Research Proposal: Blending Machine Learning with Scientific Computation

There are four ongoing projects towards which this fund will help accelerating the progress, which are listed in detail below. The first three have preliminary work already submitted, and the fund will be able to extend them further beyond these initial results. The last one is under preparation for a first submission, and the fund will enable its execution towards the completion. All the research mentioned below are initiated since the PI (Qijia Jiang) started at UC Davis in July 2024.

These efforts share a unifying goal — to enable data-efficient, interpretable, and computationally scalable algorithms for understanding complex dynamical and stochastic systems. The proposed work combines rigorous mathematical theory with computational innovation, spanning the spectrum from theoretical analysis to practical applications in physics, chemistry, and biology.

Project 1: Leveraging Optimization Perspective on Entropic Optimal Transport for Better Algorithm Design

The goal of this project is to build a general framework that bridges two subfields within applied mathematics: optimization and Optimal Transport, that were previously independently developed but never brought together. A recent discovery established that one can view the popular Sinkhorn algorithm traditionally used for solving the Entropic Optimal Transport problem as performing a Mirror Descent update, a well-known optimization algorithm, in an infinite dimensional probability space. In [1], this connection allows us to design better algorithms in terms of convergence rate compared to the state-of-the-art (exploiting a primal-dual perspective), while allowing a streamlined theoretical analysis. Planned future work involves relaxing some of the technical assumptions in the analysis, an improvement that will make the contribution stronger and showcase the strength of this new angle on this decades old problem, which finds wide applications in probabilistic modeling, operations research and image processing.

Project 2: Sampling Algorithms over Path Space

The goal of this project is to design algorithms to simulate time-series trajectories from a given posterior density over the path space – a problem with well-established roots in

computational chemistry (e.g., Transition Path Sampling). In work [2], we put forth 2 algorithms based on controlled measure transport, and Wasserstein gradient flow, respectively, that can be integrated with neural network training for this challenging problem, which traditionally required significant compute and memory footprint. The algorithm proposed therein works efficiently for a special case of the problem with a particular type of prior, which we'd like to generalize to a broader class as follow-up work, along with larger scale experiments to test the robustness of the proposed methodology.

Project 3: Neural PDE Solver

This project investigates how PDE solvers can benefit from neural networks. More specifically, we study algorithms that have a learning component baked in, instead of using generic hand-crafted update rules – this paradigm falls under the broader theme of meta-learning, or learning to learn. We show in [3] that given access to data (solutions to similar PDEs previously solved using traditional numerical methods), one can theoretically discover a better data-driven update with higher fidelity. At the same time, such methods can be trained end-to-end efficiently with the current AutoDiff pipeline in deep learning. With the main methodological framework laid out in the workshop presentation [3], ongoing future work involves conducting more numerical experiments on different types of PDEs and benchmark the results against established solvers. Being a cornerstone of applied math, PDEs are widely used for modeling complex dynamics across engineering disciplines (e.g., fluid dynamics).

Project 4: Transferrable Generative Model for Molecular Dynamics Trajectories

The goal of this project is to design efficient algorithms for MD simulations. We leverage Generative Model to amortize the computation cost by building numerical surrogates using existing data from traditional MD methods. Ongoing work [4] leverages Sequential Mote Carlo and Reinforcement Learning to perform sample-efficient distribution learning in a data-scarce regime without sacrificing out-of-domain generalization capabilities. This development will expand our capability to (1) do longer time scale dynamical simulation; (2) move beyond the static picture afforded by AlphaFold by capturing the true conformational ensembles. These are crucial for a deeper understanding of many phenomena in biology and chemistry, with transformative implications for protein design and drug discovery.

In conclusion, the award, if received, will enable the achievements of the goal of building an algorithmic foundation for scientific computation, and facilitate the broader agenda across the whole spectrum of data science, from fundamental theory and methodology development to application with real-world scientific impact in the next academic year.

References

[1] Vishwak Srinivasan, Qijia Jiang, “Designing Algorithms for Entropic Optimal Transport from an Optimisation Perspective”
[2] Qijia Jiang, Reuben Cohn-Gordon, “Machine-Learned Sampling of Conditioned Path Measures”
[3] Qijia Jiang, “When Differentiable Programming Meets Spectral PDE Solver”
[4] “Sample-Efficient Generative Model for Molecular Dynamics Trajectories via Twisted SMC”