6. Conclusions and Future Directions

In this paper, we have introduced the fundamental concepts of ODEs and explored their application in modeling both discrete processes, such as neural network architectures, and continuous processes, such as flows. We further extended the discussion to the broader framework of differential equations, including SDEs, focusing on how these mathematical tools define generation processes in vision and language diffusion modeling. Along the way, we addressed practical challenges such as the stiffness of dynamical systems and the complexities of discrete diffusion in token spaces.

While the intersection of differential equations and deep learning has already produced remarkable results, several promising directions remain for future exploration:

- One direction is to consider the issues of scaling and numerical stability. As models scale to billions of parameters, ensuring the numerical stability of the underlying ODEs becomes increasingly difficult. Future research could investigate more robust adaptive solvers and regularization techniques to handle stiff dynamics in large-scale architectures without excessive computational overhead.

- It is also valuable to consider efficient Inference methods. Reducing the number of sampling steps remains a major priority for real-world deployment. Techniques like trajectory rectification, consistency distillation, and specialized high-order solvers offer paths to achieving high-fidelity generation in just one or a few steps.

- Applying continuous modeling to discrete data like text is not straightforward. Exploring native discrete flow matching and more sophisticated continuous relaxations may lead to more effective non-autoregressive language models that rival the quality of their autoregressive counterparts.

- The perspective of dynamical systems provides a unique lens for model interpretability. Analyzing fixed points, attractors, and the geometry of learned trajectories could provide deeper insights into how neural networks work on real-world problems.

- It is natural to extend the dynamical-systems viewpoint to a broader range of learning tasks, not only to the problems we discussed in this paper. For example, viewing optimization itself as a dynamical system may help connect training stability and generalization with classical results in stability and perturbation theory, and may provide a unified way to understand how algorithmic choices (e.g., step size schedules, momentum, or implicit updates) interact with model architecture and data.

Acknowledgements

This work was supported in part by the National Science Foundation of China (Nos. 62276056 and U24A20334), the Yunnan Fundamental Research Projects (No.202401BC070021), the Yunnan Science and Technology Major Project (No. 202502AD080014), the Fundamental Research Funds for the Central Universities (Nos. N25BSS054 and N25BSS094), and the Program of Introducing

Talents of Discipline to Universities, Plan 111 (No.B16009). The authors thank Junxiang Zhang and Hengyu Li for their valuable comments, which helped improve the manuscript.