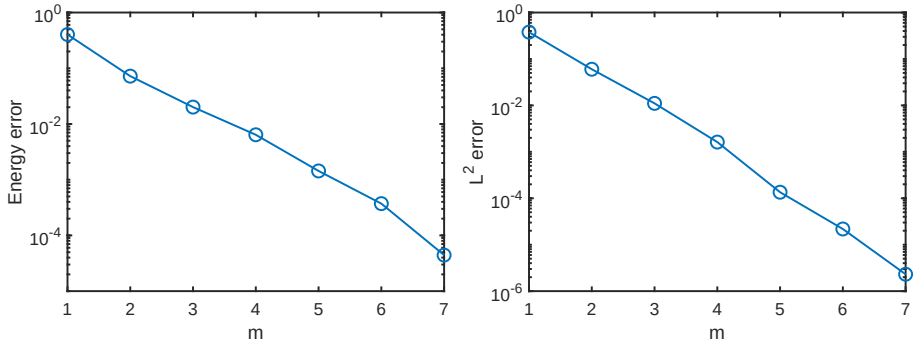


Figure 7.6: Numerical results for the mixed boundary and rough field example. Left: $e_{\mathcal{H}}$ versus m; right: e_{L^2} versus m.

7.5 Discussions

In this section, we discuss related multiscale methods in the literature; for a more specific review under the context of the elliptic and Helmholtz equations, see [81, 83]. We also outline future possibilities and open questions about ExpMsFEM at the end of this section.

7.5.1 Related literature

There is a vast amount of literature on multiscale methods and numerical homogenization.

Earlier work mainly focuses on structured $A(x)$ such as in periodic media and with scale separation; some examples include the generalized finite element methods (GFEM) [16], the multiscale finite element method (MsFEM) [220, 221, 134], the variational multiscale methods (VMS) [227], and the heterogeneous multiscale method (HMM) [1].

Later on, people are interested in multiscale methods that can address more general rough coefficients that lie in $L^{\infty}(\Omega)$ only; see, for example, the work of optimal basis using partition of unity functions [14, 18, 303, 304], harmonic coordinates [360], local orthogonal decomposition (LOD) [307, 193, 256, 189, 305], Gamblets related approaches [359, 361, 356, 357, 222, 358, 80], and generalizations of MsFEM [216, 87, 271, 154]. Different methods differ in how to find an accurate function representation. In deriving the function representation in ExpMsFEM, the solution is first decomposed into a harmonic part and a bubble part. For elliptic equations, this decomposition is the same as the orthogonal decomposition in previous work of MsFEM [216] and approximate component mode synthesis [202, 201].

To the best of our knowledge, among all the previous work, the optimal basis framework using partition of unity functions (and its variant) is the only one that achieves nearly exponential accuracy regarding the number of basis functions. Our ExpMsFEM [81, 83] is motivated by the argument of Caccioppoli's inequality used in the optimal basis framework. ExpMsFEM is the first framework that achieves exponential accuracy without using partition of unity functions and is a direct generalization of MsFEM.

We comment in more detail on the differences and similarities between the optimal basis framework and ExpMsFEM. In the optimal basis framework, the exponentially accurate representation is obtained through the partition of unity functions rather than the edge localization and coupling in ExpMsFEM. More precisely, one can write

$$u = \sum_i \eta_i u = \sum_i \eta_i u^{\mathsf{h}}_{\omega_i} + \sum_i \eta_i u^{\mathsf{b}}_{\omega_i}, \tag{7.5.1}$$

where $\{\eta_i\}_i$ are partition of unity functions subordinate to an overlapped domain decomposition $\{\omega_i\}_i$ and $u^{\mathsf{h}}_{\omega_i}, u^{\mathsf{b}}_{\omega_i}$ are obtained by the harmonic-bubble splitting in ω_i. The part $\eta_i u^{\mathsf{h}}_{\omega_i}$ can be seen as a "restriction" of harmonic-type functions. Thus, the argument using Caccioppoli's inequality implies that this part can be approximated by basis functions with a nearly exponential convergence rate.

Compared to (7.3.9), the representation (7.5.1) admits better geometric flexibility since by using partition of unity functions, such representation can work for problems in general dimensions. The representation (7.3.9) produced by ExpMsFEM is tied to the mesh structure. When $d = 2$, we have nodal and edge basis functions in the representation (7.3.9). When $d \geq 3$, we need facial basis functions and so on to represent the solution; for details see section 7 in [83]. In this sense, ExpMsFEM removes the partition of unity functions in the overlapped domain decomposition but pays the design cost of using a more complicated geometric structure in the non-overlapped domain decomposition. Nevertheless, the benefit of non-overlapped domain decomposition is that the basis functions are more localized since the local domain is smaller. Also, ExpMsFEM does not have the additional parameter of the partition of unity functions. Some basic numerical comparisons between ExpMsFEM and optimal basis using partition of unity functions are presented in [83]. We need a more in-depth comparison between the two approaches to identify their trade-offs more clearly.

7.5.2 Future directions

To now, ExpMsFEM has been successfully applied to solve elliptic and Helmholtz equations. Moving forward, one can extend this idea to advection-dominated diffusion problems, time-dependent problems such as Schrödinger's equations, and many other linear equations. Extension to nonlinear equations appears to be nontrivial since the decomposition used in ExpMsFEM requires linearity of the equation. It could be interesting to explore the combination of ExpMsFEM and linearization to provide nonlinear homogenization of these equations.

For the current ExpMsFEM framework, we observe its robustness regarding the high contrast in the media numerically (subsection 7.4.2), but a rigorous understanding of such robustness is still lacking. Moreover, a discrete-level analysis of ExpMsFEM could be helpful for its practical use.

In essence, both ExpMsFEM and optimal basis using partition of unity functions take advantage of the low approximation complexity structures of the restriction operator on harmonic-type functions. Finding other novel low complexity structures is crucial to advance multiscale computation and model reduction.

ExpMsFEM and optimal basis using partition of unity functions imply that nonlinear model reduction can break the Kolmogorov barrier and achieve remarkable exponential convergence. Embedding this idea to data-driven model reduction or operator learning also represents an exciting avenue for future work.

BIBLIOGRAPHY

[1] Assyr Abdulle et al. “The heterogeneous multiscale method”. In: *Acta Numerica* 21 (2012), pp. 1–87.

[2] Diab W Abueidda, Panos Pantidis, and Mostafa E Mobasher. “Deepokan: Deep operator network based on kolmogorov arnold networks for mechanics problems”. In: *arXiv preprint arXiv:2405.19143* (2024).

[3] Robert A Adams and John JF Fournier. *Sobolev spaces*. Elsevier, 2003.

[4] Rishabh Agarwal et al. “Neural additive models: Interpretable machine learning with neural nets”. In: *Advances in neural information processing systems* 34 (2021), pp. 4699–4711.

[5] Alireza Afzal Aghaei. “fKAN: Fractional Kolmogorov-Arnold Networks with trainable Jacobi basis functions”. In: *arXiv preprint arXiv:2406.07456* (2024).

[6] Alireza Afzal Aghaei. “rKAN: Rational Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2406.14495* (2024).

[7] Tashin Ahmed and Md Habibur Rahman Sifat. “GraphKAN: Graph Kolmogorov Arnold Network for Small Molecule-Protein Interaction Predictions”. In: *ICML’24 Workshop ML for Life and Material Science: From Theory to Industry Applications*. 2024.

[8] Sokratis J Anagnostopoulos et al. “Residual-based attention in physics-informed neural networks”. In: *Computer Methods in Applied Mechanics and Engineering* 421 (2024), p. 116805.

[9] Igor S Aranson and Lorenz Kramer. “The world of the complex Ginzburg-Landau equation”. In: *Reviews of modern physics* 74.1 (2002), p. 99.

[10] Basim Azam and Naveed Akhtar. “Suitability of KANs for Computer Vision: A preliminary investigation”. In: *arXiv preprint arXiv:2406.09087* (2024).

[11] Shayan Aziznejad and Michael Unser. “Deep spline networks with control of Lipschitz regularity”. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2019, pp. 3242–3246.

[12] M Al-Baali. “Numerical experience with a class of self-scaling quasi-Newton algorithms”. In: *Journal of optimization theory and applications* 96 (1998), pp. 533–553.

[13] Mehiddin Al-Baali, Emilio Spedicato, and Francesca Maggioni. “Broyden’s quasi-Newton methods for a nonlinear system of equations and unconstrained optimization: a review and open problems”. In: *Optimization Methods and Software* 29.5 (2014), pp. 937–954.

[14] Ivo Babuška and Robert Lipton. “Optimal local approximation spaces for generalized finite element methods with application to multiscale problems”. In: *Multiscale Modeling & Simulation* 9.1 (2011), pp. 373–406.

[15] Ivo Babuška and John Osborn. “Can a finite element method perform arbitrarily badly?” In: *Mathematics of Computation* 69.230 (2000), pp. 443–462.

[16] Ivo Babuška and John E Osborn. “Generalized finite element methods: their performance and their relation to mixed methods”. In: *SIAM Journal on Numerical Analysis* 20.3 (1983), pp. 510–536.

[17] Ivo Babuška and Stefan Sauter. “Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?” In: *SIAM Journal on numerical analysis* 34.6 (1997), pp. 2392–2423.

[18] Ivo Babuška et al. “Multiscale-Spectral GFEM and optimal oversampling”. In: *Computer Methods in Applied Mechanics and Engineering* 364 (2020), p. 112960.

[19] Yasaman Bahri et al. “Explaining neural scaling laws”. In: *arXiv preprint arXiv:2102.06701* (2021).

[20] Dominique Bakry and Michel Émery. “Diffusions hypercontractives”. In: *Seminaire de probabilités XIX 1983/84*. Springer, 1985, pp. 177–206.

[21] Andrew R Barron. “Universal approximation bounds for superpositions of a sigmoidal function”. In: *IEEE Transactions on Information theory* 39.3 (1993), pp. 930–945.

[22] Peter L Bartlett et al. “Nearly-tight VC-dimension and pseudodimension bounds for piecewise linear neural networks”. In: *Journal of Machine Learning Research* 20.63 (2019), pp. 1–17.

[23] Nuri Benbarka, Timon Höfer, Andreas Zell, et al. “Seeing implicit neural representations as fourier series”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022, pp. 2041–2050.

[24] Jan Bouwe van den Berg, John R King, and Josephus Hulshof. “Formal asymptotics of bubbling in the harmonic map heat flow”. In: *SIAM Journal on Applied Mathematics* 63.5 (2003), pp. 1682–1717.

[25] Kay Bergemann and Sebastian Reich. “A mollified ensemble Kalman filter”. In: *Quarterly Journal of the Royal Meteorological Society* 136.651 (2010), pp. 1636–1643.

[26] Kay Bergemann and Sebastian Reich. “An ensemble Kalman-Bucy filter for continuous data assimilation”. In: *Meteorologische Zeitschrift* 21.3 (2012), p. 213.

[27] Marsha Berger and Robert V Kohn. “A rescaling algorithm for the numerical calculation of blowing-up solutions”. In: *Communications on pure and applied mathematics* 41.6 (1988), pp. 841–863.

[28] Michael Betancourt. “A conceptual introduction to Hamiltonian Monte Carlo”. In: *arXiv preprint arXiv:1701.02434* (2017).

[29] Michael Betancourt et al. “The geometric foundations of Hamiltonian Monte Carlo”. In: *Bernoulli* 23.4A (2017), pp. 2257–2298.

[30] Fabrice Bethuel, Haim Brezis, Frédéric Hélein, et al. *Ginzburg-landau vortices*. Vol. 13. Springer, 1994.

[31] Piotr Biler. *Singularities of solutions to chemotaxis systems*. Vol. 6. De Gruyter Series in Mathematics and Life Sciences. De Gruyter, Berlin, 2020, pp. xxiv+205.

[32] Piotr Biler, Ignacio Guerra, and Grzegorz Karch. “Large global-in-time solutions of the parabolic-parabolic Keller-Segel system on the plane”. In: *Commun. Pure Appl. Anal.* 14.6 (2015), pp. 2117–2126. ISSN: 1534-0392,1553-5258. DOI: 10.3934/cpaa.2015.14.2117. URL: https://doi.org/10.3934/cpaa.2015.14.2117.

[33] Garrett Bingham and Risto Miikkulainen. “Discovering parametric activation functions”. In: *Neural Networks* 148 (2022), pp. 48–65.

[34] Adrien Blanchet, Jean Dolbeault, and Benoît Perthame. “Two-dimensional Keller-Segel model: optimal critical mass and qualitative properties of the solutions”. In: *Electron. J. Differential Equations* (2006), No. 44, 32.

[35] Alexander Dylan Bodner et al. “Convolutional Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2406.13155* (2024).

[36] Pakshal Bohra et al. “Learning activation functions in deep (spline) neural networks”. In: *IEEE Open Journal of Signal Processing* 1 (2020), pp. 295–309.

[37] Nawaf Bou-Rabee and Andreas Eberle. “Two-scale coupling for preconditioned Hamiltonian Monte Carlo in infinite dimensions”. In: *Stochastics and Partial Differential Equations: Analysis and Computations* 9.1 (2021), pp. 207–242.

[38] Nawaf Bou-Rabee and Jesús María Sanz-Serna. “Geometric integrators and the Hamiltonian Monte Carlo method”. In: *Acta Numerica* 27 (2018), pp. 113–206.

[39] Nawaf Bou-Rabee and Jesús María Sanz-Serna. “Randomized Hamiltonian Monte Carlo”. In: *The Annals of Applied Probability* 27.4 (2017), pp. 2159–2194.

[40] Zavareh Bozorgasl and Hao Chen. “Wav-kan: Wavelet kolmogorov-arnold networks”. In: *arXiv preprint arXiv:2405.12832* (2024).

[41] Helmut R Brand, Peter S Lomdahl, and Alan C Newell. "Benjamin-Feir turbulence in convective binary fluid mixtures". In: *Physica D: Nonlinear Phenomena* 23.1-3 (1986), pp. 345–361.

[42] Michael P. Brenner et al. "Diffusion, attraction and collapse". In: *Nonlinearity* 12.4 (1999), pp. 1071–1098. ISSN: 0951-7715,1361-6544. DOI: 10.1088/0951-7715/12/4/320. URL: https://doi.org/10.1088/0951-7715/12/4/320.

[43] Susanne C Brenner, L Ridgway Scott, and L Ridgway Scott. *The mathematical theory of finite element methods*. Vol. 3. Springer, 2008.

[44] Roman Bresson et al. "Kagnns: Kolmogorov-arnold networks meet graph learning". In: *arXiv preprint arXiv:2406.18380* (2024).

[45] J. Bricmont and A. Kupiainen. "Universality in blow-up for nonlinear heat equations". In: *Nonlinearity* 7.2 (1994), p. 539.

[46] Steve Brooks et al. *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[47] Tristan Buckmaster, Gonzalo Cao-Labora, and Javier Gómez-Serrano. "Smooth imploding solutions for 3D compressible fluids". English. In: *Forum Math. Pi* 13 (2025). Id/No e6, p. 139. ISSN: 2050-5086. DOI: 10.1017/fmp.2024.12.

[48] Tristan Buckmaster, Steve Shkoller, and Vlad Vicol. "Formation of point shocks for 3D compressible Euler". In: *Communications on Pure and Applied Mathematics* 76.9 (2023), pp. 2073–2191.

[49] Tristan Buckmaster, Steve Shkoller, and Vlad Vicol. "Formation of shocks for 2D isentropic compressible Euler". In: *Communications on Pure and Applied Mathematics* 75.9 (2022), pp. 2069–2120.

[50] Chris J Budd, Vivi Rottschäfer, and JF Williams. "Multibump, Blow-Up, Self-Similar Solutions of the Complex Ginzburg–Landau Equation". In: *SIAM Journal on Applied Dynamical Systems* 4.3 (2005), pp. 649–678.

[51] Andreas Buhr and Kathrin Smetana. "Randomized local model order reduction". In: *SIAM journal on scientific computing* 40.4 (2018), A2120–A2151.

[52] Federico Buseghin et al. "Existence of finite time blow-up in Keller-Segel system". In: *arXiv preprint arXiv:2312.01475* (2023).

[53] L Caffarelli, R Kohn, and L Nirenberg. "Partial regularity of suitable weak solutions of the Navier-Stokes equations". In: *Communications on Pure and Applied Mathematics* 35.6 (1982), pp. 771–831.

[54] Shengze Cai et al. "Physics-informed neural networks (PINNs) for fluid mechanics: A review". In: *Acta Mechanica Sinica* 37.12 (2021), pp. 1727–1738.

[55] Wei Cai and Zhi-Qin John Xu. "Multi-scale deep neural networks for solving high dimensional pdes". In: *arXiv preprint arXiv:1910.11710* (2019).

[56] Yu Cao, Jianfeng Lu, and Lihan Wang. "On explicit L^2-convergence rate estimate for underdamped Langevin dynamics". In: *arXiv preprint arXiv:1908.04746* (2019).

[57] Gonzalo Cao-Labora et al. *Non-radial implosion for compressible Euler and Navier-Stokes in T^d and R^d*. Preprint, arXiv:2310.05325 [math.AP] (2023). 2023. URL: https://arxiv.org/abs/2310.05325.

[58] Gonzalo Cao-Labora et al. *Non-radial implosion for the defocusing nonlinear Schrödinger equation in T^d and R^d*. Preprint, arXiv:2410.04532 [math.AP] (2024). 2024. URL: https://arxiv.org/abs/2410.04532.

[59] JA Carrillo et al. "Consensus-based sampling". In: *Studies in Applied Mathematics* 148.3 (2022), pp. 1069–1140.

[60] José A Carrillo et al. "An analytical framework for consensus-based global optimization method". In: *Mathematical Models and Methods in Applied Sciences* 28.06 (2018), pp. 1037–1066.

[61] José A Carrillo et al. "Particle, kinetic, and hydrodynamic models of swarming". In: *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer, 2010, pp. 297–336.

[62] José A. Carrillo, Katy Craig, and Yao Yao. "Aggregation-diffusion equations: dynamics, asymptotics, and singular limits". In: *Active particles. Vol. 2. Advances in theory, models, and applications*. Model. Simul. Sci. Eng. Technol. Birkhäuser/Springer, Cham, 2019, pp. 65–108.

[63] Thierry Cazenave, João-Paulo Dias, and Mário Figueira. "Finite-time blowup for a complex Ginzburg–Landau equation with linear driving". In: *Journal of Evolution Equations* 14.2 (2014), pp. 403–415.

[64] Thierry Cazenave, Flavio Dickstein, and Fred B Weissler. "Finite-time blowup for a complex Ginzburg–Landau equation". In: *SIAM Journal on Mathematical Analysis* 45.1 (2013), pp. 244–266.

[65] Martin Chak et al. "Optimal friction matrix for underdamped Langevin sampling". In: *arXiv preprint arXiv:2112.06844* (2021).

[66] S Jonathan Chapman, Sam D Howison, and John R Ockendon. "Macroscopic models for superconductivity". In: *Siam Review* 34.4 (1992), pp. 529–560.

[67] Jiajie Chen. "On the regularity of the De Gregorio model for the 3D Euler equations". In: *J. Eur. Math. Soc., to appear, arXiv preprint arXiv:2107.04777* (2021).

[68] Jiajie Chen. "On the Slightly Perturbed De Gregorio Model on S^1". In: *Archive for Rational Mechanics and Analysis* 241.3 (2021), pp. 1843–1869.

[69] Jiajie Chen. "Remarks on the smoothness of the $C^{1,\alpha}$ asymptotically self-similar singularity in the 3D Euler and 2D Boussinesq equations". In: *arXiv preprint arXiv:2309.00150* (2023).

[70] Jiajie Chen. "Singularity formation and global well-posedness for the generalized Constantin–Lax–Majda equation with dissipation". In: *Nonlinearity* 33.5 (2020), p. 2502.

[71] Jiajie Chen. *Vorticity blowup in compressible Euler equations in R^d, $d \geq 3$*. Preprint, arXiv:2408.04319 [math.AP] (2024). 2024. URL: https://arxiv.org/abs/2408.04319.

[72] Jiajie Chen and Thomas Y Hou. "Finite time blowup of 2D Boussinesq and 3D Euler equations with $C^{1,\alpha}$ velocity and boundary". In: *Communications in Mathematical Physics* 383.3 (2021), pp. 1559–1667.

[73] Jiajie Chen and Thomas Y Hou. "Stable nearly self-similar blowup of the 2D Boussinesq and 3D Euler equations with smooth data". In: *arXiv preprint arXiv:2210.07191* (2022).

[74] Jiajie Chen and Thomas Y Hou. "Stable nearly self-similar blowup of the 2D Boussinesq and 3D Euler equations with smooth data II: rigorous numerics". In: *Multiscale Modeling & Simulation* 23.1 (2025), pp. 25–130.

[75] Jiajie Chen, Thomas Y Hou, and De Huang. "Asymptotically self-similar blowup of the Hou–Luo model for the 3D Euler equations". In: *Annals of PDE* 8.2 (2022), p. 24.

[76] Jiajie Chen, Thomas Y Hou, and De Huang. "On the finite time blowup of the De Gregorio model for the 3D Euler equations". In: *Communications on pure and applied mathematics* 74.6 (2021), pp. 1282–1350.

[77] Jiajie Chen, Thomas Y Hou, Van Tien Nguyen, and Yixuan Wang. "On the stability of blowup solutions to the complex Ginzburg-Landau equation in R^d". In: *Annals of PDE* 11.2 (2025), p. 29. DOI: 10.1007/s40818-025-00223-1.

[78] Jiajie Chen et al. "Vorticity blowup in 2D compressible Euler equations". In: *arXiv preprint arXiv:2407.06455* (2024).

[79] Ke Chen et al. "Randomized sampling for basis function construction in generalized finite element methods". In: *Multiscale Modeling & Simulation* 18.2 (2020), pp. 1153–1177.

[80] Yifan Chen and Thomas Y Hou. "Multiscale elliptic PDE upscaling and function approximation via subsampled data". In: *Multiscale Modeling & Simulation* 20.1 (2022), pp. 188–219.

[81] Yifan Chen, Thomas Y Hou, and Yixuan Wang. "Exponential convergence for multiscale linear elliptic PDEs via adaptive edge basis functions". In: *Multiscale Modeling & Simulation* 19.2 (2021), pp. 980–1010. DOI: 10.1137/20M1352922.

[82] Yifan Chen, Thomas Y Hou, and Yixuan Wang. "Exponentially convergent multiscale finite element method". In: *Communications on Applied Mathematics and Computation* 6.2 (2024), pp. 862–878. DOI: 10.1007/s42967-023-00260-2.

[83] Yifan Chen, Thomas Y Hou, and Yixuan Wang. "Exponentially convergent multiscale methods for 2d high frequency heterogeneous Helmholtz equations". In: *Multiscale Modeling & Simulation* 21.3 (2023), pp. 849–883. DOI: 10.1137/22M1507802.

[84] Zhuo Chen et al. "TENG: Time-evolving natural gradient for solving PDEs with deep neural nets toward machine precision". In: *arXiv preprint arXiv:2404.10771* (2024).

[85] Minjong Cheon. "Demonstrating the Efficacy of Kolmogorov-Arnold Networks in Vision Tasks". In: *arXiv preprint arXiv:2406.14916* (2024).

[86] Minjong Cheon. "Kolmogorov-Arnold Network for Satellite Image Classification in Remote Sensing". In: *arXiv preprint arXiv:2406.00600* (2024).

[87] Eric T Chung, Yalchin Efendiev, and Wing Tat Leung. "Constraint energy minimizing generalized multiscale finite element method". In: *Computer Methods in Applied Mechanics and Engineering* 339 (2018), pp. 298–319.

[88] Emmet Cleary et al. "Calibrate, emulate, sample". In: *Journal of Computational Physics* 424 (2021), p. 109716.

[89] C. Collot. "Nonradial type II blow up for the energy-supercritical semilinear heat equation". In: *Anal. PDE* 10.1 (2017), pp. 127–252. ISSN: 2157-5045. DOI: 10.2140/apde.2017.10.127. URL: https://doi.org/10.2140/apde.2017.10.127.

[90] C. Collot, F. Merle, and P. Raphael. "Strongly anisotropic type II blow up at an isolated point". In: *Journal of the AMS* (2019). DOI: https://doi.org/10.1090/jams/941. URL: https://arxiv.org/abs/1709.04941.

[91] C. Collot, F. Merle, and P. Raphaël. "Dynamics near the ground state for the energy critical nonlinear heat equation in large dimensions". In: *Comm. Math. Phys.* 352.1 (2017), pp. 215–285. ISSN: 0010-3616. DOI: 10.1007/s00220-016-2795-4. URL: https://doi.org/10.1007/s00220-016-2795-4.

[92] C. Collot et al. "Refined Description and Stability for Singular Solutions of the 2D Keller-Segel System". In: *Communications on Pure and Applied Mathematics* 75.7 (2022), pp. 1419–1516.

[93] Charles Collot, Slim Ibrahim, and Quyuan Lin. "Stable singularity formation for the inviscid primitive equations". In: *Annales de l'Institut Henri Poincaré C* (2023).

[94] Charles Collot, Pierre Raphaël, and Jeremie Szeftel. *On the stability of type I blow up for the energy super critical heat equation*. English. Vol. 1255. Mem. Am. Math. Soc. Providence, RI: American Mathematical Society (AMS), 2019. DOI: 10.1090/memo/1255.

[95] Charles Collot et al. "Collapsing-ring blowup solutions for the Keller-Segel system in three dimensions and higher". In: *J. Funct. Anal.* 285.7 (2023), Paper No. 110065, 41. ISSN: 0022-1236,1096-0783. DOI: 10.1016/j.jfa.2023.110065. URL: https://doi.org/10.1016/j.jfa.2023.110065.

[96] Charles Collot et al. *Singularity formed by the collision of two collapsing solitons in interaction for the 2D Keller-Segel system*. Preprint, arXiv:2409.05363 [math.AP] (2024). 2024. URL: https://arxiv.org/abs/2409.05363.

[97] P Constantin. "On the Euler equations of incompressible fluids". In: *Bulletin of the American Mathematical Society* 44.4 (2007), pp. 603–621.

[98] Peter Constantin. "Note on loss of regularity for solutions of the 3D incompressible Euler and related equations". In: *Commun. Math. Phys.* 104 (1986), pp. 311–326.

[99] Peter Constantin, Peter D Lax, and Andrew Majda. "A simple one-dimensional model for the three-dimensional vorticity equation". In: *Communications on pure and applied mathematics* 38.6 (1985), pp. 715–724.

[100] Diego Córdoba and Luis Martínez-Zoroa. "Blow-up for the incompressible 3D-Euler equations with uniform $C^{1,\frac{1}{2}-\epsilon} \cap L^2$ force". In: *arXiv preprint arXiv:2309.08495* (2023).

[101] Diego Córdoba, Luis Martínez-Zoroa, and Fan Zheng. "Finite time singularities to the 3D incompressible Euler equations for solutions in $C^\infty(R^3\backslash 0) \cap C^{1,\alpha} \cap L^2$". In: *arXiv preprint arXiv:2308.12197* (2023).

[102] L. Corrias, B. Perthame, and H. Zaag. "Global solutions of some chemotaxis and angiogenesis systems in high space dimensions". In: *Milan J. Math.* 72 (2004), pp. 1–28. ISSN: 1424-9286,1424-9294. DOI: 10.1007/s00032-003-0026-x. URL: https://doi.org/10.1007/s00032-003-0026-x.

[103] Simon L Cotter et al. "MCMC methods for functions: modifying old algorithms to make them faster". In: *Statistical Science* 28.3 (2013), pp. 424–446.

[104] Miles Cranmer. "Interpretable machine learning for science with PySR and SymbolicRegression. jl". In: *arXiv preprint arXiv:2305.01582* (2023).

[105] Jessica Craven, Vishnu Jejjala, and Arjun Kar. "Disentangling a deep learned volume formula". In: *JHEP* 06 (2021), p. 040. DOI: 10.1007/JHEP06(2021)040. arXiv: 2012.03955 [hep-th].

[106] Jessica Craven et al. "Illuminating new and known relations between knot invariants". In: (Nov. 2022). arXiv: 2211.01404 [math.GT].

[107] Hoagy Cunningham et al. “Sparse autoencoders find highly interpretable features in language models”. In: *arXiv preprint arXiv:2309.08600* (2023).

[108] George Cybenko. “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.

[109] A Davey, LM Hocking, and K Stewartson. “On the nonlinear evolution of three-dimensional disturbances in plane Poiseuille flow”. In: *Journal of Fluid Mechanics* 63.3 (1974), pp. 529–536.

[110] Alex Davies et al. “Advancing mathematics by guiding human intuition with AI”. In: *Nature* 600.7887 (2021), pp. 70–74.

[111] Carl De Boor. *A practical guide to splines*. Vol. 27. springer-verlag New York, 1978.

[112] Gianluca De Carlo, Andrea Mastropietro, and Aris Anagnostopoulos. “Kolmogorov-arnold graph neural networks”. In: *arXiv preprint arXiv:2406.18354* (2024).

[113] Salvatore De Gregorio. “A partial differential equation arising in a 1D model for the 3D vorticity equation”. In: *Mathematical methods in the applied sciences* 19.15 (1996), pp. 1233–1255.

[114] Salvatore De Gregorio. “On a one-dimensional model for the three-dimensional vorticity equation”. In: *Journal of statistical physics* 59.5 (1990), pp. 1251–1263.

[115] Steven Dejak et al. “Blow-up in nonlinear heat equations”. In: *Advances in Applied Mathematics* 40.4 (2008), pp. 433–481.

[116] Pierre Del Moral, Aline Kurtzmann, and Julian Tugaut. “On the stability and the uniform propagation of chaos of a class of extended ensemble Kalman–Bucy filters”. In: *SIAM Journal on Control and Optimization* 55.1 (2017), pp. 119–155.

[117] Ronald DeVore, Boris Hanin, and Guergana Petrova. “Neural network approximation”. In: *Acta Numerica* 30 (2021), pp. 327–444.

[118] Ronald A DeVore. “Nonlinear approximation”. In: *Acta numerica* 7 (1998), pp. 51–150.

[119] Ronald A DeVore and George G Lorentz. “Constructive Approximation”. In: *Grundlehren der mathematischen Wissenschaften* (1993).

[120] RC Di Prima and Harry L Swinney. “Instabilities and transition in flow between concentric rotating cylinders”. In: *Hydrodynamic instabilities and the transition to turbulence*. Springer, 2005, pp. 139–180.

[121] Zhiyan Ding and Qin Li. “Ensemble Kalman sampler: Mean-field limit and convergence analysis”. In: *SIAM Journal on Mathematical Analysis* 53.2 (2021), pp. 1546–1578.

[122] RC DiPrima, W Eckhaus, and LA Segel. "Non-linear wave-number interaction in near-critical two-dimensional flows". In: *Journal of Fluid Mechanics* 49.4 (1971), pp. 705–744.

[123] Jean Dolbeault and Benoît Perthame. "Optimal critical mass in the two-dimensional Keller-Segel model in R^2". In: *C. R. Math. Acad. Sci. Paris* 339.9 (2004), pp. 611–616. ISSN: 1631-073X. DOI: 10.1016/j.crma.2004.08.011. URL: https://doi.org/10.1016/j.crma.2004.08.011.

[124] Roland Donninger and Birgit Schörkhuber. "On blowup in supercritical wave equations". In: *Comm. Math. Phys.* 346.3 (2016), pp. 907–943. ISSN: 0010-3616. DOI: 10.1007/s00220-016-2610-2. URL: https://doi.org/10.1007/s00220-016-2610-2.

[125] Qiang Du, Max D Gunzburger, and Janet S Peterson. "Analysis and approximation of the Ginzburg–Landau model of superconductivity". In: *Siam Review* 34.1 (1992), pp. 54–81.

[126] Simon Duane et al. "Hybrid Monte Carlo". In: *Physics letters B* 195.2 (1987), pp. 216–222.

[127] Renáta Dubcáková. "Eureqa: software review". In: *Genetic Programming and Evolvable Machines* 12 (2011), pp. 173–178. URL: https://api.semanticscholar.org/CorpusID:36698573.

[128] Owen Dugan et al. "OccamNet: A Fast Neural Model for Symbolic Regression at Scale". In: *arXiv preprint arXiv:2007.10784* (2020).

[129] Andrew B Duncan, Nikolas Nüsken, and Grigorios A Pavliotis. "Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions". In: *Journal of Statistical Physics* 169.6 (2017), pp. 1098–1131.

[130] Giao Ky Duong, Nejla Nouaili, and Hatem Zaag. *Construction of blowup solutions for the complex Ginzburg-Landau equation with critical parameters*. Vol. 285. 1411. American Mathematical Society, 2023.

[131] Giao Ky Duong, Nejla Nouaili, and Hatem Zaag. "Flat blow-up solutions for the complex ginzburg landau equation". In: *Archive for Rational Mechanics and Analysis* 248.6 (2024), p. 117.

[132] Valentin Duruisseaux et al. "Towards enforcing hard physics constraints in operator learning frameworks". In: *ICML 2024 AI for Science Workshop*. 2024.

[133] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. "Couplings and quantitative contraction rates for Langevin dynamics". In: *The Annals of Probability* 47.4 (2019), pp. 1982–2010.

[134] Yalchin R Efendiev, Thomas Y Hou, and Xiao-Hui Wu. "Convergence of a nonconforming multiscale finite element method". In: *SIAM Journal on Numerical Analysis* 37.3 (2000), pp. 888–910.

[135] Tarek M Elgindi. "Finite-time singularity formation for $C^{1,\alpha}$ solutions to the incompressible Euler equations on R^3". In: *Annals of Mathematics* 194.3 (2021), pp. 647–727.

[136] Tarek M Elgindi and In-Jee Jeong. "Finite-time singularity formation for strong solutions to the axi-symmetric 3D Euler equations". In: *Annals of PDE* 5.2 (2019), p. 16.

[137] Tarek M Elgindi and In-Jee Jeong. "On the effects of advection and vortex stretching". In: *Archive for Rational Mechanics and Analysis* 235.3 (2020), pp. 1763–1817.

[138] Tarek M Elgindi and In-Jee Jeong. "The incompressible Euler equations under octahedral symmetry: singularity formation in a fundamental domain". In: *Advances in Mathematics* 393 (2021), p. 108091.

[139] Tarek M Elgindi and Karim R Shikh Khalil. "Strong Ill-Posedness in L^∞ for the Riesz Transform Problem". In: *arXiv preprint arXiv:2207.04556* (2022).

[140] Nelson Elhage et al. "Softmax Linear Units". In: *Transformer Circuits Thread* (2022). https://transformer-circuits.pub/2022/solu/index.html.

[141] Nelson Elhage et al. "Toy models of superposition". In: *arXiv preprint arXiv:2209.10652* (2022).

[142] Klaus-Jochen Engel and Rainer Nagel. *One-parameter semigroups for linear evolution equations*. Vol. 194. Graduate Texts in Mathematics. Springer-Verlag, New York, 2000, pp. xxii+586. ISBN: 0-387-98463-1.

[143] Björn Engquist and Hongkai Zhao. "Approximate separability of the Green's function of the Helmholtz equation in the high frequency limit". In: *Communications on Pure and Applied Mathematics* 71.11 (2018), pp. 2220–2274.

[144] Geir Evensen et al. *Data assimilation: the ensemble Kalman filter*. Vol. 2. Springer, 2009.

[145] Daniele Fakhoury, Emanuele Fakhoury, and Hendrik Speleers. "ExSpliNet: An interpretable and expressive spline-based neural network". In: *Neural Networks* 152 (2022), pp. 332–346.

[146] Charles L Fefferman. "Existence and smoothness of the Navier-Stokes equation". In: *The millennium prize problems* 57.67 (2006), p. 22.

[147] Gadi Fibich. *The nonlinear Schrödinger equation*. Vol. 192. Springer, 2015.

[148] Gadi Fibich and Doron Levy. "Self-focusing in the complex Ginzburg-Landau limit of the critical nonlinear Schrödinger equation". In: *Physics Letters A* 249.4 (1998), pp. 286–294.

[149] S. Filippas, M. A. Herrero, and J. J. L. Velázquez. "Fast blow-up mechanisms for sign-changing solutions of a semilinear parabolic equation with critical nonlinearity". In: *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* 456.2004 (2000), pp. 2957–2982. ISSN: 1364-5021. DOI: 10.1098/rspa.2000.0648. URL: http://dx.doi.org/10.1098/rspa.2000.0648.

[150] S. Filippas and Robert V Kohn. "Refined asymptotics for the blowup of $u_t - \Delta u = u^p$". In: *Communications on pure and applied mathematics* 45 (1992), pp. 821–869.

[151] Stathis Filippas and Wenxiong Liu. "On the blowup of multidimensional semilinear heat equations". In: *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*. Vol. 10. Elsevier. 1993, pp. 313–344.

[152] Sara Fridovich-Keil, Raphael Gontijo Lopes, and Rebecca Roelofs. "Spectral bias in practice: The role of function frequency in generalization". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 7368–7382.

[153] A. Friedman and B. McLeod. "Blow-up of positive solutions of semilinear heat equations". In: *Indiana Univ. Math. J.* 34.2 (1985), pp. 425–447. ISSN: 0022-2518. DOI: 10.1512/iumj.1985.34.34025. URL: http://dx.doi.org/10.1512/iumj.1985.34.34025.

[154] Shubin Fu, Eric Chung, and Guanglian Li. "Edge multiscale methods for elliptic problems with heterogeneous coefficients". In: *Journal of Computational Physics* 396 (2019), pp. 228–242.

[155] Mario Fuest. "Approaching optimality in blow-up results for Keller-Segel systems with logistic-type dampening". In: *NoDEA Nonlinear Differential Equations Appl.* 28.2 (2021), Paper No. 16, 17. ISSN: 1021-9722,1420-9004. DOI: 10.1007/s00030-021-00677-9. URL: https://doi.org/10.1007/s00030-021-00677-9.

[156] H. Fujita. "On the blowing up of solutions of the Cauchy problem for $u_t = \Delta u + u^{1+\alpha}$". In: *J. Fac. Sci. Univ. Tokyo Sect. I* 13 (1966), 109–124 (1966). ISSN: 0040-8980.

[157] Alfredo Garbuno-Inigo, Nikolas Nüsken, and Sebastian Reich. "Affine invariant interacting Langevin dynamics for Bayesian inference". In: *SIAM Journal on Applied Dynamical Systems* 19.3 (2020), pp. 1633–1658.

[158] Alfredo Garbuno-Inigo et al. "Interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler". In: *SIAM Journal on Applied Dynamical Systems* 19.1 (2020), pp. 412–441.

[159] Remi Genet and Hugo Inzirillo. "A Temporal Kolmogorov-Arnold Transformer for Time Series Forecasting". In: *arXiv preprint arXiv:2406.02486* (2024).

[160] Remi Genet and Hugo Inzirillo. “Tkan: Temporal kolmogorov-arnold networks”. In: *arXiv preprint arXiv:2405.07344* (2024).

[161] Tej-Eddine Ghoul, Van Tien Nguyen, and Hatem Zaag. “Construction and stability of blowup solutions for a non-variational semilinear parabolic system”. In: *Annales de l’Institut Henri Poincare C, Analyse non lineaire*. Vol. 35. 6. Elsevier. 2018, pp. 1577–1630.

[162] JD Gibbon. “The three-dimensional Euler equations: Where do we stand?” In: *Physica D: Nonlinear Phenomena* 237.14 (2008), pp. 1894–1904.

[163] Y. Giga and R. V. Kohn. “Asymptotically self-similar blow-up of semilinear heat equations”. In: *Comm. Pure Appl. Math.* 38.3 (1985), pp. 297–319. ISSN: 0010-3640. DOI: 10.1002/cpa.3160380304. URL: http://dx.doi.org/10.1002/cpa.3160380304.

[164] Y. Giga and R. V. Kohn. “Characterizing blowup using similarity variables”. In: *Indiana Univ. Math. J.* 36.1 (1987), pp. 1–40. ISSN: 0022-2518. DOI: 10.1512/iumj.1987.36.36001. URL: http://dx.doi.org/10.1512/iumj.1987.36.36001.

[165] Y. Giga and R. V. Kohn. “Nondegeneracy of blowup for semilinear heat equations”. In: *Comm. Pure Appl. Math.* 42.6 (1989), pp. 845–884. ISSN: 0010-3640. DOI: 10.1002/cpa.3160420607. URL: http://dx.doi.org/10.1002/cpa.3160420607.

[166] Y. Giga, S. Matsui, and S. Sasayama. “Blow up rate for semilinear heat equations with subcritical nonlinearity”. In: *Indiana Univ. Math. J.* 53.2 (2004), pp. 483–514. ISSN: 0022-2518. DOI: 10.1512/iumj.2004.53.2401. URL: http://dx.doi.org/10.1512/iumj.2004.53.2401.

[167] Yoshikazu Giga, Shin’ya Matsui, and Satoshi Sasayama. “On blow-up rate for sign-changing solutions in a convex domain”. In: *Math. Methods Appl. Sci.* 27.15 (2004), pp. 1771–1782. ISSN: 0170-4214,1099-1476. DOI: 10.1002/mma.562. URL: https://doi.org/10.1002/mma.562.

[168] J Ginibre and G Velo. “The Cauchy problem in local spaces for the complex Ginzburg-Landau equation”. In: *Differential equations, asymptotic analysis, and mathematical physics (Potsdam, 1996)* (1997), pp. 138–152.

[169] J Ginibre and G Velo. “The Cauchy problem in local spaces for the complex Ginzburg-Landau equation I. Compactness methods”. In: *Physica D: Nonlinear Phenomena* 95.3-4 (1996), pp. 191–228.

[170] J Ginibre and G Velo. “The Cauchy Problem in Local Spaces for the Complex Ginzburg—Landau Equation II. Contraction Methods”. In: *Communications in mathematical physics* 187.1 (1997), pp. 45–79.

[171] Vitaly L Ginzburg and Lev D Landau. “On the theory of superconductivity”. In: *On superconductivity and superfluidity: a scientific autobiography*. Springer, 2009, pp. 113–137.

[172] Federico Girosi and Tomaso Poggio. “Representation properties of networks: Kolmogorov’s theorem is irrelevant”. In: *Neural Computation* 1.4 (1989), pp. 465–469.

[173] Irfan Glogić and Birgit Schörkhuber. “Stable singularity formation for the Keller-Segel system in three dimensions”. English. In: *Arch. Ration. Mech. Anal.* 248.1 (2024). Id/No 4, p. 40. ISSN: 0003-9527. DOI: 10.1007/s00205-023-01947-9.

[174] Oscar Gonzalez and Andrew M Stuart. *A first course in continuum mechanics*. Vol. 42. Cambridge University Press, 2008.

[175] Jonathan Goodman and Jonathan Weare. “Ensemble samplers with affine invariance”. In: *Communications in applied mathematics and computational science* 5.1 (2010), pp. 65–80.

[176] Mitchell A Gordon, Kevin Duh, and Jared Kaplan. “Data and Parameter Scaling Laws for Neural Machine Translation”. In: *ACL Rolling Review - May 2021*. 2021. URL: https://openreview.net/forum?id=IKA7MLxsLSu.

[177] Mohit Goyal, Rajan Goyal, and Brejesh Lall. “Learning activation functions: A new paradigm for understanding neural networks”. In: *arXiv preprint arXiv:1906.09529* (2019).

[178] *GPLearn*. https://github.com/trevorstephens/gplearn. Accessed: 2024-04-19.

[179] Robert Graham. “Covariant formulation of non-equilibrium statistical thermodynamics”. In: *Zeitschrift für Physik B Condensed Matter* 26.4 (1977), pp. 397–405.

[180] Sergei Gukov, James Halverson, and Fabian Ruehle. “Rigor with machine learning from field theory to the Poincaréconjecture”. In: *Nature Reviews Physics* (2024). DOI: 10.1038/s42254-024-00709-0. URL: https://doi.org/10.1038/s42254-024-00709-0.

[181] Sergei Gukov et al. “Learning to Unknot”. In: *Mach. Learn. Sci. Tech.* 2.2 (2021), p. 025035. DOI: 10.1088/2632-2153/abe91f. arXiv: 2010.16263 [math.GT].

[182] Sergei Gukov et al. *Searching for ribbons with machine learning*. 2023. arXiv: 2304.09304 [math.GT].

[183] Yan Guo, Mahir Hadžić, and Juhi Jang. “Larson-Penston Self-similar Gravitational Collapse”. English. In: *Commun. Math. Phys.* 386 (2021), pp. 1551–1601. DOI: 10.1007/s00220-021-04175-y.

[184] Yan Guo et al. “Gravitational collapse for polytropic gaseous stars: self-similar solutions”. English. In: *Arch. Ration. Mech. Anal.* 246.2-3 (2022), pp. 957–1066. ISSN: 0003-9527. DOI: 10.1007/s00205-022-01827-8.

[185] Eldad Haber, Felix Lucka, and Lars Ruthotto. "Never look back-A modified EnKF method and its application to the training of neural networks without back propagation". In: *arXiv preprint arXiv:1805.08034* (2018).

[186] Patrick S Hagan. "Spiral waves in reaction-diffusion equations". In: *SIAM journal on applied mathematics* 42.4 (1982), pp. 762–786.

[187] J. Harada. "A higher speed type II blowup for the five dimensional energy critical heat equation". In: *Ann. Inst. H. Poincaré C Anal. Non Linéaire* 37.2 (2020), pp. 309–341. ISSN: 0294-1449,1873-1430. DOI: 10.1016/j.anihpc.2019.09.006. URL: https://doi.org/10.1016/j.anihpc.2019.09.006.

[188] Junichi Harada. "A type II blowup for the six dimensional energy critical heat equation". In: *Ann. PDE* 6.2 (2020), Paper No. 13, 63. ISSN: 2524-5317,2199-2576. DOI: 10.1007/s40818-020-00088-6. URL: https://doi.org/10.1007/s40818-020-00088-6.

[189] Moritz Hauck and Daniel Peterseim. "Super-localization of elliptic multiscale problems". In: *arXiv preprint arXiv:2107.13211* (2021).

[190] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[191] Y.H. He. *Machine Learning in Pure Mathematics and Theoretical Physics*. G - Reference,Information and Interdisciplinary Subjects Series. World Scientific, 2023. ISBN: 9781800613690. URL: https://books.google.com/books?id=6a5gzwEACAAJ.

[192] Tom Henighan et al. "Scaling laws for autoregressive generative modeling". In: *arXiv preprint arXiv:2010.14701* (2020).

[193] Patrick Henning and Daniel Peterseim. "Oversampling for the multiscale finite element method". In: *Multiscale Modeling & Simulation* 11.4 (2013), pp. 1149–1175.

[194] Luis Fernando Herbozo Contreras et al. "KAN-EEG: Towards Replacing Backbone-MLP for an Effective Seizure Detection System". In: *medRxiv* (2024), pp. 2024–06.

[195] M. A. Herrero, E. Medina, and J. J. L. Velázquez. "Self-similar blow-up for a reaction-diffusion system". In: *J. Comput. Appl. Math.* 97.1-2 (1998), pp. 99–119. ISSN: 0377-0427,1879-1778. DOI: 10.1016/S0377-0427(98)00104-6. URL: https://doi.org/10.1016/S0377-0427(98)00104-6.

[196] M. A. Herrero and J. J. L. Velázquez. "A blow up result for semi-linear heat equations in the supercritical case". In: *Preprint* (1994).

[197] M. A. Herrero and J. J. L. Velázquez. "Blow-up behaviour of one-dimensional semilinear parabolic equations". In: *Ann. Inst. H. Poincaré Anal. Non Linéaire* 10.2 (1993), pp. 131–189. ISSN: 0294-1449.

[198] M. A. Herrero and J. J. L. Velázquez. “Explosion de solutions d’équations paraboliques semilinéaires supercritiques”. In: *C. R. Acad. Sci. Paris Sér. I Math.* 319.2 (1994), pp. 141–145. ISSN: 0764-4442.

[199] M. A. Herrero and J. J. L. Velázquez. “Generic behaviour of one-dimensional blow up patterns”. In: *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* 19.3 (1992), pp. 381–450. ISSN: 0391-173X. URL: http://www.numdam.org/item?id=ASNSP%5C_1992%5C_4%5C_19%5C_3%5C_381%5C_0.

[200] Joel Hestness et al. “Deep learning scaling is predictable, empirically”. In: *arXiv preprint arXiv:1712.00409* (2017).

[201] Ulrich Hetmaniuk and Axel Klawonn. “Error estimates for a two-dimensional special finite element method based on component mode synthesis”. In: *Electron. Trans. Numer. Anal* 41 (2014), pp. 109–132.

[202] Ulrich Hetmaniuk and Richard Lehoucq. “A special finite element method based on component mode synthesis”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.3 (2010), pp. 401–420.

[203] T Hillen and KJ Painter. “A user’s guide to PDE models for chemotaxis”. In: *J. Math. Biol.* 58.1-2 (2009), pp. 183–217. ISSN: 0303-6812,1432-1416. DOI: 10.1007/s00285-008-0201-3. URL: https://doi.org/10.1007/s00285-008-0201-3.

[204] Sean Hon and Haizhao Yang. “Simultaneous neural network approximation for smooth functions”. In: *Neural Networks* 154 (2022), pp. 152–164.

[205] Qingguo Hong et al. “On the activation function dependence of the spectral bias of neural networks”. In: *arXiv preprint arXiv:2208.04924* (2022).

[206] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural networks* 2.5 (1989), pp. 359–366.

[207] Joel L Horowitz and Enno Mammen. “Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions”. In: *Annals of Statistics* 35.6 (2007), pp. 2589–2619.

[208] Dirk Horstmann. “From 1970 until present: the Keller-Segel model in chemotaxis and its consequences. I”. In: *Jahresber. Deutsch. Math.-Verein.* 105.3 (2003), pp. 103–165. ISSN: 0012-0456.

[209] Dirk Horstmann. “From 1970 until present: the Keller-Segel model in chemotaxis and its consequences. II”. In: *Jahresber. Deutsch. Math.-Verein.* 106.2 (2004), pp. 51–69. ISSN: 0012-0456.

[210] Thomas Hou, Yixuan Wang, and Changhe Yang. “Nonuniqueness of Leray-Hopf solutions to the unforced incompressible 3D Navier-Stokes Equation”. In: *arXiv preprint arXiv:2509.25116* (2025).

[211] Thomas Y Hou. "Nearly self-similar blowup of generalized axisymmetric navier-stokes and boussinesq equations". In: *arXiv preprint arXiv:2405.10916* (2024).

[212] Thomas Y Hou. "Potential singularity of the 3D Euler equations in the interior domain". In: *Foundations of Computational Mathematics* (2022), pp. 1–47.

[213] Thomas Y Hou. "Potentially Singular Behavior of the 3D Navier–Stokes Equations". In: *Foundations of Computational Mathematics* (2022), pp. 1–49.

[214] Thomas Y Hou and Zhen Lei. "On the stabilizing effect of convection in three-dimensional incompressible flows". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 62.4 (2009), pp. 501–564.

[215] Thomas Y Hou and Congming Li. "Dynamic stability of the three-dimensional axisymmetric Navier-Stokes equations with swirl". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 61.5 (2008), pp. 661–697.

[216] Thomas Y Hou and Pengfei Liu. "Optimal local multi-scale basis functions for linear elliptic equations with rough coefficient". In: *Discrete and Continuous Dynamical Systems* 36.8 (2016), pp. 4451–4476.

[217] Thomas Y Hou, Van Tien Nguyen, and Peicong Song. "Axisymmetric type II blowup solutions to the three dimensional Keller-Segel system". In: *arXiv preprint arXiv:2502.19775* (2025).

[218] Thomas Y Hou, Van Tien Nguyen, and Yixuan Wang. "L^2-based stability of blowup with log correction for semilinear heat equation". In: *Archive for Rational Mechanics and Analysis* 250.3 (2026), p. 28. DOI: 10.1007/s00205-026-02191-7.

[219] Thomas Y Hou and Yixuan Wang. "Blowup analysis for a quasi-exact 1D model of 3D Euler and Navier–Stokes". In: *Nonlinearity* 37.3 (2024), p. 035001. DOI: 10.1088/1361-6544/ad1c2f.

[220] Thomas Y Hou and Xiao-Hui Wu. "A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media". In: *Journal of Computational Physics* 134.1 (1997), pp. 169–189. ISSN: 0021-9991.

[221] Thomas Y Hou, Xiao-Hui Wu, and Zhiqiang Cai. "Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients". In: *Mathematics of computation* 68.227 (1999), pp. 913–943.

[222] Thomas Y Hou and Pengchuan Zhang. "Sparse operator compression of higher-order elliptic operators with rough coefficients". en. In: *Research in the Mathematical Sciences* 4.1 (Dec. 2017). ISSN: 2197-9847. (Visited on 12/21/2018).

[223] Daniel Z Huang, Tapio Schneider, and Andrew M Stuart. "Unscented Kalman inversion". In: *arXiv preprint arXiv:2102.01580* (2021).

[224] Daniel Zhengyu Huang et al. "Efficient derivative-free bayesian inference for large-scale inverse problems". In: *arXiv preprint arXiv:2204.04386* (2022).

[225] De Huang et al. "Self-similar finite-time blowups with smooth profiles of the generalized Constantin-Lax-Majda model". In: *arXiv preprint arXiv:2305.05895* (2023).

[226] Mark C Hughes. "A neural network approach to predicting and computing knot invariants". In: *Journal of Knot Theory and Its Ramifications* 29.03 (2020), p. 2050005.

[227] Thomas JR Hughes et al. "The variational multiscale method—a paradigm for computational mechanics". In: *Computer Methods in Applied Mechanics and Engineering* 166.1 (1998), pp. 3–24. ISSN: 0045-7825.

[228] Marco A Iglesias, Kody JH Law, and Andrew M Stuart. "Ensemble Kalman methods for inverse problems". In: *Inverse Problems* 29.4 (2013), p. 045001.

[229] G Iooss and A Mielke. "Time-periodic Ginzburg-Landau equations for one dimensional patterns with large wave length". In: *Zeitschrift für angewandte Mathematik und Physik ZAMP* 43.1 (1992), pp. 125–138.

[230] Aysu Ismayilova and Vugar E Ismailov. "On the Kolmogorov neural networks". In: *Neural Networks* (2024), p. 106333.

[231] Pierre-Emmanuel Jabin and Zhenfu Wang. "Mean field limit for stochastic particle systems". In: *Active Particles, Volume 1*. Springer, 2017, pp. 379–402.

[232] Ameya D Jagtap, Kenji Kawaguchi, and George Em Karniadakis. "Locally adaptive activation functions with slope recovery for deep and physics-informed neural networks". In: *Proceedings of the Royal Society A* 476.2239 (2020), p. 20200334.

[233] Ameya D Jagtap, Kenji Kawaguchi, and George Em Karniadakis. "Adaptive activation functions accelerate convergence in deep and physics-informed neural networks". In: *Journal of Computational Physics* 404 (2020), p. 109136.

[234] Juhi Jang, Jiaqi Liu, and Matthew Schrecker. "Converging/Diverging Self-Similar Shock Waves: From Collapse to Reflection". In: *SIAM Journal on Mathematical Analysis* 57.1 (2025), pp. 190–232. DOI: 10.1137/24M1653240. URL: https://doi.org/10.1137/24M1653240.

[235] Juhi Jang, Jiaqi Liu, and Matthew Schrecker. "On self-similar converging shock waves". English. In: *Arch. Ration. Mech. Anal.* 249.24 (2025). DOI: 10.1007/s00205-025-02096-x.

[236] Hao Jia and Vladimir Sverak. “Are the incompressible 3d Navier-Stokes equations locally ill-posed in the natural energy space?” In: *J. Funct. Anal.* 268.12 (2015), pp. 3734–3766. ISSN: 0022-1236,1096-0783. DOI: 10.1016/j.jfa.2015.04.006. URL: https://doi.org/10.1016/j.jfa.2015.04.006.

[237] Da-Quan Jiang and Donghua Jiang. *Mathematical theory of nonequilibrium steady states: on the frontier of probability and dynamical systems*. Springer Science & Business Media, 2004.

[238] Anas Jnini, Flavio Vella, and Marius Zeinhofer. “Gauss-newton natural gradient descent for physics-informed computational fluid dynamics”. In: *arXiv preprint arXiv:2402.10680* (2024).

[239] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Vol. 160. Springer Science & Business Media, 2006.

[240] Kyungkeun Kang and Angela Stevens. “Blowup and global solutions in a chemotaxis-growth system”. In: *Nonlinear Anal.* 135 (2016), pp. 57–72. ISSN: 0362-546X,1873-5215. DOI: 10.1016/j.na.2016.01.017. URL: https://doi.org/10.1016/j.na.2016.01.017.

[241] Jared Kaplan et al. “Scaling laws for neural language models”. In: *arXiv preprint arXiv:2001.08361* (2020).

[242] S. Kaplan. “On the growth of solutions of quasi-linear parabolic equations”. In: *Comm. Pure Appl. Math.* 16 (1963), pp. 305–330. ISSN: 0010-3640.

[243] Kari Karhunen. “Under lineare methoden in der wahr scheinlichkeitsrechnung”. In: *Annales Academiae Scientiarun Fennicae Series A1: Mathematia Physica* 47 (1947).

[244] George Em Karniadakis et al. “Physics-informed machine learning”. In: *Nature Reviews Physics* 3.6 (2021), pp. 422–440.

[245] L. H. Kauffman, N. E. Russkikh, and I. A. Taimanov. *Rectangular knot diagrams classification with deep learning*. 2020. arXiv: 2011.03498 [math.GT].

[246] Carlos E Kenig and Frank Merle. “Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case”. In: *Inventiones mathematicae* 166.3 (2006), pp. 645–675.

[247] Mehrdad Kiamari, Mohammad Kiamari, and Bhaskar Krishnamachari. “GKAN: Graph Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2406.06470* (2024).

[248] Jihoi Kim. “Self-similar blow up for energy supercritical semilinear wave equation”. In: *arXiv preprint arXiv:2211.13699* (2022).

[249] Alexander Kiselev. “Small scales and singularity formation in fluid dynamics”. In: *Proceedings of the International Congress of Mathematicians*. Vol. 3. 2018.

[250] Elham Kiyani et al. “Which Optimizer Works Best for Physics-Informed Neural Networks and Kolmogorov-Arnold Networks?” In: *arXiv preprint arXiv:2501.16371* (2025).

[251] Jason M Klusowski and Andrew R Barron. “Approximation by Combinations of ReLU and Squared ReLU Ridge Functions With ℓ^1 and ℓ^0 Controls”. In: *IEEE Transactions on Information Theory* 64.12 (2018), pp. 7649–7656.

[252] Michael Kohler and Sophie Langer. “On the rate of convergence of fully connected deep neural network regression estimates”. In: *The Annals of Statistics* 49.4 (2021), pp. 2231–2249.

[253] Paul Kolodner, D Bensimon, and CM Surko. “Traveling-wave convection in an annulus”. In: *Physical review letters* 60.17 (1988), p. 1723.

[254] Paul Kolodner et al. “Characterization of dispersive chaos and related states of binary-fluid convection”. In: *Physica D: Nonlinear Phenomena* 85.1-2 (1995), pp. 165–224.

[255] Mario Köppen. “On the training of a Kolmogorov Network”. In: *Artificial Neural Networks—ICANN 2002: International Conference Madrid, Spain, August 28–30, 2002 Proceedings 12*. Springer. 2002, pp. 474–479.

[256] Ralf Kornhuber, Daniel Peterseim, and Harry Yserentant. “An analysis of a class of variational multiscale methods based on subspace decomposition”. In: *Mathematics of Computation* 87.314 (2018), pp. 2765–2774.

[257] Nikola Kovachki et al. “Neural operator: Learning maps between function spaces with applications to pdes”. In: *Journal of Machine Learning Research* 24.89 (2023), pp. 1–97.

[258] Nikola B Kovachki and Andrew M Stuart. “Ensemble Kalman inversion: a derivative-free technique for machine learning tasks”. In: *Inverse Problems* 35.9 (2019), p. 095005.

[259] Akash Kundu, Aritra Sarkar, and Abhishek Sadhu. “Kanqas: Kolmogorov arnold network for quantum architecture search”. In: *arXiv preprint arXiv:2406.17630* (2024).

[260] David Lafontaine, Euan A Spence, and Jared Wunsch. “For most frequencies, strong trapping has a weak effect in frequency-domain scattering”. In: *arXiv preprint arXiv:1903.12172* (2019).

[261] Ming-Jun Lai and Zhaiming Shen. “The kolmogorov superposition theorem can break the curse of dimensionality when approximating high dimensional functions”. In: *arXiv preprint arXiv:2112.09963* (2021).

[262] Shiwei Lan et al. “Emulation of higher-order tensors in manifold Monte Carlo methods for Bayesian inverse problems”. In: *Journal of Computational Physics* 308 (2016), pp. 81–101.

[263] Michael J Landman et al. “Rate of blowup for solutions of the nonlinear Schrödinger equation at critical dimension”. In: *Physical Review A* 38.8 (1988), p. 3837.

[264] Jin Lee, Ziming Liu, Xinling Yu, Yixuan Wang, Haewon Jeong, Murphy Yuezhen Niu, and Zheng Zhang. “KANO: Kolmogorov-Arnold neural operator”. In: *The Fourteenth International Conference on Learning Representations*. 2026. URL: https://openreview.net/forum?id=2QmiKXfsIr.

[265] Benedict Leimkuhler, Charles Matthews, and Jonathan Weare. “Ensemble preconditioning for Markov chain Monte Carlo simulation”. In: *Statistics and Computing* 28.2 (2018), pp. 277–290.

[266] Pierre-Emmanuel Leni, Yohan D Fougerolle, and Frédéric Truchetet. “The kolmogorov spline network for image processing”. In: *Image Processing: Concepts, Methodologies, Tools, and Applications*. IGI Global, 2013, pp. 54–78.

[267] Moshe Leshno et al. “Multilayer feedforward networks with a nonpolynomial activation function can approximate any function”. In: *Neural networks* 6.6 (1993), pp. 861–867.

[268] H. A. Levine. “Some nonexistence and instability theorems for solutions of formally parabolic equations of the form $Pu_t = -Au + F(u)$”. In: *Arch. Rational Mech. Anal.* 51 (1973), pp. 371–386. ISSN: 0003-9527.

[269] Chenxin Li et al. “U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation”. In: *arXiv preprint arXiv:2406.02918* (2024).

[270] Congming Li and Kai Zhang. “A note on the Gagliardo-Nirenberg inequality in a bounded domain”. In: *arXiv preprint arXiv:2110.12967* (2021).

[271] Guanglian Li. “On the convergence rates of GMsFEMs for heterogeneous elliptic problems without oversampling techniques”. In: *Multiscale Modeling & Simulation* 17.2 (2019), pp. 593–619.

[272] Xinhe Li et al. “COEFF-KANs: A Paradigm to Address the Electrolyte Field with KANs”. In: *arXiv preprint arXiv:2407.20265* (2024).

[273] Zexing Li. “Mode stability for self-similar blowup of L^2 slightly supercritical NLS”. In: *in preparation* ().

[274] Zexing Li and Tao Zhou. *Finite-time blowup for Keller-Segel-Navier-Stokes system in three dimensions*. Preprint, arXiv:2404.17228 [math.AP] (2024). 2024. URL: https://arxiv.org/abs/2404.17228.

[275] Zexing Li and Tao Zhou. "Nonradial stability of self-similar blowup to Keller-Segel equation in three dimensions". In: *arXiv preprint arXiv:2501.07073* (2025).

[276] Ziyao Li. "Kolmogorov-arnold networks are radial basis function networks". In: *arXiv preprint arXiv:2405.06721* (2024).

[277] Zongyi Li et al. "Fourier neural operator for parametric partial differential equations". In: *arXiv preprint arXiv:2010.08895* (2020).

[278] Zongyi Li et al. "Fourier neural operator with learned deformations for pdes on general geometries". In: *Journal of Machine Learning Research* 24.388 (2023), pp. 1–26.

[279] Zongyi Li et al. "Geometry-informed neural operator for large-scale 3d pdes". In: *Advances in Neural Information Processing Systems* 36 (2023), pp. 35836–35854.

[280] Zongyi Li et al. "Physics-informed neural operator for learning partial differential equations". In: *ACM/JMS Journal of Data Science* (2021).

[281] Zongyi Li, Samuel Lanthaler, Catherine Deng, Yixuan Wang, Kamyar Azizzadenesheli, and Anima Anandkumar. "Scale-consistent learning with neural operators". In: *Neurips 2024 Workshop Foundation Models for Science: Progress, Opportunities, and Challenges*. 2024.

[282] Henry W Lin, Max Tegmark, and David Rolnick. "Why does deep and cheap learning work so well?" In: *Journal of Statistical Physics* 168 (2017), pp. 1223–1247.

[283] Ji-Nan Lin and Rolf Unbehauen. "On the realization of a Kolmogorov network". In: *Neural Computation* 5.1 (1993), pp. 18–20.

[284] Michael Lindsey, Jonathan Weare, and Anna Zhang. "Ensemble Markov chain Monte Carlo with teleporting walkers". In: *arXiv preprint arXiv:2106.02686* (2021).

[285] Hao Liu, Jin Lei, and Zhongzhou Ren. *From Complexity to Clarity: Kolmogorov-Arnold Networks in Nuclear Binding Energy Prediction*. 2024. arXiv: 2407.20737 [nucl-th]. URL: https://arxiv.org/abs/2407.20737.

[286] Jiaqi Liu, Yixuan Wang, and Tao Zhou. "Finite time blowup for Keller-Segel equation with logistic damping in three dimensions". In: *arXiv preprint arXiv:2504.12231* (2025).

[287] Mengxi Liu et al. "iKAN: Global Incremental Learning with KAN for Human Activity Recognition Across Heterogeneous Datasets". In: *arXiv preprint arXiv:2406.01646* (2024).

[288] Mengxi Liu et al. "Initial Investigation of Kolmogorov-Arnold Networks (KANs) as Feature Extractors for IMU Based Human Activity Recognition". In: *arXiv preprint arXiv:2406.11914* (2024).

[289] Pengfei Liu. *Spatial profiles in the singular solutions of the 3D Euler equations and simplified models*. California Institute of Technology, 2017.

[290] Ziming Liu, Eric Gan, and Max Tegmark. "Seeing is believing: Brain-inspired modular training for mechanistic interpretability". In: *Entropy* 26.1 (2023), p. 41.

[291] Ziming Liu, Andrew M Stuart, and Yixuan Wang. "Second order ensemble Langevin method for sampling and inverse problems". In: *Communications in Mathematical Sciences* 23.5 (2025), pp. 1299–1317. DOI: 10.4310/CMS.250517001811.

[292] Ziming Liu, Pingchuan Ma, Yixuan Wang, Wojciech Matusik, and Max Tegmark. "Kan 2.0: Kolmogorov-arnold networks meet science". In: *Physical Review X* 15.4 (2025), p. 041051. DOI: 10.1103/4t7t-v19l.

[293] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljacic, Thomas Y. Hou, and Max Tegmark. "KAN: Kolmogorov–Arnold Networks". In: *The Thirteenth International Conference on Learning Representations*. 2025. URL: https://openreview.net/forum?id=Ozo7qJ5vZi.

[294] Samuel Livingstone et al. "On the geometric ergodicity of Hamiltonian Monte Carlo". In: *Bernoulli* 25.4A (2019), pp. 3109–3138.

[295] Jianfeng Lu et al. "Deep network approximation for smooth functions". In: *SIAM Journal on Mathematical Analysis* 53.5 (2021), pp. 5465–5506.

[296] Lu Lu et al. "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators". In: *Nature machine intelligence* 3.3 (2021), pp. 218–229.

[297] Lu Lu et al. "Physics-informed neural networks with hard constraints for inverse design". In: *SIAM Journal on Scientific Computing* 43.6 (2021), B1105–B1132.

[298] G Luo and TY Hou. "Toward the finite-time blowup of the 3D incompressible Euler equations: a numerical investigation". In: *SIAM Multiscale Modeling and Simulation* 12.4 (2014), pp. 1722–1776.

[299] Guo Luo and Thomas Y Hou. "Potentially singular solutions of the 3D axisymmetric Euler equations". In: *Proceedings of the National Academy of Sciences* 111.36 (2014), pp. 12968–12973.

[300] Pavel M Lushnikov, Denis A Silantyev, and Michael Siegel. "Collapse Versus Blow-Up and Global Existence in the Generalized Constantin–Lax–Majda Equation". In: *Journal of Nonlinear Science* 31.5 (2021), p. 82.

[301] Yi-An Ma, Tianqi Chen, and Emily Fox. "A complete recipe for stochastic gradient MCMC". In: *Advances in Neural Information Processing Systems*. 2015, pp. 2917–2925.

[302] Chupeng Ma, Christian Alber, and Robert Scheichl. "Wavenumber explicit convergence of a multiscale GFEM for heterogeneous Helmholtz problems". In: *arXiv preprint arXiv:2112.10544* (2021).

[303] Chupeng Ma and Robert Scheichl. "Error estimates for fully discrete generalized FEMs with locally optimal spectral approximations". In: *arXiv preprint arXiv:2107.09988* (2021).

[304] Chupeng Ma, Robert Scheichl, and Tim Dodwell. "Novel design and analysis of generalized FE methods based on locally optimal spectral approximations". In: *arXiv preprint arXiv:2103.09545* (2021).

[305] Roland Maier. "A high-order approach to elliptic multiscale problems with general unstructured coefficients". In: *SIAM Journal on Numerical Analysis* 59.2 (2021), pp. 1067–1089.

[306] AJ Majda and AL Bertozzi. *Vorticity and incompressible flow*. Vol. 27. Cambridge University Press, 2002.

[307] Axel Målqvist and Daniel Peterseim. "Localization of elliptic multiscale problems". en. In: *Mathematics of Computation* 83.290 (June 2014), pp. 2583–2603. ISSN: 0025-5718, 1088-6842. (Visited on 08/23/2019).

[308] Yvan Martel, Frank Merle, and Pierre Raphaël. "Blow up for the critical generalized Korteweg–de Vries equation. I: Dynamics near the soliton". In: *Acta Mathematica* 212.1 (2014), pp. 59–140.

[309] Georg Martius and Christoph H Lampert. "Extrapolation and learning equations". In: *arXiv preprint arXiv:1610.02995* (2016).

[310] Nader Masmoudi and Hatem Zaag. "Blow-up profile for the complex Ginzburg–Landau equation". In: *Journal of Functional Analysis* 255.7 (2008), pp. 1613–1666.

[311] H. Matano and F. Merle. "Classification of type I and type II behaviors for a supercritical nonlinear heat equation". In: *J. Funct. Anal.* 256.4 (2009), pp. 992–1064. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2008.05.021. URL: http://dx.doi.org/10.1016/j.jfa.2008.05.021.

[312] H. Matano and F. Merle. "On nonexistence of type II blowup for a supercritical nonlinear heat equation". In: *Comm. Pure Appl. Math.* 57.11 (2004), pp. 1494–1541. ISSN: 0010-3640. DOI: 10.1002/cpa.20044. URL: http://dx.doi.org/10.1002/cpa.20044.

[313] Haydn Maust, Zongyi Li, Yixuan Wang, Daniel Leibovici, Oscar Bruno, Thomas Hou, and Anima Anandkumar. "Fourier continuation for exact derivative computation in physics-informed neural operators". In: *arXiv preprint arXiv:2211.15960* (2022).

[314] Levi D McClenny and Ulisses M Braga-Neto. "Self-adaptive physics-informed neural networks". In: *Journal of Computational Physics* 474 (2023), p. 111722.

[315] DW McLaughlin et al. “Focusing singularity of the cubic Schrödinger equation”. In: *Physical Review A* 34.2 (1986), p. 1200.

[316] Jens M Melenk. “On n-widths for elliptic problems”. In: *Journal of mathematical analysis and applications* 247.1 (2000), pp. 272–289.

[317] Jens M Melenk and Stefan Sauter. “Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions”. In: *Mathematics of Computation* 79.272 (2010), pp. 1871–1914.

[318] Kevin Meng et al. “Locating and editing factual associations in GPT”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 17359–17372.

[319] F. Merle et al. “On the implosion of a compressible fluid I: Smooth self-similar inviscid profiles”. In: *Annals of Mathematics* 196.2 (2022), pp. 567–778. DOI: 10.4007/annals.2022.196.2.3. URL: https://doi.org/10.4007/annals.2022.196.2.3.

[320] F. Merle et al. “On the implosion of a compressible fluid II: Singularity formation”. In: *Annals of Mathematics* 196.2 (2022), pp. 779–889. DOI: 10.4007/annals.2022.196.2.4. URL: https://doi.org/10.4007/annals.2022.196.2.4.

[321] Frank Merle and Pierre Raphael. “The blow-up dynamic and upper bound on the blow-up rate for critical nonlinear Schrödinger equation”. In: *Annals of mathematics* (2005), pp. 157–222.

[322] Frank Merle, Pierre Raphaël, and Igor Rodnianski. “Type II blow up for the energy supercritical NLS”. In: *Cambridge Journal of Mathematics* 3.4 (2015), pp. 439–617.

[323] Frank Merle, Pierre Raphaël, and Jeremie Szeftel. “Stable self-similar blow-up dynamics for slightly L^2 super-critical NLS equations”. In: *Geom. Funct. Anal.* 20.4 (2010), pp. 1028–1071. ISSN: 1016-443X. DOI: 10.1007/s00039-010-0081-8. URL: https://doi.org/10.1007/s00039-010-0081-8.

[324] Frank Merle and Hatem Zaag. “Stability of the blow-up profile for equations of the type $u_t = \Delta u + |u|^{p-1}u$”. In: *Duke Math. J* 86.1 (1997), pp. 143–195.

[325] Frank Merle et al. “On blow up for the energy super critical defocusing nonlinear Schrödinger equations”. English. In: *Invent. Math.* 227.1 (2022), pp. 247–413. ISSN: 0020-9910. DOI: 10.1007/s00222-021-01067-9.

[326] Eric J Michaud, Ziming Liu, and Max Tegmark. “Precision machine learning”. In: *Entropy* 25.1 (2023), p. 175.

[327] Eric J Michaud et al. “The Quantization Model of Neural Scaling”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023. URL: https://openreview.net/forum?id=3tbTw2ga8K.

[328] Alexander Mielke. “The Ginzburg-Landau equation in its role as a modulation equation”. In: *Handbook of dynamical systems*. Vol. 2. Elsevier, 2002, pp. 759–834.

[329] Alexander Mielke, Philip Holmes, and J Nathan Kutz. “Global existence and uniqueness for an optical fibre laser model”. In: *Nonlinearity* 11.6 (1998), pp. 1489–1504.

[330] Evan Miller. “Finite-time blowup for the inviscid vortex stretching equation”. In: *Nonlinearity* 36.8 (2023), p. 4086.

[331] N. Mizoguchi. “Rate of type II blowup for a semilinear heat equation”. In: *Math. Ann.* 339.4 (2007), pp. 839–877. ISSN: 0025-5831. DOI: 10.1007/s00208-007-0133-z. URL: http://dx.doi.org/10.1007/s00208-007-0133-z.

[332] Arvind T Mohan et al. “Embedding hard physical constraints in neural network coarse-graining of three-dimensional turbulence”. In: *Physical Review Fluids* 8.1 (2023), p. 014604.

[333] Hadrien Montanelli and Haizhao Yang. “Error bounds for deep ReLU networks using the Kolmogorov–Arnold superposition theorem”. In: *Neural Networks* 129 (2020), pp. 1–6.

[334] Johannes Müller and Marius Zeinhofer. “Achieving high accuracy with PINNs via energy natural gradient descent”. In: *International Conference on Machine Learning*. PMLR. 2023, pp. 25471–25485.

[335] Terrell N. Mundhenk et al. “Symbolic Regression via Deep Reinforcement Learning Enhanced Genetic Programming Seeding”. In: *Advances in Neural Information Processing Systems*. Ed. by A. Beygelzimer et al. 2021. URL: https://openreview.net/forum?id=tjwQaOI9tdy.

[336] Toshitaka Nagai. “Behavior of solutions to a parabolic-elliptic system modelling chemotaxis”. English. In: *J. Korean Math. Soc.* 37.5 (2000), pp. 721–733. ISSN: 0304-9914.

[337] Toshitaka Nagai. “Blow-up of radially symmetric solutions to a chemotaxis system”. In: *Adv. Math. Sci. Appl.* 5.2 (1995), pp. 581–601. ISSN: 1343-4373.

[338] Yūki Naito and Takashi Suzuki. “Self-similarity in chemotaxis systems”. In: *Colloq. Math.* 111.1 (2008), pp. 11–34. ISSN: 0010-1354,1730-6302. DOI: 10.4064/cm111-1-2. URL: https://doi.org/10.4064/cm111-1-2.

[339] Neel Nanda et al. “Progress measures for grokking via mechanistic interpretability”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: https://openreview.net/forum?id=9XFSbDPmdW.

[340] Radford M Neal. “An improved acceptance procedure for the hybrid Monte Carlo algorithm”. In: *Journal of Computational Physics* 111.1 (1994), pp. 194–203.

[341] J Necas, M Ruzicka, and V Sverak. “On Leray’s self-similar solutions of the Navier-Stokes equations”. In: (1996).

[342] George Nehma and Madhur Tiwari. “Leveraging KANs For Enhanced Deep Koopman Operator Discovery”. In: *arXiv preprint arXiv:2406.02875* (2024).

[343] AC Newell and JA Whitehead. “Review of the finite bandwidth concept”. In: *Instability of Continuous Systems: Symposium Herrenalb (Germany) September 8–12, 1969*. Springer. 1971, pp. 284–289.

[344] Alan C Newell and John A Whitehead. “Finite bandwidth, finite amplitude convection”. In: *Journal of Fluid Mechanics* 38.2 (1969), pp. 279–303.

[345] Jakin Ng, Yongji Wang, and Ching-Yao Lai. “Spectrum-Informed Multistage Neural Networks: Multiscale Function Approximators of Machine Precision”. In: *arXiv preprint arXiv:2407.17213* (2024).

[346] Van Tien Nguyen, Zhi-An Wang, and Kaiqiang Zhang. “Infinitely many self-similar blow-up profiles for the Keller-Segel system in dimensions 3 to 9”. In: *arXiv preprint arXiv:2503.02263* (2025).

[347] VT Nguyen, N Nouaili, and H Zaag. “Construction of type I-Log blowup for the Keller-Segel system in dimensions 3 and 4”. In: *to appear in Annal of PDE* (2023). Available at arXiv:2309.13932.

[348] Nejla Nouaili and Hatem Zaag. “Construction of a blow-up solution for the complex Ginzburg–Landau equation in a critical case”. In: *Archive for Rational Mechanics and Analysis* 228.3 (2018), pp. 995–1058.

[349] Tiago Novello, Diana Aldana, and Luiz Velho. “Taming the Frequency Factory of Sinusoidal Networks”. In: *arXiv preprint arXiv:2407.21121* (2024).

[350] Nikolas Nüsken and Sebastian Reich. “Note on interacting Langevin diffusions: Gradient structure and ensemble Kalman sampler by Garbuno-Inigo, Hoffmann, Li and Stuart”. In: *arXiv preprint arXiv:1908.10890* (2019).

[351] Takayoshi Ogawa and Hiroshi Wakui. “Non-uniform bound and finite time blow up for solutions to a drift–diffusion equation in higher dimensions”. In: *Anal. Appl. (Singap.)* 14.1 (2016), pp. 145–183. ISSN: 0219-5305,1793-6861. DOI: 10.1142/S0219530515400060. URL: https://doi.org/10.1142/S0219530515400060.

[352] Sung-Jin Oh and Federico Pasqualotto. “Gradient blow-up for dispersive and dissipative perturbations of the Burgers equation”. In: *Arch. Ration. Mech. Anal.* 248.3 (2024), Paper No. 54, 61. ISSN: 0003-9527,1432-0673. DOI: 10.1007/s00205-024-01985-x. URL: https://doi.org/10.1007/s00205-024-01985-x.

[353] Hisashi Okamoto, Takashi Sakajo, and Marcus Wunsch. “On a generalization of the Constantin–Lax–Majda equation”. In: *Nonlinearity* 21.10 (2008), p. 2447.

[354] Catherine Olsson et al. “In-context learning and induction heads”. In: *arXiv preprint arXiv:2209.11895* (2022).

[355] Michela Ottobre and GA Pavliotis. “Asymptotic analysis for the generalized Langevin equation”. In: *Nonlinearity* 24.5 (2011), p. 1629.

[356] Houman Owhadi. “Bayesian numerical homogenization”. In: *Multiscale Modeling & Simulation* 13.3 (2015), pp. 812–828.

[357] Houman Owhadi. “Multigrid with Rough Coefficients and Multiresolution Operator Decomposition from Hierarchical Information Games”. en. In: *SIAM Review* 59.1 (Jan. 2017), pp. 99–149. ISSN: 0036-1445, 1095-7200. (Visited on 11/19/2018).

[358] Houman Owhadi and Clint Scovel. *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*. Vol. 35. Cambridge University Press, 2019.

[359] Houman Owhadi and Lei Zhang. “Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast”. In: *Multiscale Modeling & Simulation* 9.4 (2011), pp. 1373–1398.

[360] Houman Owhadi and Lei Zhang. “Metric-based upscaling”. In: *Communications on Pure and Applied Mathematics* 60.5 (2007), pp. 675–723.

[361] Houman Owhadi, Lei Zhang, and Leonid Berlyand. “Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 48.2 (2014), pp. 517–552.

[362] K. J. Painter, P. K. Maini, and H. G. Othmer. “Development and applications of a model for cellular response to multiple chemotactic cues”. In: *J. Math. Biol.* 41.4 (2000), pp. 285–314. ISSN: 0303-6812,1432-1416. DOI: 10.1007/s002850000035. URL: https://doi.org/10.1007/s002850000035.

[363] Adam Paszke et al. “Pytorch: An imperative style, high-performance deep learning library”. In: *Advances in neural information processing systems* 32 (2019).

[364] Grigorios A Pavliotis. *Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations*. Vol. 60. Springer, 2014.

[365] Benjamin Peherstorfer. “Breaking the Kolmogorov Barrier with Nonlinear Model Reduction”. In: *Notices of the American Mathematical Society* 69.5 (2022), pp. 725–733.

[366] Yanhong Peng et al. “Predictive Modeling of Flexible EHD Pumps using Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2405.07488* (2024).

[367] P. Petersen. *Riemannian Geometry*. Graduate Texts in Mathematics. Springer New York, 2006. ISBN: 9780387294032. URL: https://books.google.com/books?id=9cekXdo52hEC.

[368] Allan Pinkus. *N-widths in Approximation Theory*. Vol. 7. Springer Science & Business Media, 2012.

[369] M. del Pino, M. Musso, and J. Wei. "Geometry driven type II higher dimensional blow-up for the critical heat equation". In: *J. Funct. Anal.* 280.1 (2021), Paper No. 108788, 49. ISSN: 0022-1236,1096-0783. DOI: 10.1016/j.jfa.2020.108788. URL: https://doi.org/10.1016/j.jfa.2020.108788.

[370] M. del Pino et al. "Type II finite time blow-up for the energy critical heat equation in $\mathbb{R}^4$". In: *Discrete Contin. Dyn. Syst.* 40.6 (2020), pp. 3327–3355. ISSN: 1078-0947,1553-5231. DOI: 10.3934/dcds.2020052. URL: https://doi.org/10.3934/dcds.2020052.

[371] M. del Pino et al. "Type II finite time blow-up for the three dimensional energy critical heat equation". In: *Preprint* (2020). URL: arXiv:2002.05765.

[372] Manuel del Pino, Monica Musso, and Jun Cheng Wei. "Type II blow-up in the 5-dimensional energy critical heat equation". In: *Acta Math. Sin. (Engl. Ser.)* 35.6 (2019), pp. 1027–1042. ISSN: 1439-8516,1439-7617. DOI: 10.1007/s10114-019-8341-5. URL: https://doi.org/10.1007/s10114-019-8341-5.

[373] Petr Plecháč and Vladimír Šverák. "On self-similar singular solutions of the complex Ginzburg-Landau equation". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 54.10 (2001), pp. 1215–1242.

[374] Tomaso Poggio. "How deep sparse networks avoid the curse of dimensionality: Efficiently computable functions are compositionally sparse". In: *CBMM Memo* 10 (2022), p. 2022.

[375] Tomaso Poggio, Andrzej Banburski, and Qianli Liao. "Theoretical issues in deep networks". In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30039–30045.

[376] Tomaso Poggio et al. "Theory of deep learning iii: the non-overfitting puzzle". In: *CBMM Memo* 73.1-38 (2018), p. 2.

[377] Michael Poluektov and Andrew Polar. "A new iterative method for construction of the Kolmogorov-Arnold representation". In: *arXiv preprint arXiv:2305.08194* (2023).

[378] Pawel Pratyush et al. “CaLMPhosKAN: Prediction of General Phosphorylation Sites in Proteins via Fusion of Codon Aware Embeddings with Amino Acid Aware Embeddings and Wavelet-based Kolmogorov Arnold Network”. In: *bioRxiv* (2024), pp. 2024–07.

[379] Pavol Quittner. “Optimal Liouville theorems for superlinear parabolic problems”. In: *Duke Math. J.* 170.6 (2021), pp. 1113–1136. ISSN: 0012-7094,1547-7398. DOI: 10.1215/00127094-2020-0096. URL: https://doi.org/10.1215/00127094-2020-0096.

[380] Pavol Quittner and Philippe Souplet. *Superlinear parabolic problems*. Springer, 2019.

[381] Nasim Rahaman et al. “On the spectral bias of neural networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 5301–5310.

[382] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *Journal of Computational physics* 378 (2019), pp. 686–707.

[383] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. “Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations”. In: *Science* 367.6481 (2020), pp. 1026–1030.

[384] Prajit Ramachandran, Barret Zoph, and Quoc V Le. “Searching for activation functions”. In: *arXiv preprint arXiv:1710.05941* (2017).

[385] Pierre Raphaël and Rémi Schweyer. “On the stability of critical chemotactic aggregation”. In: *Mathematische Annalen* 359 (2014), pp. 267–377.

[386] Pratik Rathore et al. “Challenges in Training PINNs: A Loss Landscape Perspective”. In: *Forty-first International Conference on Machine Learning*. 2024.

[387] Pratik Rathore et al. “Challenges in training PINNs: A loss landscape perspective”. In: *arXiv preprint arXiv:2402.01868* (2024).

[388] Sebastian Reich. “A dynamical systems framework for intermittent data assimilation”. In: *BIT Numerical Mathematics* 51.1 (2011), pp. 235–249.

[389] Jack Richter-Powell, Yaron Lipman, and Ricky TQ Chen. “Neural conservation laws: A divergence-free perspective”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 38075–38088.

[390] Spyros Rigas et al. “Adaptive Training of Grid-Dependent Physics-Informed Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2407.17611* (2024).

[391] Spyros Rigas, Dhruv Verma, Georgios Alexandridis, and Yixuan Wang. “Initialization schemes for Kolmogorov-Arnold networks: An empirical study”. In: *The Fourteenth International Conference on Learning Representations*. 2026. URL: https://openreview.net/forum?id=dwNXKkiP51.

[392] Hannes Risken. “Fokker-planck equation”. In: *The Fokker-Planck Equation*. Springer, 1996, pp. 63–95.

[393] Gareth O Roberts and Jeffrey S Rosenthal. “Optimal scaling for various Metropolis-Hastings algorithms”. In: *Statistical science* 16.4 (2001), pp. 351–367.

[394] Basri Ronen et al. “The convergence rate of neural networks for learned functions of different frequencies”. In: *Advances in Neural Information Processing Systems* 32 (2019).

[395] Vivi Rottschäfer. “Asymptotic analysis of a new type of multi-bump, self-similar, blowup solutions of the Ginzburg–Landau equation”. In: *European Journal of Applied Mathematics* 24.1 (2013), pp. 103–129.

[396] Fabian Ruehle. “Data science applications to string theory”. In: *Phys. Rept.* 839 (2020), pp. 1–117. DOI: 10.1016/j.physrep.2019.09.005.

[397] Jesus-Maria Sanz-Serna and Mari-Paz Calvo. *Numerical Hamiltonian Problems*. Courier Dover Publications, 2018.

[398] Alejandro Sarria and Ralph Saxton. “Blow-up of solutions to the generalized inviscid Proudman–Johnson equation”. In: *Journal of Mathematical Fluid Mechanics* 15.3 (2013), pp. 493–523.

[399] Arnd Scheel. “Bifurcation to spiral waves in reaction-diffusion systems”. In: *SIAM journal on mathematical analysis* 29.6 (1998), pp. 1399–1418.

[400] Claudia Schillings and Andrew M Stuart. “Analysis of the ensemble Kalman filter for inverse problems”. In: *SIAM Journal on Numerical Analysis* 55.3 (2017), pp. 1264–1290.

[401] Julia Schleuß and Kathrin Smetana. “Optimal local approximation spaces for parabolic problems”. In: *arXiv preprint arXiv:2012.02759* (2020).

[402] Johannes Schmidt-Hieber. “Nonparametric regression using deep neural networks with ReLU activation function”. In: *The Annals of Statistics* 48.4 (2020), p. 1875.

[403] Johannes Schmidt-Hieber. “The Kolmogorov–Arnold representation theorem revisited”. In: *Neural networks* 137 (2021), pp. 119–126.

[404] Guido Schneider. “Hopf bifurcation in spatially extended reaction—diffusion systems”. In: *Journal of Nonlinear Science* 8.1 (1998), pp. 17–41.

[405] R. Schweyer. “Type II blow-up for the four dimensional energy critical semi linear heat equation”. In: *J. Funct. Anal.* 263.12 (2012), pp. 3922–3983. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2012.09.015. URL: http://dx.doi.org/10.1016/j.jfa.2012.09.015.

[406] Seyd Teymoor Seydi. “Exploring the Potential of Polynomial Basis Functions in Kolmogorov-Arnold Networks: A Comparative Study of Different Groups of Polynomials”. In: *arXiv preprint arXiv:2406.02583* (2024).

[407] Seyd Teymoor Seydi. "Unveiling the Power of Wavelets: A Wavelet-based Kolmogorov-Arnold Network for Hyperspectral Image Classification". In: *arXiv preprint arXiv:2406.07869* (2024).

[408] Utkarsh Sharma and Jared Kaplan. "A neural scaling law from the dimension of the data manifold". In: *arXiv preprint arXiv:2004.10802* (2020).

[409] Zuowei Shen, Haizhao Yang, and Shijun Zhang. "Optimal approximation rate of ReLU networks in terms of width and depth". In: *Journal de Mathématiques Pures et Appliquées* 157 (2022), pp. 101–135.

[410] Nanako Shigesada, Kohkichi Kawasaki, and Ei Teramoto. "Spatial segregation of interacting species". In: *J. Theoret. Biol.* 79.1 (1979), pp. 83–99. ISSN: 0022-5193,1095-8541. DOI: 10.1016/0022-5193(79)90258-3. URL: https://doi.org/10.1016/0022-5193(79)90258-3.

[411] Khemraj Shukla et al. "A comprehensive and FAIR comparison between MLP and KAN representations for differential equations and operator networks". In: *arXiv preprint arXiv:2406.02917* (2024).

[412] Jonathan W Siegel. "Optimal approximation rates for deep ReLU neural networks on Sobolev and Besov spaces". In: *Journal of Machine Learning Research* 24.357 (2023), pp. 1–52.

[413] Jonathan W Siegel and Jinchao Xu. "Approximation rates for neural networks with general activation functions". In: *Neural Networks* 128 (2020), pp. 313–321.

[414] Jonathan W Siegel and Jinchao Xu. "Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks". In: *Foundations of Computational Mathematics* (2022), pp. 1–57.

[415] Vincent Sitzmann et al. "Implicit neural representations with periodic activation functions". In: *Advances in neural information processing systems* 33 (2020), pp. 7462–7473.

[416] Kathrin Smetana and Anthony T Patera. "Optimal local approximation spaces for component-based static condensation procedures". In: *SIAM Journal on Scientific Computing* 38.5 (2016), A3318–A3356.

[417] A. Soffer. "Soliton dynamics and scattering". In: *International congress of mathematicians*. Vol. 3. 2006, pp. 459–471.

[418] Avy Soffer and Michael I Weinstein. "Multichannel nonlinear scattering for nonintegrable equations". In: *Communications in mathematical physics* 133 (1990), pp. 119–146.

[419] Huan Song et al. "Optimizing kernel machines using deep learning". In: *IEEE transactions on neural networks and learning systems* 29.11 (2018), pp. 5528–5540.

[420] Jinyeop Song et al. “A Resource Model For Neural Scaling Law”. In: *arXiv preprint arXiv:2402.05164* (2024).

[421] Philippe Souplet and Michael Winkler. “Blow-up profiles for the parabolic-elliptic Keller-Segel system in dimensions $n \geq 3$”. In: *Comm. Math. Phys.* 367.2 (2019), pp. 665–681. ISSN: 0010-3616,1432-0916. DOI: 10.1007/s00220-018-3238-1. URL: https://doi.org/10.1007/s00220-018-3238-1.

[422] David A Sprecher and Sorin Draghici. “Space-filling curves and Kolmogorov superposition-based neural networks”. In: *Neural Networks* 15.1 (2002), pp. 57–67.

[423] Sidharth SS. “Chebyshev polynomial-based kolmogorov-arnold networks: An efficient architecture for nonlinear function approximation”. In: *arXiv preprint arXiv:2405.07200* (2024).

[424] K Stewartson and JT Stuart. “A non-linear instability theory for a wave system in plane Poiseuille flow”. In: *Journal of Fluid Mechanics* 48.3 (1971), pp. 529–545.

[425] Alain-Sol Sznitman. “Topics in propagation of chaos”. In: *Ecole d’été de probabilités de Saint-Flour XIX—1989*. Springer, 1991, pp. 165–251.

[426] Hoang-Thang Ta. “BSRBF-KAN: A combination of B-splines and Radial Basic Functions in Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2406.11173* (2024).

[427] Amirhossein Taghvaei et al. “Kalman filter and its modern extensions for the continuous-time nonlinear filtering problem”. In: *Journal of Dynamic Systems, Measurement, and Control* 140.3 (2018).

[428] Makoto Takamoto et al. “Pdebench: An extensive benchmark for scientific machine learning”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 1596–1611.

[429] Matthew Tancik et al. “Fourier features let networks learn high frequency functions in low dimensional domains”. In: *Advances in neural information processing systems* 33 (2020), pp. 7537–7547.

[430] Michael Eugene Taylor. *Measure theory and integration*. American Mathematical Soc., 2006.

[431] J. Ignacio Tello and Michael Winkler. “A chemotaxis system with logistic source”. In: *Comm. Partial Differential Equations* 32.4-6 (2007), pp. 849–877. ISSN: 0360-5302,1532-4133. DOI: 10.1080/03605300701319003. URL: https://doi.org/10.1080/03605300701319003.

[432] Juan Diego Toscano et al. “From pinns to pikans: Recent advances in physics-informed machine learning”. In: *Machine Learning for Computational Science and Engineering* 1.1 (2025), pp. 1–43.

[433] Tai-Peng Tsai. "On Leray's self-similar solutions of the Navier-Stokes equations satisfying local energy estimates". In: *Archive for rational mechanics and analysis* 143.1 (1998), pp. 29–51.

[434] Sergei K Turitsyn. "Nonstable solitons and sharp criteria for wave collapse". In: *Physical Review E* 47.1 (1993), R13.

[435] Silviu-Marian Udrescu and Max Tegmark. "AI Feynman: A physics-inspired method for symbolic regression". In: *Science Advances* 6.16 (2020), eaay2631.

[436] Silviu-Marian Udrescu et al. "AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 4860–4871.

[437] Jorge F Urbán, Petros Stefanou, and José A Pons. "Unveiling the optimization process of Physics Informed Neural Networks: How accurate and competitive can PINNs be?" In: *Journal of Computational Physics* 523 (2025), p. 113656.

[438] Cristian J Vaca-Rubio et al. "Kolmogorov-arnold networks (kans) for time series analysis". In: *arXiv preprint arXiv:2405.08790* (2024).

[439] JJL Velázquez. "Higher dimensional blow up for semilinear parabolic equations". In: *Communications in partial differential equations* 17.9-10 (1992), pp. 1567–1596.

[440] Cédric Villani. "Hypocoercivity". In: *arXiv preprint math/0609050* (2006).

[441] Kevin Ro Wang et al. "Interpretability in the Wild: a Circuit for Indirect Object Identification in GPT-2 Small". In: *The Eleventh International Conference on Learning Representations*. 2023. URL: https://openreview.net/forum?id=NpsVSN6o4ul.

[442] Sifan Wang, Yujun Teng, and Paris Perdikaris. "Understanding and mitigating gradient flow pathologies in physics-informed neural networks". In: *SIAM Journal on Scientific Computing* 43.5 (2021), A3055–A3081.

[443] Sifan Wang, Xinling Yu, and Paris Perdikaris. "When and why PINNs fail to train: A neural tangent kernel perspective". In: *Journal of Computational Physics* 449 (2022), p. 110768.

[444] Sifan Wang et al. "Gradient Alignment in Physics-informed Neural Networks: A Second-Order Optimization Perspective". In: *arXiv preprint arXiv:2502.00604* (2025).

[445] Yixuan Wang, Ziming Liu, Zongyi Li, Anima Anandkumar, and Thomas Y Hou. "High precision PINNs in unbounded domains: application to singularity formulation in PDEs". In: *arXiv preprint arXiv:2506.19243* (2025).

[446] Yixuan Wang, Jonathan W. Siegel, Ziming Liu, and Thomas Y. Hou. "On the expressiveness and spectral bias of KANs". In: *The Thirteenth International Conference on Learning Representations*. 2025. URL: https://openreview.net/forum?id=ydlDRUuGm9.

[447] Yizheng Wang et al. "Kolmogorov Arnold Informed neural network: A physics-informed deep learning framework for solving PDEs based on Kolmogorov Arnold Networks". In: *arXiv preprint arXiv:2406.11045* (2024).

[448] Yongji Wang and Ching-Yao Lai. "Multi-stage neural networks: Function approximator of machine precision". In: *Journal of Computational Physics* 504 (2024), p. 112865.

[449] Yongji Wang et al. "Asymptotic self-similar blow-up profile for three-dimensional axisymmetric Euler equations using neural networks". In: *Physical Review Letters* 130.24 (2023), p. 244002.

[450] Yongji Wang et al. "Asymptotic self-similar blow-up profile for three-dimensional axisymmetric Euler equations using neural networks". In: *Physical Review Letters* 130.24 (2023), p. 244002.

[451] Michael I Weinstein. "Modulational stability of ground states of nonlinear Schrödinger equations". In: *SIAM journal on mathematical analysis* 16.3 (1985), pp. 472–491.

[452] Michael Winkler. "Blow-up in a higher-dimensional chemotaxis system despite logistic growth restriction". In: *J. Math. Anal. Appl.* 384.2 (2011), pp. 261–272. ISSN: 0022-247X,1096-0813. DOI: 10.1016/j.jmaa.2011.05.057. URL: https://doi.org/10.1016/j.jmaa.2011.05.057.

[453] Michael Winkler. "Finite-time blow-up in low-dimensional Keller-Segel systems with logistic-type superlinear degradation". In: *Z. Angew. Math. Phys.* 69.2 (2018), Paper No. 69, 40. ISSN: 0044-2275,1420-9039. DOI: 10.1007/s00033-018-0935-8. URL: https://doi.org/10.1007/s00033-018-0935-8.

[454] Chenxi Wu et al. "A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks". In: *Computer Methods in Applied Mechanics and Engineering* 403 (2023), p. 115671.

[455] Zixue Xiang et al. "Self-adaptive loss balanced physics-informed neural networks". In: *Neurocomputing* 496 (2022), pp. 11–34.

[456] Hongyi Xu et al. "Nonlinear material design using principal stretches". In: *ACM Transactions on Graphics (TOG)* 34.4 (2015), pp. 1–11.

[457] Jinfeng Xu et al. "FourierKAN-GCF: Fourier Kolmogorov-Arnold Network–An Effective and Efficient Feature Transformation for Graph Collaborative Filtering". In: *arXiv preprint arXiv:2406.01034* (2024).

[458] Kunpeng Xu, Lifei Chen, and Shengrui Wang. “Kolmogorov-Arnold Networks for Time Series: Bridging Predictive Power and Interpretability”. In: *arXiv preprint arXiv:2406.02496* (2024).

[459] Zhi-Qin John Xu, Yaoyu Zhang, and Yanyang Xiao. “Training behavior of deep neural network in frequency domain”. In: *Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part I 26*. Springer. 2019, pp. 264–274.

[460] Zhi-Qin John Xu et al. “Frequency principle: Fourier analysis sheds light on deep neural networks”. In: *arXiv preprint arXiv:1901.06523* (2019).

[461] Shangshang Yang, Linrui Qin, and Xiaoshan Yu. “Endowing Interpretability for Neural Cognitive Diagnosis by Efficient Kolmogorov-Arnold Networks”. In: *arXiv preprint arXiv:2405.14399* (2024).

[462] Yahong Yang and Yulong Lu. “Optimal deep neural network approximation for Korobov functions with respect to Sobolev norms”. In: *arXiv preprint arXiv:2311.04779* (2023).

[463] Yahong Yang et al. “Nearly optimal approximation rates for deep super ReLU networks on Sobolev spaces”. In: *arXiv preprint arXiv:2310.10766* (2023).

[464] Dmitry Yarotsky. “Error bounds for approximations with deep ReLU networks”. In: *Neural Networks* 94 (2017), pp. 103–114.

[465] Dmitry Yarotsky. “Optimal approximation of continuous functions by very deep ReLU networks”. In: *Conference on Learning Theory*. PMLR. 2018, pp. 639–649.

[466] Dmitry Yarotsky and Anton Zhevnerchuk. “The phase diagram of approximation rates for deep neural networks”. In: *Advances in neural information processing systems* 33 (2020), pp. 13005–13015.

[467] Bing Yu et al. “The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems”. In: *Communications in Mathematics and Statistics* 6.1 (2018), pp. 1–12.

[468] Jeremy Yu et al. “Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems”. In: *Computer Methods in Applied Mechanics and Engineering* 393 (2022), p. 114823.

[469] Lei Yuan et al. “A-PINN: Auxiliary physics informed neural networks for forward and inverse problems of nonlinear integro-differential equations”. In: *Journal of Computational Physics* 462 (2022), p. 111260.

[470] Hatem Zaag. “Blow-up results for vector-valued nonlinear heat equations with no gradient structure”. In: *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*. Vol. 15. 5. Elsevier. 1998, pp. 581–622.

[471] Manzil Zaheer et al. “Deep sets”. In: *Advances in neural information processing systems* 30 (2017).

[472] Chiyuan Zhang et al. “Understanding deep learning (still) requires rethinking generalization”. In: *Communications of the ACM* 64.3 (2021), pp. 107–115.

[473] Fan Zhang and Xin Zhang. “GraphKAN: Enhancing Feature Extraction with Graph Kolmogorov Arnold Networks”. In: *arXiv preprint arXiv:2406.13597* (2024).

[474] Shijun Zhang, Zuowei Shen, and Haizhao Yang. “Neural network architecture beyond width and depth”. In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 5669–5681.

[475] Shijun Zhang et al. “Why shallow networks struggle with approximating and learning high frequency: A numerical study”. In: *arXiv preprint arXiv:2306.17301* (2023).

[476] Xiao Zhang, Haoyi Xiong, and Dongrui Wu. “Rethink the connections among generalization, memorization and the spectral bias of DNNs”. In: *arXiv preprint arXiv:2004.13954* (2020).

[477] Ziqian Zhong et al. “The Clock and the Pizza: Two Stories in Mechanistic Explanation of Neural Networks”. In: *Thirty-seventh Conference on Neural Information Processing Systems*. 2023. URL: https://openreview.net/forum?id=S5wmbQc1We.

Appendix A

BLOWUP ANALYSIS FOR A QUASI-EXACT 1D MODEL OF 3D EULER AND NAVIER-STOKES

We study the singularity formation of a quasi-exact 1D model proposed by Hou-Li in [215]. This model is based on an approximation of the axisymmetric Navier-Stokes equations in the r direction. The solution of the 1D model can be used to construct an exact solution of the original 3D Euler and Navier-Stokes equations if the initial angular velocity, angular vorticity, and angular stream function are linear in r. This model shares many intrinsic properties similar to those of the 3D Euler and Navier-Stokes equations. It captures the competition between advection and vortex stretching as in the 1D De Gregorio [114, 113] model. We show that the inviscid model with weakened advection and smooth initial data or the original 1D model with Hölder continuous data develops a self-similar blowup. We also show that the viscous model with weakened advection and smooth initial data develops a finite time blowup. To obtain sharp estimates for the nonlocal terms, we perform an exact computation for the low-frequency Fourier modes and extract damping in leading order estimates for the high-frequency modes using singularly weighted norms in the energy estimates. The analysis for the viscous case is more subtle since the viscous terms produce some instability if we just use singular weights. We establish the blowup analysis for the viscous model by carefully designing an energy norm that combines a singularly weighted energy norm and a sum of high-order Sobolev norms.

A.1 Introduction

Whether the 3D incompressible Euler and Navier-Stokes equations can develop a finite time singularity from smooth initial data is one of the most outstanding open questions in nonlinear partial differential equations. An essential difficulty is that the vortex stretching term has a quadratic nonlinearity in terms of vorticity. A simplified 1D model was proposed by the Constantin-Lax-Majda model (CLM model for short) [99] to capture the effect of nonlocal vortex stretching. The CLM model can be solved explicitly and can develop a finite time singularity from smooth initial data. Later on, De Gregorio incorporated the advection term into the CLM model to study the competition between advection and vortex stretching

[114, 113], see [98] for singularity formation in the distorted Euler equations with transport neglected and also [215, 214] for a related study on the stabilizing effect of advection for the 3D Euler and Navier-Stokes equations. There have been recent studies on the effect of advection and vortex stretching in other related models; see [398] for the generalized inviscid Proudman-Johnson equation, [139] with a Riesz transform added to the vorticity formulation of 2D Euler equation, and [330] with advection term dropped in the vorticity formulation of 3D Euler equation. In [353], Okamoto, Sakajo, and Wunsch further introduced a parameter for the advection term to measure the relative strength of the advection in the De Gregorio model. These simplified 1D models have inspired many subsequent studies. Interested readers may consult the excellent surveys [97, 162, 249, 306] and the references therein. Very recently, the authors in [225] established self-similar blowup for the whole family of gCLM models with $a \leq 1$ using a fixed-point argument. On the other hand, these 1D scalar models are phenomenological in nature and cannot be used to recover the solution of the original 3D Euler equations.

For the line of research on the singularity formation for the 3D Euler equations, Luo-Hou [298] presented in 2014 convincing numerical evidence that the 3D axisymmetric Euler equations with smooth initial data and boundary develop a potential finite time singularity. Inspired by Elgindi's recent breakthrough for finite time singularity of the axisymmetric Euler with no swirl and $C^{1,\alpha}$ velocity [135], Chen and Hou proved the finite time blowup of the 2D Boussinesq and 3D Euler equations with $C^{1,\alpha}$ initial velocity and boundary [72]. For other recent works on singularity formation of 3D Euler with limited regularity, see also [69, 101] for initial data that is smooth except at the origin, [100] for more smooth data but with a $C^{1/2-\epsilon}$ force, and [138, 136] for settings with nonsmooth boundary. Very recently, Chen and Hou proved stable and nearly self-similar blowup of the 2D Boussinesq and 3D Euler with smooth initial data and boundary using computer assistance [73].

In 2008, Hou and Li [215] proposed a new 1D model for the 3D axisymmetric Euler and Navier-Stokes equations. This model approximates the 3D axisymmetric Euler and Navier-Stokes equations along the symmetry axis based on an approximation in the r direction. The solution of the 1D model can be used to construct an exact solution of the original 3D Navier-Stokes equations if the initial angular velocity, angular vorticity, and angular stream function are linear in r. This model shares many intrinsic properties similar to those of the 3D Navier-Stokes equations. Thus, it captures some essential nonlinear features of the 3D Euler and Navier-Stokes

equations. In the same paper [215], the authors proved the global regularity of the Hou-Li model by deriving a new Lyapunov functional, which captures the exact cancellation between advection and vortex stretching.

The purpose of this chapter is to study the singularity formation of a weak advection version of the Hou-Li model for smooth data. We introduce a parameter a to characterize the relative strength between advection and vortex stretching, just like the gCLM model. Both inviscid and viscous cases are considered. We also prove the finite time singularity formation of the original inviscid Hou-Li model ($a = 1$ and $\nu = 0$) with C^α initial data. Inspired by the recent work of Chen [67] for the De Gregorio model, we consider the case of $a < 1$ and treat $1 - a$ as a small parameter. For the C^α initial data, we consider the original Hou-Li model with $a = 1$ and $1 - \alpha$ small. By using the dynamic rescaling formulation and analyzing the stability of the linearized operator around an approximate steady state of the original Hou-Li model ($a = 1$), we prove finite time self-similar blowup.

We follow a general strategy that we have established in our previous works [76, 72]. Establishing linear stability of the approximate steady state is the most crucial step in our blowup analysis. To obtain sharp estimates for the nonlocal terms, we carry out an exact computation for the low-frequency Fourier modes and extract damping in leading order estimates for the high-frequency modes using singularly weighted norms in the energy estimates. The blowup analysis for the viscous model is more subtle since the viscous terms do not provide damping and produce some bad terms if we use a singularly weighted norm. We establish the blowup analysis for the viscous model by carefully designing an energy norm that combines a singularly weighted energy norm and a sum of high-order Sobolev norms.

A.1.1 Problem setting

In [215], Hou-Li introduced the following reformulation of the axisymmetric Navier-Stokes equation:

$$u_{1,t} + u^r u_{1,r} + u^z u_{1,z} = 2u_1\psi_{1,z} + \nu\Delta u_1\,, \tag{A.1.1}$$

$$\omega_{1,t} + u^r \omega_{1,r} + u^z \omega_{1,z} = \left(u_1^2\right)_z + \nu\Delta\omega_1\,, \tag{A.1.2}$$

$$-\left[\partial_r^2 + (3/r)\partial_r + \partial_z^2\right]\psi_1 = \omega_1\,, \tag{A.1.3}$$

where $u_1 = u^\theta/r,\ \omega_1 = \omega^\theta,\ \psi_1 = \psi^\theta/r$, and u^θ, ω^θ, and ψ^θ are the angular velocity, angular vorticity, and angular stream function, respectively. By the well-known Caffarelli-Kohn-Nirenberg partial regularity result [53], the axisymmetric Navier-

Stokes equations can develop a finite time singularity only along the symmetry axis $r = 0$. To study the potential singularity or global regularity of the axisymmetric Navier-Stokes equations, Hou-Li [215] proposed the following 1D model along the symmetry axis $r = 0$:

$$\begin{aligned} u_{1,t} + 2\psi_1 u_{1,z} &= 2\psi_{1,z} u_1 + \nu u_{1,zz}\,, \\ \omega_{1,t} + 2\psi_1 \omega_{1,z} &= \left(u_1^2\right)_z + \nu \omega_{1,zz}\,, \\ -\psi_{1,zz} &= \omega_1\,. \end{aligned} \tag{A.1.4}$$

Such a reduction is exact in the sense that if $(\omega_1,\, u_1,\, \psi_1)$ is an exact solution of the 1D model, we can obtain an exact solution of the 3D Navier-Stokes equations by using a constant extension in r. This corresponds to the case when the physical quantities $u^\theta = r u_1$, $\omega^\theta = r\omega_1$ are linear in r. We assume that the solutions are periodic in z on $[0, 2\pi]$. We already know from the original Hou-Li paper that this system is well-posed for C^m initial data with $m \geq 1$. In [215], the authors also used the well-posedness of the Hou-Li model to construct globally smooth solutions to the 3D equations with large dynamic growth.

In two recent papers by Hou [212, 213], the author presented new numerical evidence that the 3D axisymmetric Euler and Navier-Stokes equations develop potential singular solutions at the origin. This new blowup scenario is very different from the Hou-Luo blowup scenario, which occurs on the boundary. In this computation, the author observed that the axial velocity $u^z = 2\psi_1 + r\psi_{1,r}$ near the maximal point of u_1 is significantly weaker than $2\psi_1$. This is due to the fact that ψ_1 reaches the maximum at a position $r = r_\psi$ that is smaller than the position $r = r_u$ in which u_1 achieves its maximum, i.e. $r_\psi < r_u$. Therefore $\psi_{1,r}$ is negative near the maximal position of u_1. Thus the axial velocity u^z is actually weaker than $2\psi_1$, which corresponds to $u^z|_{r=0}$. Thus, the original Hou-Li model along $r = 0$ does not capture this subtle phenomenon, which is three-dimensional in nature. To gain some understanding of this potentially singular behavior, we introduce the following 1D weak advection model.

$$\begin{aligned} u_t + 2a\psi u_z &= 2u\psi_z + \nu u_{zz}\,, \\ \omega_t + 2a\psi \omega_z &= \left(u^2\right)_z + \nu \omega_{zz}\,, \\ -\psi_{zz} &= \omega\,, \end{aligned} \tag{A.1.5}$$

where a is a parameter that measures the relative strength of advection in the Hou-Li model.

Remark A.1.1. *For simplicity, we drop the subscript* 1 *in the above weak advection model. The proposed model* (A.1.5) *in the inviscid case* $\nu = 0$ *resembles the generalized Constantin-Lax-Majda model (gCLM) [353]*

$$\omega_t + au\omega_x = u_x\omega\,, \qquad u_x = H\omega\,,$$

where

$$H\omega(x) = \frac{1}{\pi} p.v. \int_{\mathbb{R}} \frac{\omega(y)}{x-y}\,\mathrm{d}y$$

is the Hilbert transform. They share similar structures of competition between advection and vortex stretching. The case when $a = 1$ *corresponds to the De Gregorio (DG) model. We obtain an explicit steady-state to the inviscid Hou-Li model* (A.1.4) $(\omega, u, \psi) = (\sin x, \sin x, \sin x)$*, similar to the steady state* $(\omega, u) = (-\sin x, \sin x)$ *of the DG model on* S^1*. Many of the results we present in this chapter have analogies for the gCLM model; see in particular [67, 68].*

A.1.2 Main results

We summarize the main results of the chapter below and devote the subsequent sections to proving these results. Our first result is on the finite-time blowup of the weak inviscid advection model; for its proof see Section A.2 and A.3.

Theorem A.1.2. *For the weak advection model* (A.1.5) *in the inviscid case* $\nu = 0$*, there exists a constant* $\delta > 0$ *such that for* $a \in (1-\delta, 1)$*, the weak advection model* (A.1.5) *develops a finite time singularity for some* C^∞ *initial data. Moreover, there exists a self-similar profile* $(\omega_\infty, \omega_\infty, \omega_\infty)$ *corresponding to a blowup that is neither expanding nor focusing. More precisely, the blowup solution to* (A.1.5) *has the form*

$$\omega(x,t) = \frac{1}{1+c_{u,\infty}t}\omega_\infty\,, u(x,t) = \frac{1}{1+c_{u,\infty}t}u_\infty\,, \psi(x,t) = \frac{1}{1+c_{u,\infty}t}\psi_\infty\,,$$

for some negative constant $c_{u,\infty}$ *with a blowup time given by* $T = \frac{-1}{c_{u,\infty}}$*.*

Remark A.1.3. *Such self-similar blowup that is neither expanding nor focusing is observed numerically for* $a \in [0.6, 0.9]$*. See also a similar phenomenon observed for the gCLM model in [300] for* $a \in [0.68, 0.95]$*. The blowup result for the gCLM model has been proved in [68] for a sufficiently close to* 1*. We remark that for a very close to* 1*, since we can show that* $c_{u,\infty} = 2(a-1) + o(a-1)$*, the blowup time becomes very large due to the very small coefficient* $1-a$ *in the vortex stretching term which slightly dominates the advection term. It would be extremely difficult to compute such singularity numerically since it takes an extremely long time for the*

singularity to develop. For a below a critical value a_0, i.e. $a < a_0$, we observe that the weak advection Hou-Li model develops a focusing singularity.

The second result is on the blowup of the original Hou-Li model with C^α initial data; for its proof see Section A.4. In [137], the authors made an important observation that advection can be weakened by C^α data. Intuitively if $u = O(x^\alpha)$ in the origin, since ψ is C^2, we have that $\psi u_x \approx \alpha \psi_x u$ near the origin, the vortex stretching term is stronger than the advection term if $\alpha < 1$. See [76, 67] on results of blowup of the DG model with Hölder continuous data.

Theorem A.1.4. *Consider the Hou-Li model (A.1.4) in the inviscid case $\nu = 0$. There exists a constant $\delta_0 > 0$ such that for $\alpha \in (1 - \delta_0, 1)$, (A.1.4) develops a finite time singularity for some C^α initial data. Moreover, there exists a C^α self-similar profile corresponding to a blowup that is neither expanding nor focusing, similar to the setting in Theorem A.1.2.*

Remark A.1.5. *This theorem establishes blowups of type C^α for any α close to 1, which of course implies blowups in less regular classes since $C^\alpha \subset C^{\alpha_1}$ for $\alpha_1 < \alpha$. The regularity of the profile determines the speed of the blowup since our constructed C^α profile has blowup time $T = O(-1/(2(\alpha - 1)))$. We remark that however, we do not have blowup for data intrinsically in a low regularity class C^ϵ for ϵ close to 0; that is, data that is C^ϵ but not in any higher C^α classes. We conjecture that such blowup might be focusing, which is beyond the scope of this chapter.*

Remark A.1.6. *The above two theorems imply that the result of the wellposedness in [215] of the Hou-Li model for C^1 initial data is sharp. As long as the advection is weakened or slightly less smooth data is allowed, we would have a self-similar blowup.*

The third result is on the finite-time blowup of the weak advection model with viscosity. The dynamic rescaling formulation implies that the viscous terms are asymptotically small. Thus, we can build on Theorem A.1.2 to establish Theorem A.1.7. We remark that there is no exact self-similar profile due to the viscous term. We will provide more details of the blowup analysis for the viscous case in Section A.5.

Theorem A.1.7. *Consider the weak advection model (A.1.5) with viscosity. There exists a constant $\delta_1 > 0$ such that for $a \in (1 - \delta_1, 1)$, the weak advection model (A.1.5) develops a finite time singularity for some C^∞ initial data.*

We use the framework of the dynamic rescaling formulation to establish the blowups. This formulation was first introduced by McLaughlin, Papanicolaou, and co-workers in their study of self-similar blowup of the nonlinear Schrödinger equation [315, 263]. This formulation was later developed into an effective modulation technique, which has been applied to analyze the singularity formation for the nonlinear Schrödinger equation [246, 321], compressible Euler equations [49], the nonlinear heat equation [324], the generalized KdV equation [308], and other dispersive problems. Recently this approach has been applied to prove singularity in various gCLM models [76, 67, 68] and in Euler equations [135, 72, 73]. Our blowup analysis consists of several steps. First, we use the dynamic rescaling formulation to link a self-similar singularity to the (stable) steady state of the dynamic rescaling formulation. Secondly, we identify, either analytically or numerically, an approximate steady state to the dynamic rescaling formulation. Thirdly, we perform energy estimates using a singularly weighted norm to establish linear and nonlinear stability of the approximate steady state. Finally, we establish exponential convergence to the steady state in the rescaled time.

The crucial ingredient of the framework is the linear stability of the approximate steady state, and we usually adopt a singularly weighted L^2-based estimate. To avoid an overestimate in the linear stability analysis, we expand the perturbation in terms of the orthonormal basis with respect to the weight L^2 norm and reduce the linear stability estimate into an estimate of a quadratic form for the Fourier coefficients. We further extract the damping effect of the linearized operator by establishing a lower bound on the eigenvalues of an infinite-dimensional symmetric matrix. We prove the positive-definiteness of this quadratic form by performing an exact computation of the eigenvalues of a small number of Fourier modes with rigorous computer-assisted bounds, and treat the high-frequency Fourier modes as a small perturbation by using the asymptotic decay of the quadratic form in the high-frequency Fourier coefficients.

A.1.3 Organization of the chapter and notations

In Section A.2, we introduce our dynamic rescaling formulation and link the blowup of the physical equation to the steady state of the dynamic rescaling formulation. The linear stability of the approximate steady state is established. In Section A.3, we establish the nonlinear stability of the approximate steady state and the exponential convergence to the steady state, which proves Theorem A.1.2 and the blowup for the weak advection model. In Section A.4, we prove Theorem A.1.4 and establish

blowup for the original model with Hölder continuous data. In Section A.5, we prove Theorem A.1.7 by designing a special energy norm to estimate the viscous terms. We provide the crucial linear damping estimates in the Appendix using computer assistance.

Throughout the article, we use $(\cdot,\cdot)$ to denote the inner product on S^1: $(f,g) = \int_{-\pi}^{\pi} fg$. We use C to denote absolute constants, which may vary from line to line, and we use $C(k)$ to denote some constant that may depend on specific parameters k we choose. We use $A \lesssim B$ for positive B to denote that there exists an absolute constant $C > 0$ such that $A \le CB$.

A.2 Dynamic Rescaling Formulation and Linear Estimates

A.2.1 Dynamic rescaling formulation

We will establish the singularity formation of the weak advection model by using the dynamic rescaling formulation. We first consider the inviscid case with $\nu = 0$. For solutions to the system (A.1.5), we introduce

$$\tilde{u}(x,\tau) = C_u(\tau)u(x,t(\tau))\,, \quad \tilde{\omega}(x,\tau) = C_u(\tau)\omega(x,t(\tau))\,, \quad \tilde{\psi}(x,\tau) = C_u(\tau)\psi(x,t(\tau))\,,$$

where

$$C_u(\tau) = \exp\left(\int_0^\tau c_u(s)ds\right)\,, \quad t(\tau) = \int_0^\tau C_u(s)ds\,.$$

We can show that the rescaled variables solve the following dynamic rescaling equation

$$\begin{aligned} \tilde{u}_\tau + 2a\tilde{\psi}\tilde{u}_x &= 2\tilde{u}\tilde{\psi}_x + c_u\tilde{u}\,, \\ \tilde{\omega}_\tau + 2a\tilde{\psi}\tilde{\omega}_x &= \left(\tilde{u}^2\right)_x + c_u\tilde{\omega}\,, \\ -\tilde{\psi}_{xx} &= \tilde{\omega}\,. \end{aligned} \tag{A.2.1}$$

Remark A.2.1. *We do not rescale the spatial variable x, since we are interested in a blowup solution that is neither focusing nor expanding within a fixed period. The scaling factors for u, ω, ψ are thus the same.*

When we establish a self-similar blowup, it suffices to show the dynamic stability of equation (A.2.1) close to an approximate steady state with scaling parameter $c_u < -\epsilon < 0$ uniformly in time for a small constant ϵ; see also [76]. In fact, it's easy to see that if $(\tilde{u},\tilde{\omega},\tilde{\psi},c_u)$ converges to a steady-state $(u_\infty,\omega_\infty,\psi_\infty,c_{u,\infty})$ of (A.2.1), then

$$\omega(x,t) = \frac{1}{1+c_{u,\infty}t}\omega_\infty\,, u(x,t) = \frac{1}{1+c_{u,\infty}t}u_\infty\,, \psi(x,t) = \frac{1}{1+c_{u,\infty}t}\psi_\infty\,,$$

is a self-similar solution of (A.1.5).

From now on, we will primarily work in the dynamic rescaling formulation and use the notations that $\tilde{u} = \bar{u} + \hat{u}$, where $\bar{u}$ is the approximate steady state that we perturb around and $\hat{u}$ is the perturbation. Notations for variables $\tilde{\omega}$ and $\tilde{\psi}$ are similar.

A.2.2 Equations governing the perturbation

We use the steady state corresponding to the case of $a = 1$ to construct an approximate steady state for (A.2.1).

$$\bar{\omega} = \sin x\,, \quad \bar{u} = \sin x\,, \quad \bar{\psi} = \sin x\,, \quad \bar{c}_u = 2(a-1)\bar{\psi}_x(0) = 2(a-1)\,.$$

We consider odd perturbations $\hat{u}$, $\hat{\omega}$, $\hat{\psi}$. The parities are preserved in time by equation (A.2.1). We use the normalization condition as $c_u = 2(a-1)\hat{\psi}_x(0)$. This normalization ensures that $\bar{u}_x(0) + \hat{u}_x(0)$ is conserved in time.

To simplify our presentation, we will drop the ˆ in the perturbation $\hat{u}$ and use u for $\hat{u}$, ω for $\hat{\omega}$, ψ for $\hat{\psi}$. Now the perturbations satisfy the following system

$$\begin{aligned} u_\tau &= -2a\sin x u_x - 2a\cos x\psi + 2u\cos x + 2\sin x\psi_x + \bar{c}_u u + c_u\bar{u} + N_1 + F_1\,,\\ \omega_\tau &= -2a\sin x\omega_x - 2a\cos x\psi + 2u\cos x + 2\sin x u_x + \bar{c}_u\omega + c_u\bar{\omega} + N_2 + F_2\,,\\ -\psi_{xx} &= \omega\,, \end{aligned} \tag{A.2.2}$$

where N_1, N_2 and F_1, F_2 are the nonlinear terms and error terms defined below:

$$N_1 = (c_u + 2\psi_x)u - 2a\psi u_x\,, \quad N_2 = c_u\omega + 2uu_x - 2a\psi\omega_x\,,$$

$$F_1 = (\bar{c}_u + 2\bar{\psi}_x)\bar{u} - 2a\bar{\psi}\bar{u}_x = 2(a-1)\sin x(1-\cos x)\,, \quad F_2 = \bar{c}_u\bar{\omega} + 2\bar{u}\bar{u}_x - 2a\bar{\psi}\bar{\omega}_x = F_1\,.$$

We further organize the system (A.2.2) into the main linearized term and a smaller term containing a factor of $a - 1$:

$$\begin{aligned} u_\tau &= L_1 + (a-1)L_1' + N_1 + F_1\,,\\ \omega_\tau &= L_2 + (a-1)L_2' + N_2 + F_2\,,\\ -\psi_{xx} &= \omega\,. \end{aligned} \tag{A.2.3}$$

where

$$\begin{aligned} L_1 &= -2\sin x u_x - 2\cos x\psi + 2u\cos x + 2\sin x\psi_x\,,\\ L_1' &= -2\sin x u_x - 2\cos x\psi + 2u + 2\psi_x(0)\sin x\,,\\ L_2 &= -2\sin x\omega_x - 2\cos x\psi + 2u\cos x + 2\sin x u_x\,, \end{aligned}$$

$$L_2' = -2\sin x\omega_x - 2\cos x\psi + 2\omega + 2\psi_x(0)\sin x\,.$$

To show that the dynamic rescaling equation is stable and converges to a steady state, we will perform a weighted-L^2 estimate with a singular weight ρ and a weighted L^2 norm

$$\rho = \frac{1}{2\pi(1-\cos x)}\,, \quad \|f\|_\rho = (f^2,\rho)^{1/2}\,.$$

For initial perturbation with $u_x(0,0) = 0$, we have $u_x(0,\tau) = 0$ for all time and

$$E^2(\tau) = \frac{1}{2}((u_x^2,\rho) + (\omega^2,\rho))\,,$$

is well-defined. We will first show that the dominant parts L_1 and L_2 provide damping. The following lemma is crucial and motivates the choice of ρ.

Lemma A.2.2. *We have the following identity*

$$(\sin x f_x, f\rho) = \frac{1}{2}(f^2,\rho)\,,$$

which can be verified directly by using integration by parts.

A.2.3 Stability of the main parts in the linearized equation

In order to extract the maximal amount of damping, we will expand the perturbed solution in the Fourier series and perform exact calculations. We first explore the orthonormal basis in $L^2(\rho)$.

Lemma A.2.3. *For the space of odd periodic functions on* $[0,2\pi]$*, we describe a complete set of orthonormal basis* $\{o^k\}$ *in* $L^2(\rho)$

$$o^k = \sin(kx) - \sin((k-1)x)\,, \quad k = 1,2,\cdots\,.$$

Similarly, for the space of even periodic functions that lie in $L^2(\rho)$*, we describe a complete set of orthonormal basis* $\{e^k\}$

$$e^k = \cos(kx) - \cos((k+1)x)\,, \quad k = 0,1,\cdots\,.$$

Now we are now ready to establish linear stability.

Proposition A.2.4. *The following energy estimate holds for the leading linearized operators*

$$dE_1 \coloneqq ((L_1)_x, u_x\rho) + (L_2,\omega\rho) \le -0.16[(u_x,u_x\rho) + (\omega,\omega\rho)]\,.$$

Proof. Consider the expansion of ω, u_x, and u in the orthonormal basis

$$\omega = \sum_{k\geq 1} a_k o^k\,, \quad u = \sum_{k\geq 1} b_k o^k\,, \quad u_x = \sum_{k\geq 1} c_k e^k\,.$$

Note the summation index for u_x satisfies $k \geq 1$ since we can easily see that

$$(u_x, e^0\rho) = \frac{1}{2\pi}\int_0^{2\pi} u_x = 0\,.$$

We first express b_k in terms of c_k. If we insert the expression of the basis into u, take derivative and compare the coefficients with the expansion of u_x, we get

$$c_i = \sum_{k=1}^{i} b_k - i b_{i+1}\,.$$

Therefore we can solve

$$b_{i+1} = b_1 - \sum_{k=1}^{i-1} \frac{c_k}{k(k+1)} - \frac{c_i}{i}\,.$$

Moreover, we have the compatibility condition $u_x(0) = \sum_{k\geq 0} b_k = 0$. Therefore we can solve b_1 and obtain

$$b_i = \sum_{k\geq i} \frac{c_k}{k(k+1)} - \frac{c_{i-1}}{i}\,, \tag{A.2.4}$$

where we define $c_0 = 0$.

Now we write out the terms explicitly using the expansions

$$\begin{aligned} dE_1 &= 2(-u\sin x - u_{xx}\sin x, u_x\rho) + 2(\sin x\psi + \sin x\psi_{xx}, u_x\rho) \\ &\quad + 2(-\sin x\omega_x - \cos x\psi, \omega\rho) + 2(u\cos x + \sin x u_x, \omega\rho) \\ &= -[(u_x, u_x\rho) + (\omega, \omega\rho) + (u, u\rho)] + 2[-(\cos x\psi, \omega\rho) + (\sin x\psi, u_x\rho) + (u\cos x, \omega\rho)]\,. \end{aligned}$$

Here we use the crucial Lemma A.2.2 to extract damping on the local terms and the Biot-Savart law $-\psi_{xx} = \omega$ to cancel the effect of the nonlocal terms $\sin x\psi_{xx}$ in $(L_1)_x$ and $\sin x u_x$ in L_2. Next, we calculate the remaining nonlocal terms explicitly.

$$-2(\cos x\psi, \omega\rho) = (2(1-\cos x)\psi, \omega\rho) - 2(\psi, \omega\rho) = \frac{1}{\pi}(\psi, \omega) - 2(\psi, \omega\rho)\,.$$

We express ω and ψ both in terms of orthonormal basis o^k corresponding to the weighted norm and the canonical basis $\sin(kx)$ corresponding to the (normalized by $\frac{1}{\pi}$) L^2 norm.

$$\omega = \sum_{k\geq 1} (a_k - a_{k+1})\sin(kx)\,,$$

where we denote $a_0 = 0$. Therefore

$$\psi = \sum_{k\geq 1} \frac{a_k - a_{k+1}}{k^2} \sin(kx) \,.$$

Furthermore, we collect

$$\psi = \sum_{k\geq 1} \sum_{j\geq k} \frac{a_j - a_{j+1}}{j^2} o^k \,.$$

Therefore we can compute explicitly that

$$-2(\cos x\psi, \omega\rho) = \sum_{k\geq 1} \frac{(a_k - a_{k+1})^2}{k^2} - 2\sum_{k\geq 1} a_k \sum_{j\geq k} \frac{a_j - a_{j+1}}{j^2} \,.$$

We use integration by parts similar to Lemma A.2.2 to obtain

$$2[(\sin x\psi, u_x\rho) + (u\cos x, \omega\rho)] = 2(\cos x\omega + \psi - \sin x\psi_x, u\rho) \coloneqq 2(T_u, u\rho) \,.$$

We further have

$$\begin{aligned}
T_u &= \sum_{k\geq 1} (a_k - a_{k+1})[\frac{k+1}{2k}\sin((k-1)x) + \frac{k-1}{2k}\sin((k+1)x) + \frac{1}{k^2}\sin(kx)] \\
&= \sum_{k\geq 1} \sin(kx)[\frac{k+2}{2(k+1)}(a_{k+1} - a_{k+2}) + \frac{a_k - a_{k+1}}{k^2} + \frac{k-2}{2(k-1)}(a_{k-1} - a_k)] \\
&= \sum_{k\geq 1} [\frac{k-2}{2(k-1)}a_{k-1} + (\frac{1}{2k(k-1)} + \frac{1}{k^2})a_k + (\frac{k+1}{2k} + \frac{1}{(k+1)^2} - \frac{1}{k^2})a_{k+1} \\
&\quad + \sum_{j>k+1} (\frac{1}{j^2} - \frac{1}{(j-1)^2})a_j]o^k \,,
\end{aligned}$$

where the terms involving $\frac{1}{k-1}$ in the summand is regarded as 0 for $k = 1$. Therefore we collect explicitly that

$$\begin{aligned}
dE_1 = \sum_{k\geq 1} \{&-(a_k^2 + b_k^2 + c_k^2) + \frac{(a_k - a_{k+1})^2}{k^2} - 2a_k \sum_{j\geq k} \frac{a_j - a_{j+1}}{j^2} + 2b_k[\frac{k-2}{2(k-1)}a_{k-1} \\
&+ (\frac{1}{2k(k-1)} + \frac{1}{k^2})a_k + (\frac{k+1}{2k} + \frac{1}{(k+1)^2} - \frac{1}{k^2})a_{k+1} + \sum_{j>k+1} (\frac{1}{j^2} - \frac{1}{(j-1)^2})a_j]\} \,.
\end{aligned}$$

Substituting (A.2.4) into the above and we can simplify

$$\begin{aligned}
dE_1 = -\sum_{k\geq 1} \{&a_k^2(1 + \frac{1}{k^2} - \frac{1}{(k-1)^2}) + c_k^2(1 + \frac{1}{k(k+1)}) + 2a_k a_{k+1}\frac{1}{(k+1)^2} \\
&+ 2a_k \sum_{j>k+1} a_j(\frac{1}{j^2} - \frac{1}{(j-1)^2}) + 2a_k c_k \frac{1 + 2k - k^2}{2k^2(k+1)} + 2a_{k+1}c_k \frac{k^2 - k - 1}{2k^2(k+1)^2} \\
&- 2a_{k+2}c_k \frac{k+2}{2(k+1)^2} + \sum_{j>k} 2a_k c_j \frac{1}{j(j+1)}\} \,.
\end{aligned}$$

After this explicit computation, we notice that the damping estimate in Proposition A.2.4 can be cast into an estimate of a quadratic form; see (A.2.5), which is equivalent to a lower bound on the eigenvalues of an infinite-dimensional symmetric matrix.

$$\begin{aligned} F(a,c) \coloneqq \sum_{k\geq 1}\{a_k^2(0.84+\frac{1}{k^2}-\frac{1}{(k-1)^2})+c_k^2(0.84+\frac{1}{k(k+1)})+2a_ka_{k+1}\frac{1}{(k+1)^2} \\ +2a_k\sum_{j>k+1}a_j(\frac{1}{j^2}-\frac{1}{(j-1)^2})+2a_kc_k\frac{1+2k-k^2}{2k^2(k+1)}+2a_{k+1}c_k\frac{k^2-k-1}{2k^2(k+1)^2} \\ -2a_{k+2}c_k\frac{k+2}{2(k+1)^2}+\sum_{j>k}2a_kc_j\frac{1}{j(j+1)}\}\geq 0\,. \end{aligned} \tag{A.2.5}$$

We notice that the entries decay fast. Therefore the strategy to prove (A.2.5) is to combine a computer-assisted estimate of the eigenvalues of its finite truncation with a decay estimate of the remaining part. We will defer the proof of (A.2.5) to the Appendix, see Lemma A.6.1 and the proof. Thereby we conclude the linear estimate. □

A.3 Nonlinear Estimates and Convergence to Self-similar Profile

A.3.1 Nonlinear stability

By Proposition A.2.4 and equation (A.2.3), we have

$$\begin{aligned} \frac{1}{2}\frac{d}{d\tau}E^2(\tau)\leq -0.16E^2(\tau)+(a-1)[((L_1')_x,u_x\rho)+(L_2',\omega\rho)] \\ +((N_1)_x,u_x\rho)+(N_2,\omega\rho)+((F_1)_x,u_x\rho)+(F_2,\omega\rho)\,. \end{aligned} \tag{A.3.1}$$

We first provide some estimates about the weighted L^2 norm and L^∞ norm of some lower-order terms.

Lemma A.3.1. *The following estimates hold*

1. *Weighted L^2 norm:*

$$\|\psi\|_\rho,\|\psi_x-\psi_x(0)\|_\rho,\|u\|_\rho\lesssim E\,.$$

2. *L^∞ norm:*

$$\|\psi_x\|_\infty,\|\frac{\psi}{\sin x}\|_\infty,\|u\|_\infty\lesssim E\,.$$

Proof. For (1), we use the setting of the Fourier series approach as in the proof of Proposition A.2.4 and pick up the notation there.

$$\begin{aligned}\|\psi_x-\psi_x(0)\|_\rho^2 &= \sum_{k\geq 1}(\sum_{j\geq k}\frac{a_j-a_{j+1}}{j})^2 \leq \sum_{k\geq 1}\sum_{j\geq k}a_j^2(\frac{1}{k^2}+\sum_{j>k}(\frac{1}{j}-\frac{1}{j+1})^2)\\ &\lesssim \sum_{j\geq 1}a_j^2\sum_{k\geq 1}\frac{1}{k^2}\lesssim \|\omega\|_\rho^2\,,\end{aligned}$$

where we have used the Cauchy-Schwarz inequality. We can similarly estimate $\|\psi\|_\rho$. Then we get

$$\|u\|_\rho^2=\sum_{j\geq 1}b_j^2=\sum_{j\geq 1}\frac{1}{j(j+1)}c_j^2\leq\sum_{j\geq 1}c_j^2=\|u_x\|_\rho^2\,.$$

For (2), we first compute using Fourier series similar to (1)

$$\psi_x(0)=\sum_{j\geq 1}\frac{a_j-a_{j+1}}{j}\lesssim\|\omega\|_\rho\,.$$

Next, we estimate

$$\|\psi_x-\psi_x(0)\|_\infty\lesssim\|\psi_{xx}\|_1\lesssim\|\omega\|_\rho\,.$$

Similarly, we obtain the estimate for $\|u\|_\infty$. For $\|\frac{\psi}{\sin x}\|_\infty$, since ψ is odd and periodic, we have $\psi(\pi)=\psi(0)=0$ and only need to estimate this norm in $[0,\pi]$. Since $\sin x\geq\frac{2}{\pi}\min\{x,\pi-x\}$ in $[0,\pi]$, we have $\|\frac{\psi}{\sin x}\|_\infty\lesssim\|\psi_x\|_\infty$ by Lagrange's mean value theorem. □

Combined with the damping in Lemma A.2.2, we further obtain

$$\begin{aligned}((L_1')_x,u_x\rho) &= 2(-\cos x u_x-\sin x u_{xx}+\sin x\psi-\cos x\psi_x+u_x+\psi_x(0)\cos x,u_x\rho)\\ &\lesssim E^2+(\|\psi\|_\rho+\|\psi_x-\psi_x(0)\|_\rho)E\lesssim E^2\,,\end{aligned}$$

$$(L_2',\omega\rho)=2(-\sin x\omega_x-\cos x\psi+\omega+\psi_x(0)\sin x,\omega\rho)\lesssim E^2+(\|\psi\|_\rho+|\psi_x(0)|)E\lesssim E^2\,.$$

$$((F_1)_x,u_x\rho)\lesssim|a-1|E\,,\quad (F_2,\omega\rho)\lesssim|a-1|E\,,$$

$$\begin{aligned}((N_1)_x,u_x\rho) &= 2(-\omega u+(1-a)(\psi_x-\psi_x(0))u_x-a\psi u_{xx},u_x\rho)\\ &\lesssim E^2(\|\psi_x\|_\infty+\|u\|_\infty)+|(u_x^2,(\psi\rho)_x)|\\ &\lesssim E^2(\|\psi_x\|_\infty+\|u\|_\infty+\|\frac{\psi}{\sin x}\|_\infty)\lesssim E^3\,,\end{aligned}$$

$$(N_2,\omega\rho)=2((a-1)\psi_x(0)\omega+uu_x-a\psi\omega_x,\omega\rho)\lesssim E^2(\|\psi_x\|_\infty+\|u\|_\infty+\|\frac{\psi}{\sin x}\|_\infty)\lesssim E^3\,.$$

Therefore we have

$$\frac{d}{d\tau}E(\tau) \leq -(0.16 - C|a-1|)E + C|a-1| + CE^2\,. \tag{A.3.2}$$

We can perform the standard bootstrap argument to show that there exist absolute constants $\delta, C > 0$ such that if $|a-1| < \delta$ and $E(0) < C|a-1|$, then we have $E(\tau) < C|a-1|$ for all time. In particular $c_u = O(|a-1|^2)$ and $c_u + \bar{c}_u < 0$. Therefore we prove that the solution blows up in finite time.

A.3.2 Estimates using a higher-order Sobolev norm

In order to establish convergence of the solution to a steady state, we need to estimate weighted norms of u_t and ω_t. As was pointed out in [68], we need to provide stability estimates of the equation in higher-order Sobolev norms to close the estimate. In particular, we choose

$$K^2(\tau) = \|D_x u_x\|_\rho^2 + \|D_x \omega\|_\rho^2\,,$$

where we denote D_x to be the operator $\sin x \partial x$.

Remark A.3.2. *This choice of weighted norms is again motivated by the local linear damping estimates. We recall that the leading order terms of the local terms in the linearized operators $(L_1)_x$, L_2 are $-2D_x u_x$ and $-2D_x w$, and we have $2(D_x f, f\rho) = (f, f\rho)$. Therefore in this new weighted norm, the combined terms would again give damping*

$$(-2D_x D_x u_x, D_x u_x \rho) + (-2D_x D_x w, D_x w \rho) = -K^2\,. \tag{A.3.3}$$

We now obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}K^2(\tau) \leq\ & (D_x(L_1)_x, D_x u_x \rho) + (D_x L_2, D_x \omega\rho) + (D_x(N_1)_x, D_x u_x \rho)\\ & + (D_x N_2, D_x \omega\rho) + (a-1)[(D_x(L_1')_x, D_x u_x\rho) + (D_x L_2', D_x\omega\rho)]\\ & + (D_x(F_1)_x, D_x u_x\rho) + (D_x F_2, D_x\omega\rho)\,.\end{aligned}$$

We will denote the terms that have $\|\cdot\|_\rho$ norm bounded by E as *l.o.t.*. The bound

$$\|D_x[fg]\|_\rho \lesssim (\|f_x\|_2 + \|f\|_2) \quad \text{for } g = 1, \cos x, \sin x$$

combined with the oddness of ψ and u would imply that $D_x[fg]$ is *l.o.t.* for $f = \psi, \psi_x, u$ and $g = \sin x, \cos x, 1$. Therefore combined with (A.3.3), we have the following estimate for the main term

$$dK_1 \coloneqq (D_x(L_1)_x, D_x u_x\rho) + (D_x L_2, D_x\omega\rho)\,,$$

$$\begin{aligned} dK_1 &\leq -K^2 - 2(D_x[\sin x\omega], D_x u_x \rho) + 2(D_x D_x u, D_x \omega \rho) + CEK \\ &= -K^2 - (\sin 2x\omega, D_x u_x \rho) + (\sin 2x u_x, D_x \omega \rho) + CEK \\ &\leq -K^2 + CEK\,, \end{aligned}$$

where we have again used a crucial cancellation in the equality, similar to that of dE_1 in Subsection A.2.3. We estimate the rest of the terms similar to the nonlinear stability estimates in (A.3.1).

$$(D_x(L_1')_x, D_x u_x \rho) + (D_x L_2', D_x \omega \rho) \lesssim K^2 + EK\,,$$

$$(D_x(F_1)_x, D_x u_x \rho) + (D_x F_2, D_x \omega \rho) \lesssim |a-1|K\,,$$

$$\begin{aligned} (D_x(N_1)_x, D_x u_x \rho) &\lesssim EK^2 + |(-2wD_x u + 2(1-a)D_x\psi_x u_x - 2a\psi D_x u_{xx}, D_x u_x \rho)| \\ &\lesssim EK^2 + \|\sin x u_x\|_\infty EK + |(u_{xx}^2, (\psi \sin^2 x\rho)_x)| \\ &\lesssim EK(K+E) + K^2 \|\frac{\psi}{\sin x}\|_\infty \lesssim EK(K+E)\,, \end{aligned}$$

$$(D_x N_2, D_x \omega \rho) \lesssim EK^2 + |(2D_x u u_x - 2a\psi D_x \omega_x, D_x \omega \rho)| \lesssim EK(K+E)\,,$$

where we have used integration by parts and the estimate

$$\|\sin x u_x\|_\infty \lesssim \|\sin x u_{xx}\|_1 + \|\cos x u_x\|_1 \lesssim \|D_x u_x\|_\rho + \|u_x\|_\rho \lesssim E + K\,.$$

We can finally prove that

$$\frac{d}{d\tau} K(\tau) \leq -(1 - C|a-1|)K + CE + C|a-1| + CE(E+K)\,.$$

Therefore combined with (A.3.2), we can find an absolute constant $\mu > 1$ such that

$$\frac{d}{d\tau}(K + \mu E) \leq -(0.1 - C|a-1|)(K+\mu E) + C|a-1| + C(K+\mu E)^2\,.$$

By using a standard bootstrap argument, there exist absolute constants $\delta_0 < \delta, C > 0$, if $|a-1| < \delta_0$ and $K(0) + \mu E(0) < C|a-1|$, then $K(\tau) + \mu E(\tau) < C|a-1|$ for all time.

A.3.3 Convergence to the steady state

We estimate the weighted norm of ω_τ and $u_{\tau,x}$ and then use the standard convergence in time argument as in [76, 68].

$$J^2(\tau) = \frac{1}{2}((u_{\tau,x}^2, \rho) + (\omega_\tau^2, \rho))\,.$$

Applying the estimates of $\frac{d}{d\tau}E$ to $\frac{d}{d\tau}J$, we can get damping for the linear parts, and the small error terms corresponding to $\bar{\omega}$ and $\bar{u}$ vanishes. Therefore we yield

$$\frac{1}{2}\frac{d}{d\tau}J^2 \leq -(0.16 - C|a-1|)J^2 + ((N_1)_{\tau,x}, u_{\tau,x}\rho) + ((N_2)_\tau, \omega_t\rho)\,.$$

Using estimates similar to Lemma A.3.1 and nonlinear estimates in (A.3.1), we get

$$((N_1)_{\tau,x}, u_{\tau,x}\rho) \lesssim EJ^2 + (\psi_\tau u_{xx}, u_{\tau,x}\rho) \lesssim EJ^2 + J^2\|u_{xx}\sin x\|_\rho \leq (E+K)J^2\,.$$

$$((N_2)_\tau, \omega_\tau\rho) \lesssim EJ^2 + (\psi_\tau\omega_x, \omega_\tau\rho) \lesssim EJ^2 + J^2\|\omega_x\sin x\|_\rho \leq (E+K)J^2\,.$$

Combined with the a priori estimates on $E+K$, we can establish exponential convergence of J to zero. Then we can use the same argument as in [76, 68] to establish exponential convergence to the steady state and conclude the proof of Theorem A.1.2.

A.4 Blowup of the Original Model with Hölder Continuous Data

In this section, we follow the strategy of the linear and nonlinear estimates of the weak advection model in Sections A.2 and A.3, and establish blowup of C^α data for the original model (A.1.5) with $a = 1$. Here $\alpha < 1$ is close to 1. Many of the ideas are drawn from the paper [67] and we only outline the most important steps. Intuitively, C^α regularity of the profile weakens the advection and therefore contributes to a blowup in finite time.

A.4.1 Dynamic rescaling formulation around the approximate steady state

Before we start, we will solve the Biot-Savart law of recovering ψ from ω with odd symmetry.

Lemma A.4.1. *Suppose that ω, ψ are odd and periodic on $[-\pi,\pi]$, with $-\psi_{xx} = \omega$. Then we solve $\psi_x(0) = -\frac{1}{2\pi}\int_0^{2\pi} y\omega(y)$ and obtain*

$$\psi = \int_0^x (y-x)\omega(y)dy + x\psi_x(0)\,. \tag{A.4.1}$$

The proof of this lemma is straightforward by integration in x.

We construct the following approximate steady state with C^α regularity for (A.2.1).

$$\bar{\omega}_\alpha = sgn(x)|\sin x|^\alpha\,, \quad \bar{u}_\alpha = sgn(x)|\sin x|^{\frac{1+\alpha}{2}}\,, \quad \bar{c}_{u,\alpha} = (\alpha-1)\bar{\psi}_{\alpha,x}(0)\,,$$

where $\bar{\psi}_\alpha$ is related to $\bar{\omega}_\alpha$ via (A.4.1). We consider odd perturbations u, ω, ψ. The odd symmetry of the solution is preserved in time by equation (A.2.1). We will use

the normalization condition as $c_u = (\alpha - 1)\psi_x(0)$, which ensures that u vanishes to a higher order at all times so that we can use the same singular weight ρ. In fact we compute using (A.2.1) and the normalization conditions that

$$\lim_{x\to 0} \frac{\bar{u}_\alpha(x) + u(x,\tau)}{x^{\frac{1+\alpha}{2}}} = \lim_{x\to 0} \frac{\bar{u}_\alpha(x) + u(x,0)}{x^{\frac{1+\alpha}{2}}} .$$

Therefore if we make the initial perturbation $u(x,0)$ vanish to order $1+\alpha$ around the origin, $u(x,\tau)$ will also vanish to order $1+\alpha$ for all time.

Now similar to what we obtain in (A.2.2), the perturbations satisfy the following system

$$\begin{aligned} u_\tau &= L_1 + R_{1,\alpha} + N_{1,\alpha} + F_{1,\alpha} , \\ \omega_\tau &= L_2 + R_{2,\alpha} + N_{2,\alpha} + F_{2,\alpha} , \\ -\psi_{xx} &= \omega , \end{aligned} \tag{A.4.2}$$

where we extract the same leading order linear parts as in (A.2.3), while the nonlinear and error terms change and the residual error terms $R_{1,\alpha}, R_{2,\alpha}$ model the discrepancy between our approximate profile with C^α regularity and the steady state profile $\bar{\omega} = \bar{u} = \bar{\psi} = \sin x$. Define $\psi_{\text{res}} = \bar{\psi}_\alpha - \bar{\psi}, \omega_{\text{res}} = \bar{\omega}_\alpha - \bar{\omega}, u_{\text{res}} = \bar{u}_\alpha - \bar{u}$. We can express $R_{i,\alpha}$ and $F_{i,\alpha}$ as follows

$$R_{1,\alpha} = -2\psi_{\text{res}} u_x - 2u_{\text{res},x}\psi + 2u\psi_{\text{res},x} + 2u_{\text{res}}\psi_x + \bar{c}_{u,\alpha} u + c_u \bar{u}_\alpha .$$

$$R_{2,\alpha} = -2\psi_{\text{res}}\omega_x - 2\omega_{\text{res},x}\psi + 2uu_{\text{res},x} + 2u_{\text{res}}u_x + \bar{c}_{u,\alpha}\omega + c_u\bar{\omega}_\alpha ,$$

$$N_{1,\alpha} = (c_u + 2\psi_x)u - 2\psi u_x , \quad N_{2,\alpha} = c_u\omega + 2uu_x - 2\psi\omega_x ,$$

$$F_{1,\alpha} = (\bar{c}_{u,\alpha} + 2\bar{\psi}_{\alpha,x})\bar{u}_\alpha - 2\bar{\psi}_\alpha\bar{u}_{\alpha,x} , \quad F_{2,\alpha} = \bar{c}_{u,\alpha}\bar{\omega}_\alpha + 2\bar{u}_\alpha\bar{u}_{\alpha,x} - 2\bar{\psi}_\alpha\bar{\omega}_{\alpha,x} .$$

Before we perform our energy estimates, we will obtain some basic estimates of the residues.

Lemma A.4.2. *The following estimates hold for* $\kappa = \frac{7}{8} < \frac{9}{10} < \alpha < 1$.

1. *Pointwise estimates of the residues:*

$$|\partial_x^i \omega_{res}| + |\partial_x^i u_{res}| \lesssim |\alpha - 1||\sin x|^{\kappa - i} , \quad i = 0, 1, 2, 3,$$

$$\|\psi_{res}\|_\infty + \|\psi_{res,x}\|_\infty \lesssim |\alpha - 1| .$$

2. *Refined estimates using cancellations:*

$$|\frac{\alpha-1}{2}\bar{u}_{\alpha,x} - \sin x u_{res,xx}| + |\sin x \partial x[\frac{\alpha-1}{2}\bar{u}_{\alpha,x} - \sin x u_{res,xx}]| \lesssim |\alpha-1|^{1/2}|x||\sin x|^{\alpha-1} .$$

Proof. The first part of (1) and (2) can be proved by using direct calculations, and we refer to Lemma 6.1 in [67] for details. There seems to be a typo in (6.11) in [67] where $\bar{\omega}_\alpha$ should have been $\bar{\omega}_{\alpha,x}$. Furthermore, by the expression in Lemma A.4.1 we get the second part of (1). □

Similar to the weak advection case, we define the energy $E^2(\tau) = \frac{1}{2}((u_x^2, \rho) + (\omega^2, \rho))$. We will estimate the growth of $E(\tau)$. The leading order linear estimates L_1, L_2 can be obtained in Proposition A.2.4. The estimates for the nonlinear terms $N_{1,\alpha}, N_{2,\alpha}$ follow almost exactly the same as the weak advection case by using Lemma A.3.1.

A.4.2 Nonlinear stability

By the computation in the previous subsection, we get

$$\frac{1}{2}\frac{d}{d\tau}E^2(\tau) \le -0.16E^2 + CE^3 + ((R_{1,\alpha})_x, u_x\rho) + (R_{2,\alpha}, \omega\rho) + ((F_{1,\alpha})_x, u_x\rho) + (F_{2,\alpha}, \omega\rho) .$$

Further, we get

$$\begin{aligned}((R_{1,\alpha})_x, u_x\rho) \le\ & (-2\psi_{\text{res}}u_{xx}, u_x\rho) + (c_u\bar{u}_{\alpha,x} - 2u_{\text{res},xx}\psi, u_x\rho) \\ & + (\|\omega_{\text{res}}\|_\infty + \|u_{\text{res}}\|_\infty + C|\alpha - 1|)E^2 .\end{aligned}$$

For the first term, we can use integration by parts and Lemma A.3.1 to obtain

$$CE^2(\|\psi_{\text{res},x}\|_\infty + \|\frac{\psi_{\text{res}}}{\sin x}\|_\infty) \lesssim \|\psi_{\text{res},x}\|_\infty E^2 .$$

For the second term, we compute

$$c_u\bar{u}_{\alpha,x} - 2u_{\text{res},xx}\psi = \psi_x(0)[(\alpha - 1)\bar{u}_{\alpha,x} - 2\sin x u_{\text{res},xx}] + 2u_{\text{res},xx}(\sin x\psi_x(0) - \psi) .$$

Thus by Lemmas A.3.1 and A.4.2, we have

$$\|c_u\bar{u}_{\alpha,x} - 2u_{\text{res},xx}\psi\|_\rho \lesssim E|\alpha - 1|^{1/2} + |\alpha - 1|\|\frac{\sin x\psi_x(0) - \psi}{|\sin x|x}\|_\infty . \tag{A.4.3}$$

Finally, for $|x| \ge \pi/2$, we have

$$|\frac{\sin x\psi_x(0) - \psi}{|\sin x|x}| \lesssim |\psi_x(0)| + \|\frac{\psi}{\sin x}\|_\infty \lesssim E .$$

For $|x| < \pi/2$, we use Lemma A.4.1 and $|\sin x| \ge 2/\pi|x|$ to obtain

$$|\frac{\sin x\psi_x(0) - \psi}{|\sin x|x}| \lesssim |\frac{\sin x\psi_x(0) - \psi}{x^2}| \le |\frac{(\sin x - x)\psi_x(0)}{x^2}| + |\frac{\int_0^x (y - x)\omega(y)}{x^2}| \lesssim E ,$$

where we have used the bound $\|\omega/x\|_1 \lesssim \|\omega/x\|_2 \lesssim \|\omega\|_\rho$ in the last inequality. Thus we yield

$$((R_{1,\alpha})_x, u_x\rho) \lesssim |\alpha-1|^{1/2}E^2\,.$$

Similarly we get

$$(R_{2,\alpha},\omega\rho) \le (-2\frac{\psi}{\sin x}\omega_{\mathrm{res},x}\sin x,\omega\rho)+(c_u\bar{\omega}_\alpha,\omega\rho)+(2uu_{\mathrm{res},x},\omega\rho)+C|\alpha-1|E^2\,.$$

For the first two terms, we can estimate them using Lemmas A.3.1 and A.4.2. For the third term, we use Hardy's inequality to derive

$$\|uu_{\mathrm{res},x}\|_\rho/|\alpha-1| \lesssim \|u/x/\sin x\|_2 \lesssim \|u/x^2\|_2+\|u/(\pi-x)\|_2 \lesssim \|u_x/x\|_2+\|u_x\|_2 \lesssim E\,.$$

Therefore we have

$$(R_{2,\alpha},\omega\rho) \lesssim |\alpha-1|E^2\,.$$

For the error terms, we can just perform standard norm estimates. We focus on the pointwise estimates for $x \ge 0$ and the case $x<0$ follows by using the odd symmetry of the solution.

$$\begin{aligned}F_{2,\alpha} &= (\alpha-1)\bar{\psi}_{\alpha,x}(0)\sin^\alpha x+(\alpha+1)\cos x\sin^\alpha x-2\alpha(\psi_{\mathrm{res}}+\sin x)\cos x\sin^{\alpha-1}x\\ &= (\alpha-1)(\bar{\psi}_{\alpha,x}(0)-\cos x)\sin^\alpha x-2\alpha\frac{\psi_{\mathrm{res}}}{\sin x}\cos x\sin^\alpha x\,.\end{aligned}$$

Therefore, we obtain

$$\|F_{2,\alpha}\|_\rho \lesssim |\alpha-1|+\|\frac{\psi_{res}}{\sin x}\|_\infty \lesssim |\alpha-1|\,.$$

Similarly, we have

$$(F_{1,\alpha})_x = \frac{\alpha^2-1}{2}\sin^{\frac{\alpha-1}{2}}x\cos x[\bar{\psi}_{\alpha,x}(0)-\bar{\psi}_\alpha\frac{\cos x}{\sin x}]+\sin^{\frac{\alpha+1}{2}}x[(\alpha+1)\bar{\psi}_\alpha-2\sin^\alpha x]\,.$$

Further, we obtain the following estimate:

$$\|(\alpha+1)\bar{\psi}_\alpha-2\sin^\alpha x\|_\infty \le |\alpha-1|\|\bar{\psi}_\alpha\|_\infty+2\|\psi_{\mathrm{res}}\|_\infty+2\|\sin x-\sin^\alpha x\|_\infty \lesssim |\alpha-1|\,,$$

$$\begin{aligned}\bar{\psi}_{\alpha,x}(0)-\bar{\psi}_\alpha\frac{\cos x}{\sin x} &= (1-\cos x)+\psi_{\mathrm{res},x}(0)-\psi_{\mathrm{res}}\frac{\cos x}{\sin x}\\ &= (1-\cos x)(1+\frac{\psi_{\mathrm{res}}}{\sin x})+\psi_{\mathrm{res},x}(0)-\frac{\psi_{\mathrm{res}}}{\sin x}\,.\end{aligned}$$

Combined with the estimate similar to that in (A.4.3), we get

$$\|[\bar{\psi}_{\alpha,x}(0)-\bar{\psi}_\alpha\frac{\cos x}{\sin x}]\rho^{1/2}\|_\infty \lesssim \|\frac{1-\cos x}{x}(1+\frac{\psi_{\mathrm{res}}}{\sin x})\|_\infty+\|[\psi_{\mathrm{res},x}(0)-\frac{\psi_{\mathrm{res}}}{\sin x}]/x\|_\infty \lesssim 1\,.$$

Therefore we yield

$$\|(F_{1,\alpha})_x\|_\rho \lesssim |\alpha - 1| \,.$$

Collecting all the estimates of the residues and the error terms, we arrive at

$$\frac{d}{d\tau}E(\tau) \leq -(0.16 - C|\alpha - 1|^{1/2}|)E + CE^2 + C|\alpha - 1| \,.$$

Similar to Subsection A.3, we can perform the bootstrap argument to conclude finite-time blowup.

A.4.3 Estimates in higher-order Sobolev norms and convergence to steady state

Following the ideas in Subsection A.3.2, we can perform estimates in higher-order Sobolev norms and then close the estimates to establish convergence to a steady state. We use the same energy K and only sketch the main steps here. We first have from Subsection A.3.2 the estimates of L_i and N_i and obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}K^2(\tau) \leq -K^2 + CEK + CEK^2 + ((R_{1,\alpha})_{xx}, \sin^2 x u_{xx}\rho) \\ + ((R_{2,\alpha})_x, \sin^2 x\omega_x\rho) + ((F_{1,\alpha})_{xx}, \sin^2 x u_{xx}\rho) + ((F_{2,\alpha})_x, \sin^2 x\omega_x\rho) \,.\end{aligned}$$

By Lemma A.4.2, $\sin x\partial x u_{\text{res}}, \sin x\partial x\omega_{\text{res}}$ shares the same pointwise estimates as $u_{\text{res}}, \omega_{\text{res}}$. We can obtain, similar to the estimates in E, the estimates

$$((R_{1,\alpha})_{xx}, \sin^2 x u_{xx}\rho) \lesssim |\alpha - 1|(E + K)K + K\|[u_{\text{res},xx}(\psi - \sin x\psi_x(0))]_x \sin x\|_\rho \,.$$

We further obtain the following estimate

$$\begin{aligned}\|[u_{\text{res},xx}(\psi - \sin x\psi_x(0))]_x \sin x\|_\rho &\lesssim |\alpha - 1|E + \|u_{\text{res},xx}(\psi_x - \cos x\psi_x(0)) \sin x\|_\rho \\ &\lesssim |\alpha - 1|E + |\alpha - 1|\|(\psi_x - \psi_x(0))/x\|_\infty \,.\end{aligned}$$

Finally by the Cauchy-Schwarz inequality, we have

$$|(\psi_x - \psi_x(0))/x| \leq \int_0^x |\omega(y)/y|dy \lesssim \|\omega/x\|_2 \lesssim \|\omega\|_\rho \lesssim E \,, \tag{A.4.4}$$

and therefore we conclude

$$((R_{1,\alpha})_{xx}, \sin^2 x u_{xx}\rho) \lesssim |\alpha - 1|(E + K)K \,.$$

By similar computations, we have

$$((R_{2,\alpha})_x, \sin^2 x\omega_x\rho) \lesssim |\alpha - 1|(E + K)K \,.$$

And for the residue terms, we use similar estimates to obtain

$$\|(F_{2,\alpha})_x \sin x\|_\rho \lesssim |\alpha - 1|\,,$$

$$\|(F_{1,\alpha})_{xx} \sin x\|_\rho \lesssim |\alpha - 1| + |\alpha - 1|\|(\frac{\psi_{\text{res}}}{\sin x})_x \rho^{1/2} \sin x\|_\infty\,.$$

We can use the triangular inequality to estimate

$$|(\frac{\psi_{\text{res}}}{\sin x})_x \rho^{1/2} \sin x| \leq |[\psi_{\text{res},x}(0) - \frac{\psi_{\text{res}}}{\sin x}]/x| + |(\psi_{\text{res},x} - \psi_{\text{res},x}(0))/x| + |\frac{1-\cos x}{x}\frac{\psi_{\text{res}}}{\sin x}|\,.$$

Combined with the estimate similar to that in (A.4.3) and (A.4.4), we conclude

$$\|(F_{1,\alpha})_{xx} \sin x\|_\rho \lesssim |\alpha - 1|\,.$$

We can finally obtain the same estimate as in Subsection A.3.2

$$\frac{d}{d\tau}K(\tau) \leq -(1 - C|a-1|)K + CE + C|a-1| + CEK\,.$$

We can again find an absolute constant $\mu_\alpha > 1$ such that

$$\frac{d}{d\tau}(K + \mu_\alpha E) \leq -(0.1 - C|a-1|)(K + \mu_\alpha E) + C|a-1| + C(K + \mu_\alpha E)^2\,.$$

And we can use the bootstrap argument on this higher-order energy $K + \mu_\alpha E$ to conclude a priori estimate in this norm. Now we can perform weighted estimates in time using the energy J as in Subsection A.3.3, where the linear estimates follow from the linear estimates of E, the nonlinear estimates follow from Subsection A.3.3, and the error term vanishes under time-differentiation. We can obtain the same estimates of J and establish exponential convergence of J to zero. Then we use the argument in [76, 68] to establish exponential convergence to the steady state and conclude the proof of Theorem A.1.4.

A.5 Blowup of the Viscous Model with Weak Advection

In this section, we follow the strategy of the linear and nonlinear estimates of the weak advection model and establish blowup for the weak advection viscous model with $a < 1$. Intuitively, the viscosity term is small in the dynamic rescaling formulation and nonlinear stability can be closed using a higher-order norm, where the viscosity term has a stability effect. Therefore we expect that the viscous weak advection model develops a finite time singularity as well.

A.5.1 Dynamic rescaling formulation

We recall the weak advection model with viscosity.

$$\begin{aligned} u_t + 2a\psi u_z &= 2u\psi_z + \nu u_{zz}\,, \\ \omega_t + 2a\psi\omega_z &= \left(u^2\right)_z + \nu\omega_{zz}\,, \\ -\psi_{zz} &= \omega\,. \end{aligned} \tag{A.5.1}$$

We use the rescaled variables

$$\tilde{u}(x,\tau) = C_u(\tau)u(x,t(\tau))\,, \quad \tilde{\omega}(x,\tau) = C_u(\tau)\omega(x,t(\tau))\,, \quad \tilde{\psi}(x,\tau) = C_u(\tau)\psi(x,t(\tau))\,,$$

where

$$C_u(\tau) = C_u(0)\exp\left(\int_0^\tau c_u(s)ds\right)\,, \quad t(\tau) = \int_0^\tau C_u(s)ds\,.$$

For solutions to (A.5.1), the rescaled variables satisfy the dynamic rescaling equation

$$\begin{aligned} \tilde{u}_\tau + 2a\tilde{\psi}\tilde{u}_x &= 2\tilde{u}\tilde{\psi}_x + c_u\tilde{u} + \nu C_u(\tau)u_{xx}\,, \\ \tilde{\omega}_\tau + 2a\tilde{\psi}\tilde{\omega}_x &= \left(\tilde{u}^2\right)_x + c_u\tilde{\omega} + \nu C_u(\tau)\omega_{xx}\,, \\ -\tilde{\psi}_{xx} &= \tilde{\omega}\,. \end{aligned} \tag{A.5.2}$$

Remark A.5.1. *Different from the rescaling in the inviscid case, we introduce an extra degree of freedom: the constant $C_u(0)$. We will choose it later to ensure that the viscous term has a relatively small scaling compared to the main terms.*

In order to establish a finite time blowup, it suffices to prove the dynamic stability of (A.5.2) with scaling parameter $c_u < -\epsilon < 0$ for all time; see also [76]. As before, we will primarily work in the dynamic rescaling formulation and use the notations $\tilde{u} = \bar{u} + u$, where $\bar{u}$ is an approximation steady state and u is the perturbation that we will control in time.

We consider the following approximate steady state.

$$\bar{\omega} = \bar{u} = \bar{\psi} = \sin x\,, \quad \bar{c}_u(\tau) = 2(a-1)\bar{\psi}_x(0) - \nu C_u(\tau)\bar{u}_{xxx}(0)/\bar{u}_x(0) = 2(a-1) + \nu C_u(\tau)\,.$$

We consider odd perturbations u, ω, ψ. The odd symmetry of the solution is preserved in time by equation (A.5.2). We use the following normalization condition: $c_u = 2(a-1)\psi_x(0) - \nu C_u(\tau)u_{xxx}(0)$. This normalization ensures that $u_x(0)$ remains 0 if the initial perturbation satisfies $u_x(0,0) = 0$. In fact, if $u_x(\tau,0) = 0$, then we obtain

$$\begin{aligned} \frac{d}{d\tau}u_x(\tau,0) = \frac{d}{d\tau}(u_x(\tau,0) + \bar{u}_x(\tau,0)) &= (2-2a)(u_x(\tau,0) + \bar{u}_x(\tau,0))(\psi_x(\tau,0) + \bar{\psi}_x(\tau,0)) \\ &+ (c_u + \bar{c}_u)(u_x(\tau,0) + \bar{u}_x(\tau,0)) + \nu C_u(\tau)(\bar{u}_{xxx}(0) + u_{xxx}(0)) = 0\,. \end{aligned}$$

This particular choice of the approximate steady state and the normalization conditions ensure that $u_x(0,\tau) = 0$ for all time provided that the initial perturbation satisfies $u_x(0,0) = 0$. We will perform the same weighted norm estimate in the singular weight ρ and the weighted norm E as in the inviscid case.

A.5.2 Estimates of the viscous terms

Now the perturbation satisfies

$$\begin{aligned} u_\tau &= L_1 + (a-1)L_1' + N_1 + F_1 + \nu C_u(\tau) V^u \,, \\ \omega_\tau &= L_2 + (a-1)L_2' + N_2 + F_2 + \nu C_u(\tau) V^\omega \,, \\ -\psi_{xx} &= \omega \,. \end{aligned} \tag{A.5.3}$$

Here the terms V^u and V^ω correspond to all of the terms containing the effect of the viscosity and we factor out explicitly the small factor $\nu C_u(\tau)$ for a fixed ν.

$$V^u = u_{xx} + \bar{u}_{xx} + (1 - u_{xxx}(0))(u + \bar{u}) = u_{xx} - u_{xxx}(0)\sin x + (1 - u_{xxx}(0))u \,,$$

$$V^\omega = \omega_{xx} - \omega_{xxx}(0)\sin x + (1 - \omega_{xxx}(0))\omega \,.$$

We invoke the nonlinear estimates in the inviscid case and obtain for the viscous model:

$$\frac{1}{2}\frac{d}{d\tau}E^2(\tau) \le -(0.16 - C|a-1|)E^2 + C|a-1|E + CE^3 + \nu C_u(\tau)[((V^u)_x, u_x\rho) + (V^\omega, \omega\rho)] \,. \tag{A.5.4}$$

We estimate the viscous terms carefully since they involve singular weights.

$$((V^u)_x, u_x\rho) = (u_{xxx} - u_{xxx}(0), u_x\rho) + u_{xxx}(0)(1 - \cos x, u_x\rho) + (1 - u_{xxx}(0))\|u_x\|_\rho^2 \,.$$

Notice that

$$|\rho_x| = \rho\left|\frac{-\sin x}{1-\cos x}\right| \lesssim \rho|\frac{1}{x}| \,, \quad |\rho_{xx}| = \rho\left|\frac{-\cos x}{1-\cos x} + \frac{2\sin^2 x}{(1-\cos x)^2}\right| \lesssim \rho|x^{-2}|$$

are singular near the origin and are smooth elsewhere. We can use integration by parts twice to compute

$$\begin{aligned} (u_{xxx} - u_{xxx}(0), u_x\rho) &= -(u_{xx} - xu_{xxx}(0), u_{xx}\rho) - (u_{xx} - xu_{xxx}(0), \frac{x^2}{2}u_{xxx}(0)\rho_x) \\ &\quad - (u_{xx} - xu_{xxx}(0), (u_x - \frac{x^2}{2}u_{xxx}(0))\rho_x) \\ &\le -\|u_{xx}\|_\rho^2 + C[|u_{xxx}(0)|\|u_{xx}\|_\rho + |u_{xxx}(0)|^2 + \|\frac{u_x}{x} - \frac{x}{2}u_{xxx}(0)\|_\rho^2] \\ &\le -\frac{1}{2}\|u_{xx}\|_\rho^2 + C|u_{xxx}(0)|^2 + C\|\frac{u_x}{x}\|_\rho^2 \,, \end{aligned}$$

where for the last inequality we use the weighted AM-GM inequality $ab \le \epsilon a^2 + \frac{1}{4\epsilon} b^2$ for a very small constant ϵ. Therefore we get

$$((V^u)_x, u_x\rho) \le -\frac{1}{2}\|u_{xx}\|_\rho^2 + C[|u_{xxx}(0)|^2 + \|\frac{u_x}{x}\|_\rho^2 + (1+|u_{xxx}(0)|)E^2] \,. \quad \text{(A.5.5)}$$

Similarly, we estimate via integration by parts

$$\begin{aligned}(\omega_{xx}, \omega\rho) &= -(\omega_x - \omega_x(0), (\omega_x - \omega_x(0))\rho) - (\omega_x - \omega_x(0), \omega_x(0)\rho) \\ &\quad - (\omega_x - \omega_x(0), x\omega_x(0)\rho_x) - (\omega_x - \omega_x(0), (\omega - x\omega_x(0))\rho_x) \\ &\le -\|\omega_x - \omega_x(0)\|_\rho^2 + C[|\omega_x(0)|\|\frac{\omega_x - \omega_x(0)}{x}\|_\rho + \|\frac{\omega}{x} - \omega_x(0)\|_\rho^2] \,.\end{aligned}$$

And we obtain

$$\begin{aligned}(V^\omega, \omega\rho) \le &-\|\omega_x - \omega_x(0)\|_\rho^2 + C[|\omega_x(0)|\|\frac{\omega_x - \omega_x(0)}{x}\|_\rho \\ &+ \|\frac{\omega}{x} - \omega_x(0)\|_\rho^2 + |\omega_{xxx}(0)|^2 + (1+|\omega_{xxx}(0)|)E^2] \,.\end{aligned} \quad \text{(A.5.6)}$$

The essential difficulty for the viscous terms is that after integration by parts, the singular weight produces various positive terms, on top of the damping terms $-\|u_{xx}\|_\rho$; see (A.5.5), (A.5.6). Fortunately, the positive terms contribute only to higher-order terms near the origin.

Consider the interval $I = [-\pi/2, \pi/2]$. ρ and $|1/x|$ are upper bounded by a positive constant outside of the interval and we have

$$\|\frac{u_x}{x}\|_\rho^2 \lesssim \|u_x\|_\rho^2 + \|\frac{u_x}{x^2}\|_{L^2(I)}^2 \lesssim E^2 + \|u_{xxx}\|_{L^\infty(I)}^2 \,,$$

$$\|\frac{\omega}{x} - \omega_x(0)\|_\rho^2 \lesssim \|\omega - x\omega_x(0)\|_\rho^2 + \|\frac{\omega}{x^2} - \omega_x(0)/x\|_{L^2(I)}^2 \lesssim E^2 + |\omega_x(0)|^2 + \|\omega_{xx}\|_{L^\infty(I)}^2 \,,$$

$$\|\frac{\omega_x - \omega_x(0)}{x}\|_\rho \lesssim \|\omega_x - \omega_x(0)\|_\rho + \|\frac{\omega_x - \omega_x(0)}{x^2}\|_{L^2(I)} \lesssim \|\omega_x - \omega_x(0)\|_\rho + \|\omega_{xxx}\|_{L^\infty(I)} \,.$$

Plugging these estimates into the (A.5.5), (A.5.6) and using again the weighted AM-GM inequality, we can yield

$$(V^\omega, \omega\rho) + ((V^u)_x, u\rho) \le -\frac{1}{2}(\|\omega_x - \omega_x(0)\|_\rho^2 + \|u_{xx}\|_\rho^2) + C[E_V^2 + (1+E_V)E^2] \,, \quad \text{(A.5.7)}$$

where

$$E_V = \|\omega_{xxx}\|_{L^\infty(I)} + \|u_{xxx}\|_{L^\infty(I)} + |\omega_x(0)| \,.$$

A.5.3 Estimates in a higher-order norm

We will use a weighted higher-order norm to close the estimates. To have a good estimate in this higher-order norm, it needs to satisfy three criteria. First, we need to extract damping in the leading order linear term. Secondly, we need to bound the terms like $\omega_{xxx}(0)$ using interpolation between the lower and the higher-order norms via the Gagliardo-Nirenberg inequality; therefore it needs to be at least as strong as a regular higher-order norm near the origin. Thirdly, we need damping for the diffusion terms to close the estimates. This motivates us to choose a combination of the $k-$th order weighted norms for $k \geq 1$:

$$E_k^2(\tau) = (u^{(k+1)}, u^{(k+1)}\rho_k) + (\omega^{(k)}, \omega^{(k)}\rho_k)\,, \quad \rho_k = (1+\cos x)^k\,,$$

where we use the notation that $f^{(k)} = \partial_x^k f$. We denote $E_0 = E$ and $\rho_0 = \rho$.

Remark A.5.2. *This weighted norm immediately satisfies criterion 2 and we will verify in the linear estimates that it satisfies criterion 1. Finally, a clever combination of the weighted norms can produce damping for the viscous terms and we make the damping terms in the estimates of the $(k-1)$-th order norms greater than the positive terms in the estimates of the k-th order norms. We will elaborate on those points and establish the nonlinear estimates.*

Now we can estimate $\frac{d}{d\tau}E_k(\tau)$ for $k > 0$ as follows

$$\begin{aligned}\frac{1}{2}\frac{d}{d\tau}E_k^2(\tau) &= (L_2^{(k)} + (a-1)(L_2')^{(k)} + N_2^{(k)} + F_2^{(k)} + \nu C_u(t)(V^\omega)^{(k)}, \omega^{(k)}\rho_k)\\ &+ (L_1^{(k+1)} + (a-1)(L_1')^{(k+1)} + N_1^{(k+1)} + F_1^{(k+1)} + \nu C_u(t)(V^u)^{(k+1)}, u^{(k+1)}\rho_k)\,,\end{aligned}$$

where the parts L_i, L_i', F_i, N_i are defined exactly the same as in the inviscid case.

We first look at the viscous terms. We have for example

$$((V^u)^{(k+1)}, u^{(k+1)}(1+\cos x)^k) \leq (u^{(k+3)}, u^{(k+1)}(1+\cos x)^k) + CE_V E_k + (1+E_V)E_k^2\,.$$

We use integration by parts twice to obtain

$$\begin{aligned}&(u^{(k+3)}, u^{(k+1)}(1+\cos x)^k) = -(u^{(k+2)}, u^{(k+2)}(1+\cos x)^k - u^{(k+1)}k(1+\cos x)^{k-1}\sin x)\\ &= -(u^{(k+2)}, u^{(k+2)}(1+\cos x)^k) + \frac{1}{2}(u^{(k+1)}, u^{(k+1)}k(1+\cos x)^{k-1}[(k-1) - k\cos x])\\ &\leq -(u^{(k+2)}, u^{(k+2)}(1+\cos x)^k) + C(k)(u^{(k+1)}, u^{(k+1)}(1+\cos x)^{k-1})\,.\end{aligned}$$

We can also get a similar bound for V^ω. Therefore combined with the leading order estimate (A.5.7) and using the idea in Remark A.5.2, we conclude that for small

enough constants $0 < \mu < \mu_0(k_0) < 1$, we have the following viscous estimate

$$\sum_{k=0}^{k_0} \mu^k [((V^u)^{(k+1)}, u^{(k+1)}\rho_k) + ((V^\omega)^{(k)}, \omega^{(k)}\rho_k)] \le C(k)[E_V^2 + (1+E_V)\sum_{k=0}^{k_0} \mu^k E_k^2] . \tag{A.5.8}$$

Here $\mu_0(k_0)$ is a generic constant depending on k_0. We can choose k_0 large enough later so that E_V can be bounded using the interpolation inequalities.

Now we look at the linear terms and extract damping. We denote the terms as lower order terms (l.o.t. for short) if their ρ_k-weighted L^2-norms are bounded by $\sum_{i=0}^{k-1} E_i$. For the terms of intermediate order, since $\rho_k \le C(k)\rho_i$ for $i < k$, combined with the classical elliptic estimate, we can show that u^j, ψ^i for $0 \le j < k+1$ and $0 \le i < k+2$ are l.o.t. Using the l.o.t. notation, we keep track only of the higher-order terms

$$(L_1^{(k+1)}, u^{(k+1)}\rho_k) = (-2\sin x u^{(k+2)} - 2k\cos x u^{(k+1)} + 2\sin x \psi^{(k+2)} + \text{l.o.t.}, u^{(k+1)}\rho_k) ,$$

$$(L_2^{(k)}, \omega^{(k)}\rho_k) = (-2\sin x \omega^{(k+1)} - 2k\cos x \omega^{(k)} + 2\sin x u^{(k+1)} + \text{l.o.t.}, \omega^{(k)}\rho_k) .$$

Again we have a crucial cancellation of the cross terms and for the leading order terms we use integration by parts to obtain for example

$$\begin{aligned}(-2\sin x u^{(k+2)} - 2k\cos x u^{(k+1)}, u^{(k+1)}\rho_k) &= (u^{(k+1)}, u^{(k+1)}(-k-(k-1)\cos x)\rho_k) \\ &\le -(u^{(k+1)}, u^{(k+1)}\rho_k) .\end{aligned}$$

Therefore we derive the following estimate

$$(L_1^{(k+1)}, u^{(k+1)}\rho_k) + (L_2^{(k)}, \omega^{(k)}\rho_k) \le -E_k^2 + C(k)\sum_{i=0}^{k-1} E_i E_k \le -\frac{1}{2}E_k^2 + C(k)\sum_{i=0}^{k-1} E_i^2 .$$

Similarly we have

$$(L_1'^{(k+1)}, u^{(k+1)}\rho_k) + (L_2'^{(k)}, \omega^{(k)}\rho_k) \le 2E_k^2 + C(k)\sum_{i=0}^{k-1} E_i^2 .$$

We have the trivial bound for the error term

$$(F_1^{(k+1)}, u^{(k+1)}\rho_k) + (F_2^{(k)}, \omega^{(k)}\rho_k) \le C(k)(a-1)E_k .$$

The nonlinear terms are more subtle. We will show that

$$(N_1^{(k+1)}, u^{(k+1)}\rho_k) \le C(k)\sum_{i=0}^{k} E_i^2 E_k , \tag{A.5.9}$$

and we can have the same bound for ω. In fact, for a canonical term in $N_1^{(k+1)}$ and $N_2^{(k)}$, it is of the form $\psi^{(i)}u^{(k+2-i)}$ or $\psi^{(i)}\psi^{(k+3-i)}$ or $u^{(i)}u^{(k+1-i)}$. For the terms $\psi u^{(k+2)}$ and $\psi\omega^{(k+1)}$, we can use integration by parts and Lemma A.3.1 to show that

$$(\psi u^{(k+2)}, u^{(k+1)}(1+\cos x)^k) \le C(k)(\|\psi_x\|_\infty + \|\frac{\psi}{\sin x}\|_\infty)E_k^2 \le C(k)EE_k^2 .$$

The terms associated with $\psi_x u^{(k+1)}$, $\psi_x\omega^{(k)}$, $uu^{(k+1)}$, and $u\omega^{(k)}$ have the same bound trivially. We can then focus on controlling the weighted norms of $\omega^{(i)}u^{(k-i)}$, $\omega^{(i)}\omega^{(k-1-i)}$, $u^{(i+1)}u^{(k-i)}$ for indices $0 < i < k$ to establish the bound (A.5.9). For example, we get

$$\|\omega^{(i)}u^{(k-i)}(1+\cos x)^{k/2}\|_2 \le \|\omega^{(i)}(1+\cos x)^{(i+1)/2}\|_\infty E_{k-1-i} .$$

Finally, by the fundamental theorem of calculus, we can bound the L^∞-norm by

$$\begin{aligned} &C(K)[\|\omega^{(i+1)}(1+\cos x)^{(i+1)/2}\|_1 + \|\omega^{(i)}(1+\cos x)^{(i-1)/2}\sin x\|_1] \\ &\le C(K)[E_{i+1} + \|\omega^{(i)}(1+\cos x)^{(i)/2}\|_1] \le C(k)\sum_{i=0}^{k} E_i . \end{aligned}$$

Therefore we conclude that (A.5.9) holds.

A.5.4 Collection of norms and finite time blowup

We collect the bounds (A.5.8) (A.5.9) for viscous and nonlinear terms, along with the linear bounds and the leading order estimate (A.5.4). For any fixed k_0, there exists a small enough constant $0 < \mu_1(k_0) < \mu_0(k_0)$, such that the following estimate holds

$$\frac{d}{d\tau}I_{k_0}^2 \le -(0.1 - C|a-1|)I_{k_0}^2 + C|a-1|I_{k_0} + CI_{k_0}^3 + C\nu C_u(\tau)[E_V^2 + (1+E_V)I_{k_0}^2] ,$$

where the energy is defined as

$$I_{k_0}^2 = \sum_{k=0}^{k_0} \mu_1^k(k_0)E_k^2 .$$

Here the constants depend on k_0 and μ but once we first prescribe k_0 then $\mu = \mu_1(k_0)$, they become just constants. We will later make our $C_u(\tau)$ and $|a-1|$ small to close the argument.

Finally, by the Gagliardo-Nirenberg inequality, for $k = 1, 3$, we have

$$\|\omega^{(k)}\|_{L^\infty(I)} \lesssim \|\omega^{(4)}\|_{L^2(I)}^\theta \|\omega\|_{L^2(I)}^{1-\theta} , \quad \theta = \frac{k+1/2}{4} .$$

This is the classical Gagliardo-Nirenberg inequality applied to a bounded domain, and we can just use the extension technique to prove it; see for example [270]. We get similar bounds involving u and conclude that $E_V \lesssim I_{k_0}$, for any fixed $k_0 \geq 4$. For example, we just take $k_0 = 4$ and obtain

$$\frac{d}{d\tau} I_4 \leq -(0.1 - C|a-1|)I_4 + C|a-1| + CI_4^2 + C\nu C_u(\tau)(1+I_4)I_4 \,.$$

Now we choose $C_u(0) = |a-1|^2$ for $|a-1| < \delta$ with a small enough $\delta > 0$. It is easy to check that the bootstrap argument for $I_4 \leq C|a-1|$ and $C_u(\tau) \leq C_u(0)\exp((a-1)t) \leq C_u(0)$ will hold for all time provided that it holds initially. We again use the estimate for the normalization constants

$$c_u + \bar{c}_u = 2(a-1) + \nu C_u(t)(1 - u_{xxx}(0)) + 2(a-1)\psi_x(0) < (a-1) < 0 \,.$$

Thus we can obtain a blowup in finite time in the physical variables.

A.6 Appendix

Lemma A.6.1. *Assume $\sum_{k\geq 1} a_k^2 + c_k^2 < \infty$, then we have the following inequality for* (A.2.5)

$$\begin{aligned} F(a,c) \coloneqq \sum_{k\geq 1} \{ & a_k^2 (0.84 + \frac{1}{k^2} - \frac{1}{(k-1)^2}) + c_k^2 (0.84 + \frac{1}{k(k+1)}) + 2a_k a_{k+1} \frac{1}{(k+1)^2} \\ & + 2a_k \sum_{j>k+1} a_j (\frac{1}{j^2} - \frac{1}{(j-1)^2}) + 2a_k c_k \frac{1+2k-k^2}{2k^2(k+1)} + 2a_{k+1} c_k \frac{k^2-k-1}{2k^2(k+1)^2} \\ & - 2a_{k+2} c_k \frac{k+2}{2(k+1)^2} + \sum_{j>k} 2a_k c_j \frac{1}{j(j+1)} \} \geq 0 \,. \end{aligned}$$

Proof. Denote the summation of terms in $F(a,c)$ that only involve a_i, c_j for $i, j \leq N$ as $F_N(a,c)$. Here $N = 200$. This quadratic form $F_N(a,c)$ can be expressed as $a^{(N),T} F^{(N)} c^{(N)}$, where $a^{(N)}, c^{(N)}$ are two vectors with entries a_i,c_j respectively and $F^{(N)}$ is a symmetric matrix. We numerically verify using interval arithmetic in Matlab that the smallest eigenvalue of $F^{(N)}$ is greater than 0.01; see remarks after the proof for details. Therefore we have

$$F_N(a,c) \geq 0.01 \sum_{k=1}^{N} (a_k^2 + c_k^2) \,.$$

For the remainder $F(a,c) - F_N(a,c)$, we estimate it term by term via the trivial bound $2ab \geq -(a^2+b^2)$ and obtain

$$\begin{aligned}
F(a,c) - F_N(a,c) &\geq \sum_{k>N} [a_k^2(0.84 + \frac{1}{k^2} - \frac{1}{(k-1)^2}) + c_k^2(0.84 + \frac{1}{k(k+1)})] \\
&+ \sum_{k=1}^{N} a_k^2(-\frac{2}{N^2} - \frac{1}{N+1}) + \sum_{k=N-1}^{N} c_k^2(-\frac{N^2-N-1}{2N^2(N+1)^2} - \frac{N+1}{2N^2}) \\
&+ \sum_{k>N} a_k^2(-\frac{3}{N^2} - N(\frac{1}{(N)^2} - \frac{1}{(N+1)^2}) - \frac{N^2-N-1}{2N^2(N+1)^2} - \frac{N+1}{2N^2} - \frac{N^2-2N-1}{2N^2(N+1)} \\
&- \frac{1}{N+2}) + \sum_{k>N} c_k^2(-\frac{N^2-N-1}{2N^2(N+1)^2} - \frac{N+1}{2N^2} - \frac{N^2-2N-1}{2N^2(N+1)} - \frac{1}{N+2}) .
\end{aligned}$$

For $N = 200$, we estimate all of the coefficients by a lower bound and obtain

$$F(a,c) - F_N(a,c) \geq -\frac{2}{N}\sum_{k=1}^{N}(a_k^2+c_k^2) + (0.84 - \frac{3}{N})\sum_{k>N}(a_k^2+c_k^2) \geq -\frac{2}{N}\sum_{k=1}^{N}(a_k^2+c_k^2) .$$

Therefore we conclude $F(a,c) \geq 0$. □

Remark A.6.2. *We now explain how to verify that the smallest eigenvalue of the symmetric matrix $F^{(200)}$ is greater than* 0.01. *We proceed in three steps.*

1. *We first use Matlab to perform an (approximate) SVD decomposition of*

$$F^{(200)} - 0.011I \approx VDV' .$$

   Here D is the diagonal matrix consisting of (approximate) eigenvalues of $F^{(200)} - 0.011I$, and V is the unitary matrix consisting of (approximate) eigenvectors of $F^{(200)} - 0.011I$.

2. *We use interval arithmetic to verify that the maximal absolute value of entries of $F^{(200)} - 0.011I - VDV'$ is at most 10^{-10}. Therefore the spectral norm of $F^{(200)} - 0.011I - VDV'$, which is bounded by its 1-norm, is rigorously bounded from above by 200×10^{-10}.*

3. *Since D has positive entries, we know that VDV' is positive definite. We conclude that*

$$F^{(200)} - 0.01I = VDV' + 0.001I + F^{(200)} - 0.011I - VDV'$$

   is positive definite.

Appendix B

SECOND ORDER ENSEMBLE LANGEVIN METHOD

We propose a sampling method based on an ensemble approximation of second order Langevin dynamics. The log target density is appended with a quadratic term in an auxiliary momentum variable and damped-driven Hamiltonian dynamics, is introduced; the resulting stochastic differential equation is invariant to the Gibbs measure, with marginal on the position coordinates given by the target. A preconditioner based on covariance under the law of position coordinates under the dynamics does not change this invariance property, and is introduced to accelerate convergence to the Gibbs measure. The resulting mean-field dynamics may be approximated by an ensemble method; this results in a gradient-free and affine-invariant stochastic dynamical system with desirable provably uniform convergence properties across the class of all Gaussian targets. Numerical results demonstrate the potential of the method as basis for a numerical sampler in Bayesian inverse problems, beyond the Gaussian setting.

B.1 Introduction

B.1.1 Set-up

Consider sampling the density

$$\pi(q) = \frac{1}{Z_q} \exp(-\Phi(q)) \,,$$

where $\Phi : \mathbb{R}^N \to \mathbb{R}$ is termed the *potential* function and Z_q the *normalization constant.* A broad family of problems can be cast into this formulation; the Bayesian approach to inverse problems provides a particular focus for our work [239]. The point of departure for the algorithms considered in this chapter is the following mean-field Langevin equation:

$$\frac{\mathrm{d}q}{\mathrm{d}t} = -C(\rho)D\Phi(q) + \sqrt{2C(\rho)}\frac{\mathrm{d}W}{\mathrm{d}t} \,, \tag{B.1.1}$$

where D denotes the gradient operator, W is an N-dimensional standard Brownian motion, ρ is the density associated to the law of q, and $C(\rho)$ is the covariance under this density. This constitutes a mean-field generalization [158] of the standard Langevin equation [364]. Applying a particle approximation to the mean-field

model results in an interacting particle system, and coupled Langevin dynamics [265, 158, 350, 157]. The benefit of preconditioning using the covariance is that it leads to mixing rates independent of the problem, provably for quadratic Φ and empirically beyond this setting [158], because of the affine invariance [175] of the resulting algorithms [157].

In order to further accelerate mixing and achieve sampling efficiency, we introduce an auxiliary variable $p \in \mathbb{R}^N$ and consider the Hamiltonian

$$\mathcal{H}(z) = \frac{1}{2}\langle p, \mathcal{M}^{-1} p\rangle + \Phi(q)\,, \tag{B.1.2}$$

where the new state variable $z := (q^\top, p^\top)^\top \in \mathbb{R}^{2N}$. Define a measure via its density on $\mathbb{R}^{2N}$ by

$$\Pi(z) = \frac{1}{Z_{q,p}} \exp(-\mathcal{H}(z))\,, \tag{B.1.3}$$

where $Z_{q,p}$ is the normalization constant. The marginal distribution of Π in the q variable gives the desired distribution π, i.e. $\int \Pi(z)\mathrm{d}p = \pi(q)$. We now aim at sampling the joint distribution. To this end, consider the following underdamped Langevin dynamics in $\mathbb{R}^{2N}$:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathcal{J} D\mathcal{H}(z) - \mathcal{K} D\mathcal{H}(z) + \sqrt{2\mathcal{K}}\frac{\mathrm{d}W_0}{\mathrm{d}t}\,, \tag{B.1.4}$$

with the choices

$$\mathcal{J} = \begin{pmatrix} 0 & C \\ -C & 0 \end{pmatrix}, \quad \mathcal{K} = \begin{pmatrix} \mathcal{K}_1 & 0 \\ 0 & \mathcal{K}_2 \end{pmatrix}. \tag{B.1.5}$$

Here W_0 is a standard Brownian motion in $\mathbb{R}^{2N}$ with components $W', W \in \mathbb{R}^N$. Then we have the following, proved in Subsection B.8.1:

Proposition B.1.1. *Assume that $\mathcal{K}_1$, $\mathcal{K}_2$ are symmetric and non-negative definite, and that C is symmetric positive definite. Assume further that C, $\mathcal{K}$ and $\mathcal{M}$ depend on the law of z under the dynamics defined by (B.1.4) and (B.1.5), but are are independent of z: all derivatives with respect to z are zero. Then the Gibbs measure $\Pi(z)$ is invariant under the dynamics defined by (B.1.4), (B.1.5).*

In practice, to simulate from such a mean-field model, it will be necessary to consider a particle approximation of the form

$$\frac{\mathrm{d}z^{(i)}}{\mathrm{d}t} = J(Z) D_z H(z^{(i)}; Z) - K(Z) D_z H(z^{(i)}; Z) + \sqrt{2K(Z)}\frac{\mathrm{d}W_0^{(i)}}{\mathrm{d}t}\,, \tag{B.1.6}$$

for the set of I particles $Z = \{z^{(i)}\}_{i=1}^{I}$, and where $M(Z), K(Z), J(Z)$ are appropriate empirical approximations of $\mathcal{M}(\rho), \mathcal{K}(\rho), \mathcal{J}(\rho)$ based on replacing ρ by ρ^I where

$$\rho^I = \frac{1}{I}\sum_{i=1}^{I} \delta_{z^{(i)}} ,$$

and the Hamiltonian is given by

$$H(z; Z) = \frac{1}{2}\langle p, M(Z)^{-1} p\rangle + \Phi(q) . \tag{B.1.7}$$

Thus

$$D_z H(z; Z) = \left(D\Phi(q)^\top, (M(Z)^{-1}p)^\top\right)^\top .$$

Note that H is the appropriate finite particle approximation of $\mathcal{H}$, given the particle approximation M of $\mathcal{M}$. The dependence of $\mathcal{H}$ on the law of z has been replaced by dependence on the collection of particles Z.[1]

Remark B.1.2. *Unlike* (B.1.4)*, the equation* (B.1.6) *is no longer a damped-driven Hamiltonian system; this is because of the dependence of the Hamiltonian on the particle positions Z, through the mass matrix. Furthermore, its marginal on any coordinate $q^{(i)}$ does not necessarily preserve the desired target measure under the dynamics. However we expect it to do so* approximately *when I is large. This justifies the use of algorithms based on* (B.1.6).

In this chapter we will concentrate on a specific choice of mean-field operators within the above general construction, which we now describe. Let $C_q(\rho)$ denote the q-marginal in the covariance under the law of (B.1.4). We make the choices $\mathcal{K}_1 = 0$, $C = \mathcal{M} = C_q(\rho)$, and $\mathcal{K}_2 = \gamma C_q(\rho)$, for a scalar damping parameter $\gamma > 0$. Then the underdamped Langevin dynamics yields

$$\begin{aligned} \frac{\mathrm{d}q}{\mathrm{d}t} &= p , \\ \frac{\mathrm{d}p}{\mathrm{d}t} &= -C_q(\rho)D\Phi(q) - \gamma p + \sqrt{2\gamma C_q(\rho)}\frac{\mathrm{d}W}{\mathrm{d}t} . \end{aligned} \tag{B.1.8}$$

To implement a particle approximation of the mean-field dynamics (B.1.8) we introduce particles in the form $z^{(i)}(t) = \left((q^{(i)}(t))^\top, (p^{(i)}(t))^\top\right)$ and we use the

[1] The reader is asked to note that collection of particles Z is different from normalization constant $Z_{q,p}$ appearing in (B.1.3).

ensemble covariance and mean approximations

$$C_q(\rho) \approx C_q(Z) \coloneqq \frac{1}{I}\sum_{i=1}^{I}\left(q^{(i)} - \bar{q}\right) \otimes \left(q^{(i)} - \bar{q}\right), \tag{B.1.9a}$$

$$\bar{q} \coloneqq \frac{1}{I}\sum_{i=1}^{I} q^{(i)}. \tag{B.1.9b}$$

In order to obtain affine invariance, we take the generalized square root of the ensemble covariance $C_q(Z)$, similarly to [157]. We introduce the $N \times I$ matrix

$$Q \coloneqq \left(q^{(1)} - \bar{q}, q^{(2)} - \bar{q}, \cdots, q^{(I)} - \bar{q}\right),$$

which allows us to define the empirical covariance and generalized (nonsymmetric) square root via

$$C_q(Z) = \frac{1}{I}QQ^T, \quad \sqrt{C_q(Z)} \coloneqq \frac{1}{\sqrt{I}}Q.$$

Now with I independent standard Brownian motions $\{W^{(i)}\}_{i=1}^{I} \in \mathbb{R}^I$, a natural particle approximation of (B.1.8) is, for $i = 1, \ldots, I$,

$$\begin{aligned} \frac{\mathrm{d}q^{(i)}}{\mathrm{d}t} &= p^{(i)}, \\ \frac{\mathrm{d}p^{(i)}}{\mathrm{d}t} &= -C_q(Z)D\Phi(q^{(i)}) - \gamma p^{(i)} + \sqrt{2\gamma C_q(Z)}\frac{\mathrm{d}W^{(i)}}{\mathrm{d}t}. \end{aligned} \tag{B.1.10}$$

In subsequent sections we will employ an ensemble approximation of $C_q(Z)D\Phi(\cdot)$, as in [158], thereby avoiding the need to compute adjoints of the forward model; we note also that in the linear case this approximation is exact. We will show that the resulting interacting particle system has the potential to provide accurate derivative-free inference for certain classes of inverse problems.

In the remainder of this section we provide a literature review, we highlight our contributions, and we outline the structure of the chapter.

B.1.2 Literature review

The overdamped Langevin equation is the canonical SDE that is invariant with respect to a given target density. Many sampling algorithms are built upon this idea, and in particular, it is shown to govern a large class of Monte Carlo Markov Chain (MCMC) methods; see [393, 355, 301]. To enhance mixing and accelerate convergence, a second-order method, Hybrid Monte Carlo (HMC, also referred to as Hamiltonian Monte Carlo) [126, 340] has been proposed, leading to underdamped

Langevin dynamics. There have been many attempts to justify the empirically observed fast convergence speed of second-order methods in comparison with first-order methods [28]. Recently a quantitative convergence rate is established in [56], showing that the underdamped Langevin dynamics converges faster than the overdamped Langevin dynamics when the log of the target π has a small Poincaré constant; see also [133].

The idea of introducing preconditioners within the context of interacting particle systems used for sampling is developed in [265]. Preconditioning via ensemble covariance is shown to boost convergence and numerical performance [158]. Other choices of preconditioners in sampling will lead to different forms of SDEs and associated Fokker-Planck equations with different structures, which can result in different sampling methods effective in different scenarios; see for example [284]. Affine invariance is introduced in [175] where it is argued that this property leads to desirable convergence properties for interacting particle systems used for sampling: methods that satisfy affine invariance are invariant under an affine change of coordinates, and are thus uniformly effective for problems that can be rendered well-conditioned under an affine transformation; in particular for the sampling of the class of all Gaussian measures. An affine invariant version of the mean-field underdamped Langevin dynamics of [158] is proposed in [350, 157].

Kalman methods have shown wide success in state estimation problems since their introduction by Evensen; see [144] for an overview of the field, and the papers [388, 228] for discussion of their use in inverse problems. Using the ensemble covariance as a preconditioner leads to affine invariant [157] and gradient-free [158] approximations of Langevin dynamics; this is desirable in practical computations in view of the intractability of derivatives in many large-scale models arising in science and engineering [88, 185, 258, 223]. See [25, 400] for analysis of these methods and, in the context of continuous data-assimilation, see [116, 26, 427]. There are other derivative-free methods that can be derived from the mean-field perspective, and in particular consensus-based methods show promise for optimization [60] and have recently been extended to sampling in [59].

Recent work has established the convergence of ensemble preconditioning methods to mean field limits; see for example [121]. For other works on rigorous derivation of mean field limits of interacting particle systems, see [425, 61, 231]. For underpinning theory of Hamiltonian-based sampling, see [39, 38, 294, 29].

B.1.3 Our contributions

The following contributions are made in this chapter:

1. We introduce an underdamped second order mean field Langevin dynamics, with a covariance-based preconditioner.
2. In the case of Bayesian inverse problems defined by a linear forward map, we show that that this mean field model preserves Gaussian distributions under time evolution and, if initialized at a Gaussian, converges to the desired target at a rate independent of the linear map.
3. We introduce finite particle approximations of the mean field model, resulting in an affine invariant method.
4. For Bayesian inverse problems, we introduce a gradient-free approximation of the algorithm, based on ensemble Kalman methodology.
5. In the context of Bayesian inverse problems we provide numerical examples to demonstrate that the algorithm resulting from the previous considerations has desirable sampling properties.

In Section B.2, we introduce the inverse problems context that motivates us. In Section B.3 we discuss the equilibrium distribution of the mean field model. Section B.4 introduces the ensemble Kalman approximation of the finite particle system; and in that section we also demonstrate affine invariance of the resulting method. Section B.5 presents analysis of the finite particle system in the case of linear inverse problems, where the ensemble Kalman approximation is exact; we demonstrate that the relaxation time to equilibrium is independent of the specific linear inverse problem considered, a consequence of affine invariance. In Section B.6 we provide numerical results which demonstrate the efficiency and potential value of our method, and in Section B.7 we draw conclusions. Proofs of the propositions are given in the appendix, Section B.8.

B.2 Inverse Problem

Consider the Bayesian inverse problem of finding q from an observation y determined by the forward model

$$y = \mathcal{G}(q) + \eta\,.$$

Here $\mathcal{G} : \mathbb{R}^N \to \mathbb{R}^J$ is a (in general) nonlinear forward map. We assume a prior zero-mean Gaussian $\pi_0 = \mathsf{N}(0, \Gamma_0)$ on unknown q and assume that the random

variable $\eta \sim \mathsf{N}(0, \Gamma)$, representing measurement error, is independent of the prior on q. We also assume that Γ, Γ_0 are positive definite. Then by Bayes rule, the posterior density that we aim to sample is given by[2]

$$\pi(q) \propto \exp\Big(-\frac{1}{2}\|y - \mathcal{G}(q)\|_\Gamma^2\Big)\pi_0(q) \propto \exp(-\Phi(q))\,,$$

where potential function $\Phi(q)$ has the following form:

$$\Phi(q) = \frac{1}{2}\|y - \mathcal{G}(q)\|_\Gamma^2 + \frac{1}{2}\|q\|_{\Gamma_0}^2\,. \tag{B.2.1}$$

In the linear case when $\mathcal{G}(q) = Aq$, $\Phi(q)$ is quadratic and the gradient $D\Phi(q)$ can be written as a linear function:

$$\Phi(q) = \frac{1}{2}\|y - Aq\|_\Gamma^2 + \frac{1}{2}\|q\|_{\Gamma_0}^2\,, \tag{B.2.2a}$$

$$D\Phi(q) = B^{-1}q - c\,, \tag{B.2.2b}$$

$$B = (A^T\Gamma^{-1}A + \Gamma_0^{-1})^{-1}\,, \quad c = A^T\Gamma^{-1}y\,. \tag{B.2.2c}$$

In this linear setting, the posterior distribution $\pi(q)$ is the Gaussian $\mathsf{N}(Bc, B)$.

B.3 Equilibrium Distributions for the Mean Field Fokker-Planck Equation

The mean-field underdamped Langevin equation (B.1.4) has an associated nonlinear and nonlocal Fokker-Planck equation giving the evolution of the law of particles $z(t)$, denoted $\rho(z, t)$. The equation for this law is (see proof in subsection B.8.1.)

$$\partial_t\rho = \nabla \cdot \big((\mathcal{K} - J)(\rho\nabla\mathcal{H} + \nabla\rho)\big)\,. \tag{B.3.1}$$

By Proposition B.1.1 this Fokker-Planck equation has $\Pi(z)$ as its equilibrium; this follows as for standard linear Fokker-Planck equations [364] since the dependence of the sample paths on ρ involves only the mean and covariance; see the proof in subsection B.8.1 and see also [129, 179, 237, 392, 364] for discussions on how to derive Fokker-Planck equations with a prescribed stationary distribution. In the specific case (B.1.8), recall that $\mathcal{J}$ and $\mathcal{K}$ are given by

$$\mathcal{J} = \begin{pmatrix} 0 & C_q(\rho) \\ -C_q(\rho) & 0 \end{pmatrix}\,, \quad \mathcal{K} = \begin{pmatrix} 0 & 0 \\ 0 & \gamma C_q(\rho) \end{pmatrix}\,. \tag{B.3.2}$$

These choices satisfy the assumption of Proposition B.1.1. We approximate (B.1.8) by the interacting particle system (B.1.10). In this context, we note Remark B.1.2 to motivate computational methods based on integrating (B.1.10).

[2] In what follows $\|\cdot\|_C = \|C^{-\frac{1}{2}}\cdot\|$, with analogous notation for the inducing inner-product, for any positive definite covariance C and for $\|\cdot\|$ the Euclidean norm.

B.4 Ensemble Kalman Approximation

B.4.1 Derivatives via differences

We now make the ensemble Kalman approximation to approximate the gradient term by differences, as in [158]:

$$D\mathcal{G}\left(q^{(i)}\right)\left(q^{(k)}-\bar{q}\right) \approx \left(\mathcal{G}\left(q^{(k)}\right)-\bar{\mathcal{G}}\right),$$

where $\bar{\mathcal{G}} := \frac{1}{I}\sum_{k=1}^{I}\mathcal{G}\left(q^{(k)}\right)$. Invoking this approximation within (B.1.10), using the specific form (B.2.1) of Φ, yields the following system of interacting particles in $\mathbb{R}^N$, for $i = 1, \ldots, I$:

$$\begin{aligned}
\dot{q}^{(i)} &= p^{(i)}, \\
\dot{p}^{(i)} &= -C_q(Z)\Gamma_0^{-1}q^{(i)} - \frac{1}{I}\sum_{k=1}^{I}\langle \mathcal{G}(q^{(k)})-\bar{\mathcal{G}}, \mathcal{G}(q^{(i)})-y\rangle_{\Gamma} q^{(k)} - \gamma p^{(i)} + \sqrt{2\gamma C_q(Z)}\dot{W}^{(i)}.
\end{aligned} \tag{B.4.1}$$

We will use this system as the basis of all our numerical experiments.

B.4.2 Affine invariance

In this subsection, we show the affine invariance property [175, 157, 265] for the Fokker-Planck equations in the mean-field regime (B.3.1), for the particle equation in the mean-field regime (B.1.8), for the ensemble approximation (B.1.10), and the gradient-free approximation (B.4.1). For simplicity of presentation, we only state the results in the case of ensemble approximation, and the mean-field case is a straightforward analogy upon dropping all of the particle superscripts.

Definition B.4.1 (Affine invariance for particle formulation)**.** *We say a particle formulation is affine invariant, if under all affine transformations of the form*

$$q^{(i)} = Av^{(i)} + b, \quad p^{(i)} = Au^{(i)}, \tag{B.4.2}$$

the equations on the transformed particle systems are given by the same equations with $q^{(i)}$, $p^{(i)}$ replaced by $v^{(i)}$, $u^{(i)}$ respectively, and with potential Φ replaced by $\tilde{\Phi}$ via

$$\tilde{\Phi}(v^{(i)}) = \Phi(q^{(i)}) = \Phi(Av^{(i)} + b).$$

Here A is any invertible matrix and b is a vector.

Definition B.4.2 (Affine invariance for Fokker-Planck equation)**.** *We say a Fokker-Planck equation is affine invariant, if under all affine transformations of the form*

$$q^{(i)} = Av^{(i)} + b\,, \quad p^{(i)} = Au^{(i)}\,,$$

the equations on the pushforward PDF $\tilde{\rho}^I$ are given by the same equation on ρ^I with $q^{(i)}$, $p^{(i)}$ replaced by $v^{(i)}$, $u^{(i)}$ respectively, and with Hamiltonian H replaced by $\tilde{H}$ via

$$\tilde{H}(v^{(i)}, u^{(i)}) = H(q^{(i)}, p^{(i)}) = H(Av^{(i)} + b, Au^{(i)})\,.$$

Here A is any invertible matrix and b is a vector.

The key dynamical systems introduced in this chapter are affine invariant:

Proposition B.4.3. *The particle formulations* (B.1.8)*,* (B.1.10) *and* (B.4.1) *are affine invariant. The Fokker-Planck equation* (B.3.1) *is also affine invariant.*

We defer the proof to Subsection B.8.2. The significance of affine invariance is that it implies that the rate of convergence is preserved under affine transformations. The proposed methodology is thus uniformly effective for problems that become well-conditioned under an affine transformation. Proposition B.5.1, which follows in the next section, illustrates this property in the setting of linear forward map $\mathcal{G}(\cdot)$.

Remark B.4.4. *The affine invariance of the methodology introduced in [265] involves a definition different from that in Definition B.4.1. In particular* (B.4.2) *is replaced by*

$$q^{(i)} = Av^{(i)} + b\,, \quad p^{(i)} = u^{(i)}\,, \tag{B.4.3}$$

B.5 Mean Field Model for Linear Inverse Problems

We consider the mean field SDE (B.1.8) in the linear inverse problem setting of Section B.2 with $\mathcal{G}(q) = Aq$; thus (B.2.2) holds. We note that B in (B.2.2) is both well-defined and symmetric positive definite since Γ_0, Γ are assumed to be symmetric positive definite. The two Propositions B.5.1, B.5.3 demonstrate problem-independent rates of convergence, across the set of all linear Gaussian inverse problems; this is a consequence of affine invariance which in turn is a consequence of our choice of preconditioned mean field system.

In the setting of the linear inverse problem, the mean field model (B.1.8) reduces to

$$\begin{aligned} \dot{q} &= p\,, \\ \dot{p} &= -C_q(\rho)(B^{-1}q - c) - \gamma p + \sqrt{2\gamma C_q(\rho)}\dot{W}\,. \end{aligned} \tag{B.5.1}$$

We prove the following result about this system in Subsection B.8.3:

Proposition B.5.1. *Write the mean $m(\rho)$ and the covariance $C(\rho)$ of the law $\rho(z)$ of particles in equation (B.5.1) in the block form*

$$m(\rho) = \begin{pmatrix} m_q(\rho) \\ m_p(\rho) \end{pmatrix}, \quad C(\rho) = \begin{pmatrix} C_q(\rho) & C_{q,p}(\rho) \\ C_{q,p}^T(\rho) & C_p(\rho) \end{pmatrix}.$$

The evolution of the mean and covariance is prescribed by the following system of ODEs:

$$\begin{aligned} \dot{m_q} &= m_p\,, \\ \dot{m_p} &= -C_q(B^{-1}m_q - c) - \gamma m_p\,, \\ \dot{C_q} &= C_{q,p} + C_{q,p}^T\,, \\ \dot{C_p} &= -C_q B^{-1} C_{q,p} - (C_q B^{-1} C_{q,p})^T - 2\gamma C_p + 2\gamma C_q\,, \\ \dot{C}_{q,p} &= -\gamma C_{q,p} - C_q B^{-1} C_q + C_p\,. \end{aligned} \tag{B.5.2}$$

The unique steady solution with positive definite covariance is the Gibbs measure $m_q = Bc, m_p = 0, C_{q,p} = 0, C_q = C_p = B$; the marginal on q gives the solution of the linear Gaussian Bayesian inverse problem. All other steady state solutions have degenerate covariance, are unstable, and take the form $m_q = B(c+m), m_p = 0, C_{q,p} = 0, C_q = C_p = B^{1/2}XB^{1/2}$ for a projection matrix X and m in the nullspace of C_q. The equilibrium point with positive definite covariance is hyperbolic and linearly stable and hence its basin of attraction is an open set, containing the equilibrium point itself, in the set of all mean vectors and positive definite covariances. Furthermore, the mean and covariance converge to this equilibrium, from all points in its basin of attraction, with a speed independent of B and c.

Remark B.5.2. *Proposition B.5.1 demonstrates that the convergence speed to the non-degenerate equilibrium point is independent of the specific linear inverse problem to be solved; the rate does, however, depend on γ. Analysis of the linear stability problem shows that $\gamma \approx 1.83$ gives the best local convergence rate; see Remark B.8.1. In the case where mean field preconditioning is not used, the optimal choice of γ for underdamped Langevin dynamics depends on the linear inverse*

problem being solved, and can be identified explicitly in the scalar setting [364]; for analysis in the non-Gaussian setting see [65]. Motivated by analogies with the work in [364, 65] we expect the optimal choice of γ to be problem dependent in the nonlinear case, but motivated by our analysis in the linear setting we expect to see a good choice which is not too small or too large and can be identified by straightforward optimization.

Now we show that for Gaussian initial data, the solution remains Gaussian and is thus determined by the evolution of the mean and covariance via the ODE (B.5.2). We prove this by investigating the mean field Fokker-Planck equation in the linear case. The evolution in time of the law $\rho(z)$ of equation (B.5.1) is governed by the equation

$$\partial_t \rho = \nabla \cdot \left(\begin{pmatrix} -p \\ C_q(\rho)(B^{-1}q - c) + \gamma p \end{pmatrix} \rho + \begin{pmatrix} 0 & -C_q(\rho) \\ C_q(\rho) & \gamma C_q(\rho) \end{pmatrix} \nabla \rho \right). \tag{B.5.3}$$

We prove the following result in Subsection B.8.4; by virtue of Proposition B.5.1 it establishes that the eigenvalues that determine the local stability of the posterior Gaussian density are the same across all linear Gaussian inverse problems:

Proposition B.5.3. *Let $m(t)$, $C(t)$ solve the ODE (B.5.2) with initial conditions m_0 and C_0. Assume that ρ_0 is a Gaussian distribution, so that*

$$\rho_0(z) := \frac{1}{(2\pi)^N} (\det C_0)^{-1/2} \exp\left(-\frac{1}{2} \|z - m_0\|^2_{C_0} \right)$$

with mean m_0 and covariance C_0. Then the Gaussian profile

$$\rho(t, z) := \frac{1}{(2\pi)^N} (\det C(t))^{-1/2} \exp\left(-\frac{1}{2} \|z - m(t)\|^2_{C(t)} \right)$$

solves the Fokker-Planck equation (B.5.3) with initial condition $\rho(0, z) = \rho_0(z)$.

B.6 Numerical Results

We introduce, and study, a numerical method for sampling the Bayesian inverse problem of Section B.2; the method is based on numerical time stepping of the interacting particle system (B.4.1). In this section, we demonstrate that the proposed sampler, which we refer to as *ensemble Kalman hybrid Monte Carlo* (EKHMC) can effectively approximate posterior distributions for two widely studied inverse

problem test cases. We compare EKHMC with its first-order version EKS [158] and a gold standard MCMC [46]. EKHMC inherits two major advantages of EKS: (1) exact gradients are not required (i.e., derivative-free); (2) the ensemble can faithfully approximate the spread of the posterior distribution, rather than collapse to a single point as happens with the basic EKI algorithm [400]. Furthermore, we show empirically that EKHMC can obtain samples of similar quality to EKS, and has faster convergence than EKS. We detail our numerical time-stepping scheme in the first subsection, before studying two examples (one low dimensional, one a PDE inverse problem for a field) in the subsequent subsections.

B.6.1 Time integration schemes

We employ a splitting method to integrate the stochastic dynamical system given by equation (B.4.1): the first capturing the finite particle approximation of the Hamiltonian evolution, and the second capturing an OU process in momentum space. The Hamiltonian evolution follows the equation:

$$\begin{aligned} \dot{q}^{(i)} &= p^{(i)}\,, \\ \dot{p}^{(i)} &= F_H := -C_q(Z)\Gamma_0^{-1}q^{(i)} - \frac{1}{I}\sum_{k=1}^{I}\langle \mathcal{G}(q^{(k)}) - \bar{\mathcal{G}}, \mathcal{G}(q^{(i)}) - y\rangle_\Gamma q^{(k)}\,. \end{aligned} \tag{B.6.1}$$

The OU process follows the equation:

$$\begin{aligned} \dot{q}^{(i)} &= 0\,, \\ \dot{p}^{(i)} &= -\gamma p^{(i)} + \sqrt{2\gamma C_q(Z)}\dot{W}^{(i)}\,. \end{aligned} \tag{B.6.2}$$

We implement a symplectic Euler integrator [397] for the part of the particle system arising from approximation of the Hamiltonian contribution to the damped-driven mean-field equations. That is, we take a half step $\epsilon/2$ of momentum updates, then a full step ϵ of position updates, and finally a half step $\epsilon/2$ of momentum updates. With the ensemble approximation the system is only approximately Hamiltonian; the splitting used in symplectic Euler is still well-defined, however. And we also expect it to perform well in the large particle limit because the mean-field limit is itself Hamiltonian. Let Z_j be the collection of all position and momentum particles at time j: $\{q_j^{(i)}, p_j^{(i)}\}_{i=1}^I$. Starting from time j this symplectic Euler integration gives map $Z_j \mapsto \hat{Z}_j$. We set $q_{j+1}^{(i)}$ to be the i^{th} position coordinates of $\hat{Z}_j$ and then update the momentum coordinates using the OU process which provides the damped-driven component of the mean-field limiting process. The damping coefficient $\gamma > 0$ is treated as a hyperparameter of EKHMC. Vector Z is set at the value given by output

of the preceding symplectic Euler integrator, denoted by $\hat{Z}_j$. The update of the i^{th} momentum coordinate, given by solving the OU process exactly in law, is then

$$\begin{aligned} \tilde{p}_j^{(i)} &= \exp(-\gamma\epsilon)\hat{p}_j^{(i)}\,, \\ p_{j+1}^{(i)} &= \tilde{p}_j^{(i)} + \eta, \quad \eta \sim \mathsf{N}(0, (1-\exp(-2\gamma\epsilon))C_q(\hat{Z}_j))\,, \end{aligned} \tag{B.6.3}$$

where $\hat{p}_j^{(i)}$ are the momentum coordinates from $\hat{Z}_j$. Within the OU process the damping coefficient $\gamma > 0$ is treated as a hyperparameter of EKHMC.

It is posible to consider use of a Metropolis-Hastings (Metropolization) step to correct the dynamics; however because the underlying continuous time system is not (for finite number of particles) invariant with respect to the target, doing so would be very complicated and so we do not pursue this. Aside from invariance, Metropolization also imparts stability of the integrator and we may address this in different ways. Indeed, similarly to [158], and as there with the goal of improving stability and convergence speed towards posterior distribution, we implement an adaptive step size, i.e., the true step size is rescaled by the magnitude of the "force field" F_H (defined in equation (B.6.1)):

$$\tilde{\epsilon} = \frac{\epsilon}{a|F_H| + 1}\,. \tag{B.6.4}$$

B.6.2 Low dimensional parameter space

We follow the example presented in Section 4.3 of the paper [158]. We start by defining the forward map which is given by the one-dimensional elliptic boundary value problem

$$-\frac{d}{dx}\left(\exp(u_1)\frac{d}{dx}p(x)\right) = 1\,, \quad x \in (0,1)\,, \tag{B.6.5a}$$

$$p(0) = 0\,, \quad p(1) = u_2\,. \tag{B.6.5b}$$

The solution is given explicitly by

$$p(x) = u_2 x + \exp(-u_1)\left(-\frac{x^2}{2} + \frac{x}{2}\right), \tag{B.6.6}$$

The forward model operator $\mathcal{G}$ is then defined by

$$\mathcal{G}(u) = \begin{pmatrix} p(x_1) \\ p(x_2) \end{pmatrix}. \tag{B.6.7}$$

Here $u = \left(u_1, u_2\right)^T$ is a constant vector that we want to find and we assume that we are given noise measurements y of $p(\cdot)$ at locations $x_1 = 0.25$ and $x_2 = 0.75$. Parameters are chosen according to [158], but we summarize them here for completeness:

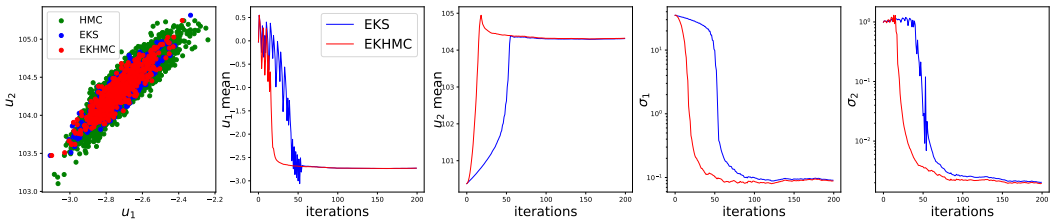


Figure B.1: The low dimensional parameter space example. From left to right: samples; mean u_1; mean u_2; the first singular value σ_1; the second singular value σ_2.

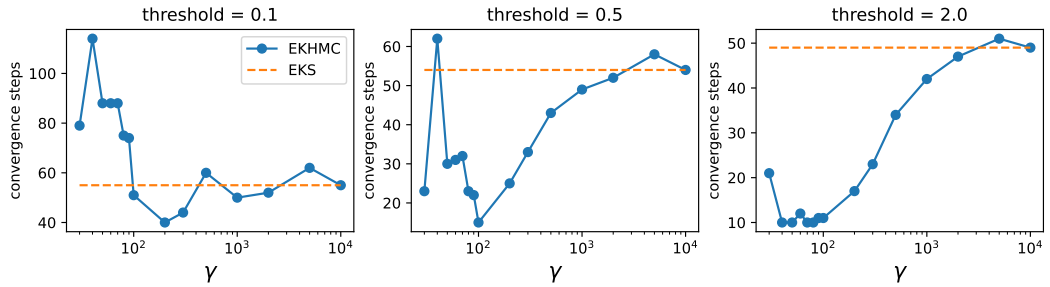


Figure B.2: Convergence time (of u_2 mean) as a function of damping coefficient γ. Left, Middle, Right: thresholds are 0.1, 0.5, 2.0, respectively. The takeaway from these plots is: large damping converges faster eventually (small threshold), while small damping converges faster initially (large threshold).

- noise $\eta \sim \mathsf{N}(0, \Gamma)$, $\Gamma = 0.1^2 I_2$;
- prior $\pi_0 = \mathsf{N}(0, \Gamma_0)$, $\Gamma_0 = \sigma^2 I_2$, $\sigma = 10$;
- measurement $y = (27.5, 79.7)^T$;
- number of particles $I = 10^3$;
- initialization: $(u_1, u_2) \sim \mathsf{N}(-3.5, 0.1^2) \times \mathsf{U}(70, 110)$.

Here U is the uniform distribution.

We choose $a = 0.01$, $\epsilon = 0.2$ in both EKS and EKHMC. We find choosing $\gamma = 1$ causes overshooting (the trajectory exhibits oscillatory rather than monotonic convergence), a phenomenon which can be ameliorated by increasing to $\gamma = 100$; indeed this latter value appears, empirically, to be close to optimal in terms of convergence speed. We conjecture this difference from the linear case, where the optimal value is $\gamma = 1.83$ (see Remark B.5.2), is due to the non-Gaussianity of the desired target: the particles have accumulated considerable momentum when entering the linear

convergence regime, and so extra damping is required to counteract this and avoid overshooting. Despite the desirable problem independence of the optimal γ in the linear case, we expect case-by-case optimization to be needed for nonlinear problems. We also empirically study how γ affects the convergence speed for this 2D problem: we sweep $\gamma \in [30, 10^4]$, compute the number of steps needed to be close to the convergent point by a threshold, shown in FIG. B.2. When the threshold is small, large γ converges faster; when the threshold is large, small γ converges faster. The takeaway is that: large γ converges faster eventually (small threshold), while small γ converges faster initially (large threshold).

We evolve the ensemble of particles for 200 iterations, and record their positions in the last iteration as the approximation of the posterior distribution, shown in Figure B.1. The EKHMC samples are quite similar to EKS samples, both of which are reasonable approximations of the samples obtained from the gold standard HMC [126] simulations – but both approximations of the gold standard miss the full spread of the true distribution, because of the ensemble Kalman approximation they invoke. We also compute four ensemble quantities to characterize the evolution of EKHMC and EKS: u_1 mean, u_2 mean, two eigenvalues σ_1 and σ_2 of the covariance matrix $C_q(Z)$. EKHMC has faster convergence towards equilibrium than EKS, benefiting from the second-order dynamics.

B.6.3 Darcy flow

This example follows the problem setting detailed in [158]. We summarize the essential problem specification here for completeness.The forward problem of porous medium flow, defined by permeability $a(\cdot)$ and source term $f(\cdot)$, is to find the pressure field $p(\cdot)$ where $p(\cdot)$ is solution of the following elliptic PDE:

$$\begin{aligned} -\nabla \cdot \Big(a(x)\nabla p(x)\Big) = f(x)\,, \quad x \in D = [0,1]^2\,, \\ p(x) = 0\,, \quad x \in \partial D\,. \end{aligned} \tag{B.6.8}$$

We assume that the permeability $a(x) = a(x;u)$ depends on some unknown parameters $u \in \mathbb{R}^d$. The resulting inverse problem is, given (noisy) pointwise measurements of $p(x)$ on a grid, to infer $a(x;u)$ or u. We model $a(x;u)$ as a log-Gaussian field. The Gaussian underlying this log-Gaussian model has mean zero and has precision operator defined as $C^{-1} = (-\triangle + \tau^2 \mathcal{I})^\alpha$; here $\triangle$ is equipped with Neumann boundary conditions on the spatial-mean zero functions. We set $\tau = 3$ and $\alpha = 2$ in the experiments. Such parametrization yields as Karhunen-Loeve (KL) expansion

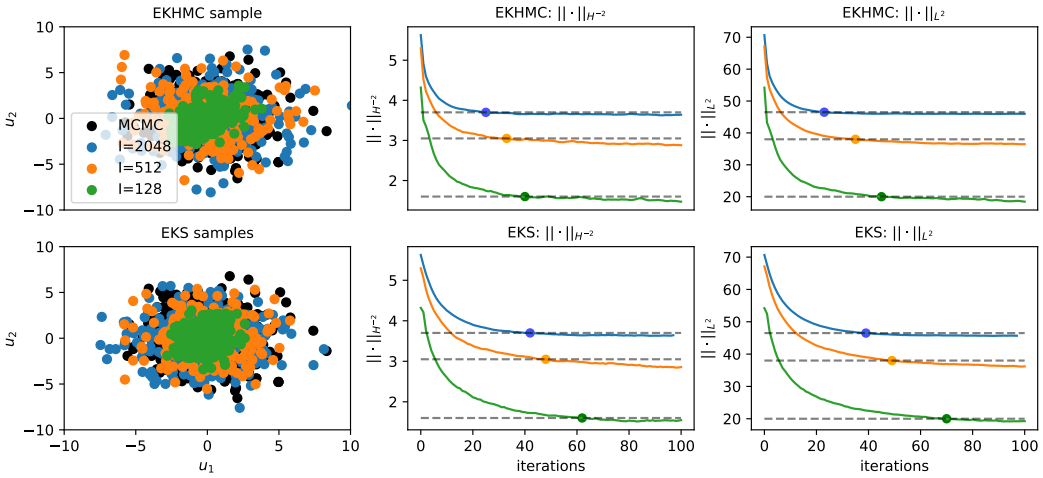


Figure B.3: Darcy flow. Left: samples obtained from EKHMC (top) and EKS (bottom), compared with MCMC. Middle: Evolution of $||u||_{H^{-2}}$ for EKHMC and EKS for different $I = 128, 512, 2048$. Right: The same as the middle, but $||u||_{L^2}$ instead of $||u||_{H^{-2}}$. EKHMC converges faster than EKS.

$$\log a(x;u) = \sum_{l \in K} u_l \sqrt{\lambda_l} \varphi_l(x)\,, \tag{B.6.9}$$

where $\varphi_l(x) = \cos(\pi\langle l, x\rangle)$, $\lambda_l = (\pi^2|\ell|^2 + \tau^2)^{-\alpha}$, $K \equiv \mathbb{Z}^2$. Draws from this random field are Hölder with any exponent less than one. In practice, we truncate K to have dimension $d = 2^8$. We generate a truth random field by sampling $u \sim \mathsf{N}(0, I_d)$. We create data y with $\eta \sim \mathsf{N}(0, 0.1^2 I_K)$. We choose the prior covariance $\Gamma_0 = 10^2 I_d$.

To characterize the evolution of EKHMC and compare it with EKS, we compute two metrics:

$$d_{H^{-2}}(\cdot) = \sqrt{\frac{1}{I}\sum_{j=1}^{I} ||u^{(j)}(t) - \cdot||^2_{H^{-2}}}\,, \quad d_{L^2}(\cdot) = \sqrt{\frac{1}{I}\sum_{j=1}^{I} ||u^{(j)}(t) - \cdot||^2_{L^2}}\,. \tag{B.6.10}$$

For these metrics, we use norms

$$||u||_{H^{-2}} = \sqrt{\sum_{l \in K_d} |u_l|^2 \lambda_l}\,, \quad ||u||_{L^2} = \sqrt{\sum_{l \in K_d} |u_l|^2}\,, \tag{B.6.11}$$

where the first is defined in the negative Sobolev space H^{-2}, whilst the second in the L^2 space.

We set $a = 0.01$ and $\epsilon = 1.0$ for both EKHMC and EKS. We set $\gamma = 1$ in EKHMC; unlike the previous example, this choice does not lead to over-shooting. We simulate

the particles for 100 iterations, and use their positions in the last iteration as the approximation to the posterior distribution. We compare the samples obtained from MCMC (we use the pCN variant on RWMH [103]), EKS ($I = 128, 512, 2048$) and EKHMC ($I = 128, 512, 2048$) in Figure B.3. The evolution of $||u||_{H^{-2}}$ and $||u||_{L^2}$ are plotted to show that convergence is achieved, and that EKHMC converges faster than EKS.

B.7 Conclusions

Gradient-free methods for inverse problems are increasingly important in large-scale multi-physics science and engineering problems; see [224] and the references therein. In this chapter we have provided an initial study of gradient-free methods which leverage the potential power of Hamiltonian based sampling. Analysis of the resulting method is hard, and our initial theoretical results leave many avenues open; in particular convergence to equilibrium is not established for the underlying nonlinear, nonlocal Fokker-Planck equation arising from the mean field model, and optimization of convergence rates over γ is not understood, except in the setting of linear Gaussian inverse problems. The Fokker-Planck equation associated with standard underdamped Langevin dynamics has been studied in the context of hypocoercity – see [20, 440] – and generalization of these methods will be of potential interest. Preconditioned HMC is also proven to converge in [37]; generalization to ensemble approximations would be of interest. On the computational side there are other approaches to avoiding gradient computations, yet leveraging Hamiltonian structure, which need to be evaluated and compared with what is proposed here [262].

B.8 Appendix

We present the proofs of various results from the chapter here.

B.8.1 Proof of Proposition B.1.1

Proof. The determination of the most general form of diffusion process which is invariant with respect to a given measure has a long history; see [129][Theorem 1] for a statement and the historical context. Note, however, that this theory is not developed in the mean-field setting and concerns only the linear Fokker-Planck equation. However, the dependence of the matrices $\mathcal{K}$, $\mathcal{J}$ and $\mathcal{M}$ on ρ readily allows use of the approach taken in the linear case. We find it expedient to use the exposition of this topic in [301]. The same derivation as used to obtain Eq. (5) from [301]

shows[3] that the density ρ associated with the dynamics (B.1.4), (B.1.5) satisfies Eq. (B.3.1); this follows from the (resp. skew-) symmetry properties of (resp. $\mathcal{J}$) $\mathcal{K}$ and the fact that $\mathcal{J}$, $\mathcal{K}$ and $\mathcal{M}$ are assumed independent of z, despite their dependence on ρ. It is also manifestly the case that the Gibbs measure is invariant for (B.3.1) since the mass term $\mathcal{M}$ appearing in $\mathcal{H}$ is independent of z so that

$$\rho \nabla \mathcal{H} + \nabla \rho = 0$$

when ρ is the Gibbs measure and (B.3.1) shows that then $\partial_t \rho = 0$. □

B.8.2 Proof of Proposition B.4.3

Proof. In fact, consider the affine transformation

$$q^{(i)} = Av^{(i)} + b\,, \quad p^{(i)} = Au^{(i)}\,,$$

along with

$$\tilde{\Phi}(v^{(i)}) = \Phi(q^{(i)}) = \Phi(Av^{(i)}+b)\,, \quad \tilde{H}(v^{(i)}, u^{(i)}) = H(q^{(i)}, p^{(i)}) = H(Av^{(i)}+b, Au^{(i)})\,.$$

Then we have that the gradient term scales like $\nabla_{v^{(i)}} \tilde{\Phi}(v^{(i)}) = A^T \nabla_{q^{(i)}} \Phi(q^{(i)})$ and the covariance preconditioner scales like $C_v = A^{-1} C_q A^{-1^T}$. The generalized square root scales like $\sqrt{C_v} = A^{-1}\sqrt{C_q}$. Therefore we can check that affine invariance holds for the particle systems (B.1.10) and (B.1.8). Affine invariance for the gradient free approximation (B.4.1) is more easily seen to be true. Finally for the Fokker-Planck equation (B.3.1), we can check similarly via the scaling of the terms. See a similar argument in the proof of Lemma 4.7 in [157]. □

B.8.3 Proof of Proposition B.5.1

Proof. We can take expectations in (B.5.1) to obtain the first two ODEs in (B.5.2). Now define

$$\hat{z} = z - m(\rho) \triangleq \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix}\,,$$

to obtain the following evolution for $\hat{z}$:

$$\begin{aligned} \dot{\hat{q}} &= \hat{p}\,, \\ \dot{\hat{p}} &= -C_q(\rho) B^{-1} \hat{q} - \gamma \hat{p} + \sqrt{2\gamma C_q(\rho)}\dot{W}\,. \end{aligned}$$

[3] We use the notational convention concerning divergence of matrix fields that is standard in continuum mechanics [174]; this differs from the notational convention adopted in [301].

Note that $C_q = \mathbb{E}[\hat{q} \otimes \hat{q}]$, $C_p = \mathbb{E}[\hat{p} \otimes \hat{p}]$, and $C_{q,p} = \mathbb{E}[\hat{q} \otimes \hat{p}]$. We can use Ito's formula and the above evolution to derive the last three ODEs in (B.5.2). The form of the steady state solutions is immediate from setting the right hand side of (B.5.2) to zero.

Now, we will establish the independence of the essential dynamics of the mean and covariance on the specific choice of B, c. To this end, denote $x_1 = B^{1/2}(B^{-1}m_q - c)$, $x_2 = B^{-1/2}m_p$, $X = B^{-1/2}C_qB^{-1/2}$, $Y = B^{-1/2}C_pB^{-1/2}$, $Z = B^{-1/2}C_{q,p}B^{-1/2}$. Applying this change of variables we obtain, from (B.5.2), the system

$$\begin{aligned} \dot{x_1} &= x_2\,, \\ \dot{x_2} &= -Xx_1 - \gamma x_2\,, \\ \dot{X} &= Z + Z^T\,, \\ \dot{Y} &= -XZ - Z^TX - 2\gamma Y + 2\gamma X\,, \\ \dot{Z} &= -\gamma Z - X^2 + Y\,. \end{aligned} \tag{B.8.1}$$

We notice that the steady state solutions take the form $(x_1, x_2, X, Y, Z) = (w, 0, X, X, 0)$ for a projection matrix $X^2 = X$ and w such that $Xw = 0$. The steady state with $X = I_N$ and $w = 0$ corresponds to the desired posterior mean. The transformed equation is indeed independent of B and c. Therefore the speed of convergence $(x_1, x_2, X, Y, Z) \to (0, 0, I_N, I_N, 0)$, within its basin of attraction, is indeed independent of B and c. The same is true in the original variables.

To complete the proof of stability it suffices to establish the local exponential convergence to the steady solution $(x_1, x_2, X, Y, Z) = (0, 0, I_N, I_N, 0)$. As a first step, we show that the following system converges to $(X, Y, Z) = (I_N, I_N, 0)$:

$$\begin{aligned} \dot{X} &= Z + Z^T\,, \\ \dot{Y} &= -XZ - Z^TX - 2\gamma Y + 2\gamma X\,, \\ \dot{Z} &= -\gamma Z - X^2 + Y\,. \end{aligned} \tag{B.8.2}$$

Linearization around $(X, Y, Z) = (I_N, I_N, 0)$ in variables of entries of $(X, Y, Z+Z^T)$ gives the matrix

$$\begin{pmatrix} 0 & 0 & 2I_{N^2} \\ 2\gamma I_{N^2} & -2\gamma I_{N^2} & -2I_{N^2} \\ -2I_{N^2} & I_{N^2} & -\gamma I_{N^2} \end{pmatrix}$$

whose eigenvalues satisfy $x^3 + 3\gamma x^2 + (2\gamma^2 + 6)x + 4\gamma = 0$ which all have strictly negative real part for $\gamma > 0$, since we can show the unique real eigenvalue is in

the interval $(-\gamma, 0)$. Therefore we conclude the local exponential convergence of covariances X, Y, Z in a neighbourhood of $(X, Y, Z) = (I_N, I_N, 0)$. The exponential convergence of means $(x_1, x_2) \to (0, 0)$ then follows linearizing the first two linear ODEs in the system (B.8.1) at $X = I_N$.

Similarly, for any other steady state $(X, Y, Z) = (X, X, 0)$ with a projection matrix $X \neq I_N$, we will show that there is an unstable direction in (B.8.2). In fact, since X is symmetric with only eigenvalues 1 and 0, there exists a nonzero vector v such that $Xv = 0$. Linearization around $(X, Y, Z) = (X, X, 0)$ in the direction (avv^T, bvv^T, cvv^T) gives the following 3×3 matrix for scalars a, b, c:

$$\begin{pmatrix} 0 & 0 & 2 \\ 2\gamma & -2\gamma & 0 \\ 0 & 1 & -\gamma \end{pmatrix}$$

whose eigenvalues satisfy $x^3 + 3\gamma x^2 + 2\gamma^2 x - 4\gamma = 0$. For all $\gamma > 0$ it may be shown that there exists a real eigenvalue in the interval $(0, 1)$, since $f(0)f(1) < 0$; thus there is an unstable direction determined by (a, b, c), and therefore in the original formulation. □

Remark B.8.1. *We can further investigate the spectral gap around* $(X, Y, Z) = (I_N, I_N, 0)$*, which is the absolute value in the real root of equation* $x^3 + 3\gamma x^2 + (2\gamma^2 + 6)x + 4\gamma = 0$*. To be precise, we can show that for* $x_0 = -\sqrt{12 - \sqrt{128}}$ *and* $\gamma_0 = -(4 + 3x_0^2)/(4x_0) \approx 1.83$*, the spectral gap is maximized as* $-x_0$ *and we expect fastest convergence for our method in the linear setting.*

To establish this claim we proceed as follows. By the intermediate value theorem, in order to show there always exists a root in $[x_0, 0)$ *and the spectral gap is at most* $-x_0$*, we only need to show that*

$$x_0^3 + 3\gamma x_0^2 + (2\gamma^2 + 6)x_0 + 4\gamma \leq 0\,.$$

The claim is true because $x_0 < 0$ *and* $2x_0(x_0^3 + 6x_0) = (4 + 3x_0^2)^2/4$*. By the basic inequality* $a^2 + b^2 \geq 2ab$*, we have*

$$x_0^3 + 3\gamma x_0^2 + (2\gamma^2 + 6)x_0 + 4\gamma = 2x_0\gamma^2 + (4 + 3x_0^2)\gamma + x_0^3 + 6x_0 \leq 0\,.$$

The maximal spectral gap is attained if and only if the equality in $a^2 + b^2 \geq 2ab$ *holds, i.e. when* $\gamma = \gamma_0$*.*

We also plot the spectral gap as a function of γ *to help visualize the dependence of the rate of convergence on damping for the linear problem; see Figure B.4. The*

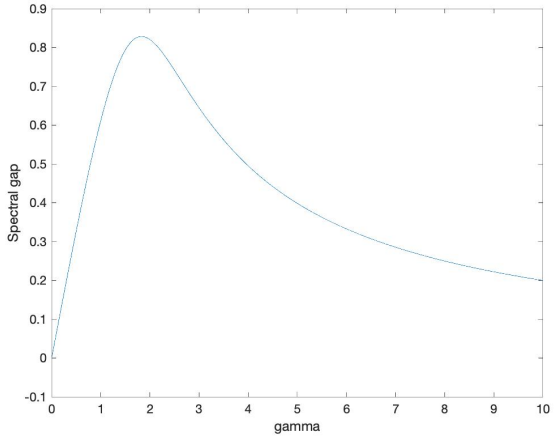


Figure B.4: Spectral gap as a function of γ

clear message from this figure is that there is a natural optimization problem for γ in this linear Gaussian setting; this is used to motivate searches for optimal parameters in the non-Gaussian setting.

B.8.4 Proof of Proposition B.5.3

Proof. Note that since $m(\rho)$, $C(\rho)$ are independent of the particle z, we have

$$\nabla\rho = -C(\rho)^{-1}(z - m(\rho))\rho\,,$$

$$D^2\rho = \Big(-C(\rho)^{-1} + \Big(C(\rho)^{-1}(z - m(\rho))\Big) \otimes \Big(C(\rho)^{-1}(z - m(\rho))\Big)\Big)\,\rho\,.$$

Using this we can substitute the Gaussian profile into equation (B.5.3). We compute the first term on the right hand to obtain

$$\left(-C(\rho)^{-1}(z - m(\rho)) \cdot \begin{pmatrix} -p \\ C_q(\rho)(B^{-1}q - c) + \gamma p \end{pmatrix} + \gamma N\right)\rho \triangleq I_1\rho\,.$$

The second term on the right hand side can be written as:

$$\begin{pmatrix} 0 & -C_q(\rho) \\ C_q(\rho) & \gamma C_q(\rho) \end{pmatrix} : \Big(-C(\rho)^{-1} + \Big(C(\rho)^{-1}(z - m(\rho))\Big) \otimes \Big(C(\rho)^{-1}(z - m(\rho))\Big)\Big)\,\rho \triangleq (I_2{+}I_3)\rho\,.$$

For the left hand side of (B.5.3), note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\det C(\rho) = \mathrm{Tr}\left[\det C(\rho)C(\rho)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}C(\rho)\right],\ \ \frac{\mathrm{d}}{\mathrm{d}t}C(\rho)^{-1} = -C(\rho)^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}C(\rho)\right)C(\rho)^{-1}\,.$$

Using the evolution of the mean (B.5.2) we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\|z-m(\rho)\|^2_{C(\rho)} =& 2\left\langle \frac{\mathrm{d}}{\mathrm{d}t}(z-m(\rho)), C(\rho)^{-1}(z-m(\rho))\right\rangle \\ &+\left\langle (z-m(\rho)), \frac{\mathrm{d}}{\mathrm{d}t}\left(C(\rho)^{-1}\right)(z-m(\rho))\right\rangle \\ =& 2\left\langle \begin{pmatrix} -m_p(\rho) \\ C_q(\rho)(B^{-1}m_q(\rho)-c)+\gamma m_p(\rho)\end{pmatrix}, C(\rho)^{-1}(z-m(\rho))\right\rangle \\ &-\left\langle C(\rho)^{-1}(z-m(\rho)), \frac{\mathrm{d}}{\mathrm{d}t}\left(C(\rho)\right)C(\rho)^{-1}(z-m(\rho))\right\rangle \triangleq 2(I_4-I_5)\,.\end{aligned}$$

Therefore we can compute the left hand side as:

$$\begin{aligned}\partial_t\rho &= \left[-\frac{1}{2}(\det C(\rho))^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}t}\det C(\rho)\right)-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|z-m(\rho)\|^2_{C(\rho)}\right]\rho \\ &= \left[-\frac{1}{2}\operatorname{Tr}\left[C(\rho)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}C(\rho)\right]-I_4+I_5\right]\rho \triangleq (I_6-I_4+I_5)\rho\,.\end{aligned}$$

In order to show that the Gaussian profile is indeed a solution to (B.5.3), we only need to show that $I_1+I_2+I_3=-I_4+I_5+I_6$. Note that

$$\begin{aligned}I_1+I_4 &= \gamma N + C(\rho)^{-1}(z-m(\rho))\cdot\begin{pmatrix} p-m_p(\rho) \\ -C_q(\rho)B^{-1}(q-m_q(\rho))-\gamma(p-m_p(\rho))\end{pmatrix} \\ &= \gamma N + C(\rho)^{-1}(z-m(\rho))\cdot\begin{pmatrix} 0 & I_N \\ -C_q(\rho)B^{-1} & -\gamma I_N\end{pmatrix}(z-m(\rho))\,.\end{aligned}$$

Defining $\hat{z}=z-m(\rho)$, we collect the terms in the to-be-proven identity $I_1+I_2+I_3=-I_4+I_5+I_6$ as:

$$\begin{aligned}&\hat{z}^T\left[C(\rho)^{-1}\begin{pmatrix} 0 & I_N \\ -C_q(\rho)B^{-1} & -\gamma I_N\end{pmatrix}+C(\rho)^{-1}\left(\begin{pmatrix} 0 & -C_q(\rho) \\ C_q(\rho) & \gamma C_q(\rho)\end{pmatrix}-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(C(\rho)\right)\right)C(\rho)^{-1}\right]\hat{z} \\ &+\gamma N-\operatorname{Tr}\left[\begin{pmatrix} 0 & -C_q(\rho) \\ C_q(\rho) & \gamma C_q(\rho)\end{pmatrix}C(\rho)^{-1}\right]+\frac{1}{2}\operatorname{Tr}\left[\frac{\mathrm{d}}{\mathrm{d}t}C(\rho)C(\rho)^{-1}\right] \triangleq \hat{z}^T M_1\hat{z}+M_2=0\,.\end{aligned}$$

We only need to show that $M_1+M_1^T=M_2=0$.

In fact, we compute $C(\rho)(M_1+M_1^T)C(\rho)$ to be:

$$\begin{pmatrix} 0 & I_N \\ -C_q(\rho)B^{-1} & -\gamma I_N\end{pmatrix}C(\rho)+C(\rho)\begin{pmatrix} 0 & -(B^{-1})^T C_q(\rho) \\ I_N & -\gamma I_N\end{pmatrix}+\begin{pmatrix} 0 & 0 \\ 0 & 2\gamma C_q(\rho)\end{pmatrix}-\frac{\mathrm{d}}{\mathrm{d}t}\left(C(\rho)\right)\,,$$

which equals zero upon blockwise computation via (B.5.2). Thus $M_1+M_1^T=0$. Moreover, note that $M_2=-\operatorname{Tr}\left[C(\rho)M_1\right]$: $M_1+M_1^T=0$ implies $C(\rho)(M_1+M_1^T)=0$ and taking the trace on both sides leads to the conclusion that $M_2=0$. Therefore the Gaussian profile indeed solves the Fokker-Planck equation (B.5.3). □