Report on Martin Šípka's dissertation, "Machine learning through geometric mechanics and thermodynamics"

Michal Pavelka*[1]

[1]Mathematical Institute, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 18675 Prague, Czech Republic, pavelka@karlin.mff.cuni.cz
*Doctoral supervisor

April 29, 2024

As the supervisor of the thesis, I have witnessed the development of this thesis and the underlying research. Martin Šípka came from an area of mathematical modeling of non-Newtonian fluids Pelech et al. [2022] to a new topic – Machine learning (ML), which he explored with great enthusiasm and many new ideas.

In his first paper Šípka et al. [2023a] (published in J. Chem. Theory Comput.), he developed an automatic machine-learning-based method for the identification of collective variables in ab initio chemical kinetics. Collective variables represent the most relevant degrees of freedom that characterize the progress of a chemical reaction. They are important because, at the atomistic scale, chemical reactions consist of a large number of degrees of freedom, and without collective variables, rare-events sampling would be overwhelmingly difficult from the computational point of view. Instead of hand-picking the collective variables (for instance an important angle of the molecular conformation), the automatic method is more versatile and has proven successful in testing benchmarks as well as in zeolites, which are important as industrial catalysts. Martin Šípka is the first author of that paper as well as the main author of the machine-learning CV identification tool. Recently, the method has been generalized to a large class of zeolites, resulting in a Nature Communications paper Erlebach et al. [2024].

His second paper Šípka et al. [2023] was published at the prestigious International Conference on Machine Learning 2023 and is written with colleagues from the Faculty of Science, Charles University, University of Vienna, and MIT. The paper contributes to machine-learning optimization in differentiable simulations, where a biasing potential is sought that makes the rare events less rare. By several mathematical and computational improvements, taking gradient along the whole trajectory becomes more stable, which is an important step towards robust rare-events (for instance chemical reaction) simulations with potentials encoded as neural networks. Martin Šípka is again the first author of that paper.

In his third paper Šípka et al. [2023b], he found a method based on neural networks that identifies a mechanical system from snapshots of its trajectories. The neural networks encode the underlying Poisson bracket and energy that generate dynamics of the mechanical system. The method comes in three flavors, with implicitly valid Jacobi identity (necessary for the system to be Hamiltonian), with weakly enforced Jacobi identity, and without Jacobi identity. The code is then applied to simple systems such as oscillators and rigid bodies. In the case of rigid body dynamics, the flavor with implicitly valid Jacobi identity finds the Poisson bracket most precisely. In contrast, if the training data are trajectories of a dissipative system, the flavor without Jacobi identity becomes the most precise, which indicates the non-Hamiltonianity of the system. Martin Šípka wrote most of the code.

All three papers are remarkably elegant in the synergy between machine learning, chemical reactions, and differential geometry, which makes the thesis unique. The thesis then summarizes the findings clearly and concisely. Martin Šípka has shown a great level of independence and ability to navigate his research behind the frontier of current knowledge to acquire new important and interesting results. He has also been able to collaborate with several scientific groups. I wholeheartedly support his dissertation and recommend him for the Ph.D. degree at the Faculty of Mathematics and Physics, Charles University.

.....
Michal Pavelka

Prague 30/4/2024

References

A. Erlebach, M. Šípka, I. Saha, P. Nachtigall, C. J. Heard, and L. Grajciar. A reactive neural network framework for water-loaded acidic zeolites. *Nature Communications*, (Accepted), 2024. arXiv: 2307.00911.

P. Pelech, K. Tůma, M. Pavelka, M. Šípka, and M. Sýkora. On compatibility of the natural configuration framework with general equation for non-equilibrium reversible–irreversible coupling (generic): Derivation of anisotropic rate-type models. *Journal of Non-Newtonian Fluid Mechanics*, 305:104808, 2022. ISSN 0377-0257. doi: https://doi.org/10.1016/j.jnnfm.2022.104808. URL `https://www.sciencedirect.com/science/article/pii/S0377025722000556`.

M. Šípka, J. C. B. Dietschreit, L. Grajciar, and R. Gómez-Bombarelli. Differentiable simulations for enhanced sampling of rare events. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

M. Šípka, A. Erlebach, and L. Grajciar. Constructing collective variables using invariant learned representations. *Journal of Chemical Theory and Computation*, 19(3):887–901, 2023a. doi: 10.1021/acs.jctc.2c00729.

M. Šípka, M. Pavelka, O. Esen, and M. Grmela. Direct Poisson neural networks: learning non-symplectic mechanical systems. *Journal of Physics A: Mathematical and Theoretical*, 56(49), 2023b. `https://github.com/enaipi/direct-Poisson-neural-networks`.