Nicholas M. Boffi: Research Statement

Pushing beyond the limitations of classical approaches in applied and computational mathematics will require the development and application of data-driven methods in novel problem settings. I am an applied mathematician, and I am interested in learning, control, modeling, and simulation of complex dynamical and spatio-temporal systems that arise in the natural sciences and engineering. I have worked in diverse fields such as optimization, machine learning, adaptive control, soft matter physics, and condensed matter physics, with my primary strengths lying in the areas of dynamical systems and control theory, optimization and numerical methods, and scientific computing. As a postdoctoral researcher, I plan to develop principled machine learning methods grounded in rigorous theory for the solution of complex high-dimensional problems in applied mathematics such as learning dynamical systems, learning solutions to partial differential equations, and solving inverse problems in science and engineering. Below, I outline several topics of past, current, and future interest.

1. **Adaptive control theory:** I have developed and analyzed a new class of algorithms for adaptive control that incorporate local geometry into the learning process in a manner similar to mirror descent algorithms in optimization, leading to a publication in *Neural Computation*. I have performed the first general analysis of discrete-time nonlinear adaptive control algorithms, bridging classical adaptive control theory with modern reinforcement learning and leading to a submission to *Learning for Dynamics and Control*.

2. **Numerical methods for elastoplasticity:** I have extended a projection algorithm for simulating the deformation of hard elastoplastic materials in the quasi-static regime to three dimensions. I also designed an efficient parallel implementation, leading to publications in *Comp. Phys. Comm.* and *Phys. Rev. E.*. I have applied the method to study shear banding in bulk metallic glasses, a common precursor to failure in such materials.

3. **Optimization and machine learning:** I have developed a data-driven method for determining stability of a nonlinear dynamical system directly from trajectory data, leading to a publication in *Conference on Robot Learning*. I have analyzed via limiting stochastic differential equations a class of distributed stochastic optimization algorithms used for parallel training of deep neural networks, leading to a publication in *Neural Computation*.

4. **Numerical methods for condensed matter physics:** I have developed an efficient algorithm for calculating the Hartree-Fock exchange operator on a real-space grid via subspace projection, leading to a publication in *J. Chem. Th. Comp.* and a publication in *J. Chem. Phys.* I have implemented the algorithm in parallel as part of the the open source electronic structure package PARSEC.

Adaptive control theory

For many nonlinear dynamical systems, adaptively updating the control input to counteract modeling errors and unknown disturbances can offer significant performance improvements at little computational expense. The field of adaptive control studies when it is rigorously possible to do so through an online learning mechanism. However, the design and understanding of adaptive control algorithms with provable guarantees is a significant theoretical challenge, and most progress revolves around a few key advances. Designing new methods and providing new insights into existing approaches from the perspective of modern optimization and machine learning has been one focus of my research.

In collaboration with Prof. Jean-Jacques Slotine at MIT, I have developed a new suite of adaptive control algorithms by analogy to mirror descent methods in optimization theory [1]. Mirror descent methods incorporate local geometry into the optimization process to improve convergence guarantees [2], and provide a unifying perspective on many first-order online convex optimization algorithms. They are also known to

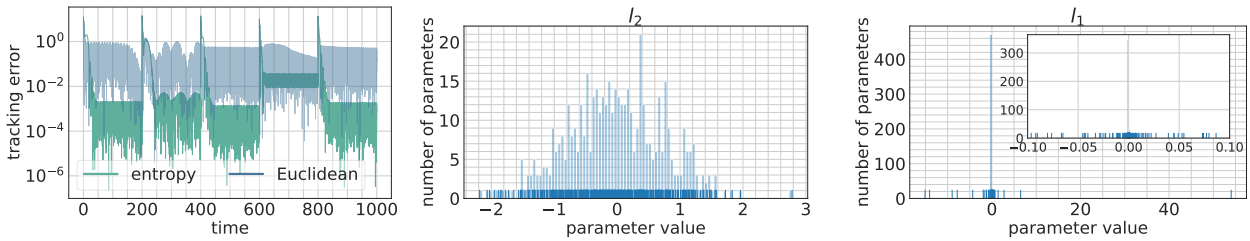

Figure 1: (Left) Euclidean adaptation compared to mirror descent-like adaptation for control with a convex combination of primitives. The problem geometry is respected by mirror descent with respect to the negative entropy, leading to several orders of magnitude improved performance of the mirror descent law. (Center/Right) Problem geometry can be imposed to learn models with desired characteristics. A standard l_2 (Euclidean) adaptive control law learns a Gaussian distributed vector of parameters (center), while a new l_1-regularized law leads to a sparse vector of learned parameters (right, inset shows blow-up). Performance is equivalent in both cases.

implicitly regularize the solution when used to solve underdetermined problems, which means that they preferentially converge to solutions with a specific structure, such as the minimum norm interpolant [3, 4].

To incorporate the advantages of mirror descent into adaptive control, I showed that replacing a quadratic with a Bregman divergence in standard Lyapunov arguments introduces local geometry into existing adaptive control algorithms. I proved that the incorporation of geometry implicitly regularizes the learned model (Figure 1, center/right), and I showed empirically that these new methods can improve convergence by exploiting knowledge of the problem geometry (Figure 1, left). Unlike in optimization, regularization cannot be directly added to adaptive control algorithms without affecting stability and performance; here, implicit regularization offers a principled way to regularize. I used implicit regularization to design methods for online identification of sparse, parsimonious models consistent with trajectory data, and I considered chemical reaction networks and Hamiltonian systems as two concrete examples. This result also answers a long-standing question about what parameters are actually found by classic adaptive control algorithms, and shows that such methods find the parameters of minimum l_2 norm that can control the system.

In a second project, I worked with Dr. Stephen Tu, a Research Scientist at Google Brain, to revisit nonlinear adaptive control theory from a modern reinforcement learning (RL) perspective. RL has been the subject of significant research interest after being used to surpass human performance in the game of Go [5] and after demonstrating strong promise in continuous control tasks such as robotics [6]. Both RL and adaptive control study the same fundamental problem – learning to control an unknown dynamical system – but do so from different perspectives. Adaptive control provides asymptotic guarantees for control of the system along a single trajectory, and analysis is typically performed in continuous time. Results in RL are generally obtained in the episodic setting, where multiple trajectories of the system are sampled, and discrete-time systems are analyzed. Convergence results are given through non-asymptotic guarantees on the *regret*, a characterization of performance defined by comparison to the best the learner could have done given all knowledge in hindsight.

To make results between the two fields comparable, we computed regret bounds for adaptive control algorithms in discrete time, providing the first finite-time guarantees for discrete-time adaptive control [7]. Discrete-time adaptive control algorithms are generally challenging to analyze due to the appearance of higher-order terms that disappear in continuous time. To sidestep this issue, we leveraged results from online learning theory not yet applied in the adaptive control context. In addition to proving regret bounds for existing methods, we developed a new class of online convex optimization-inspired algorithms, which were proven to maintain stability in the presence of noise and shown empirically to provide excellent control performance.

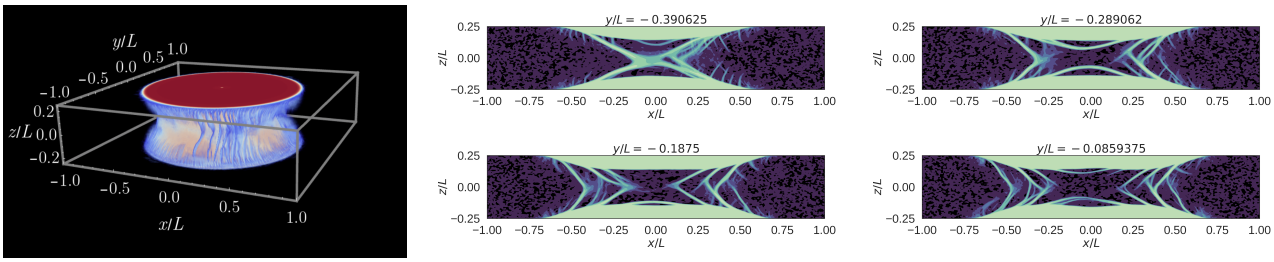

Figure 2: A simulation inspired by friction welding of a metallic glass, whereby disks of material on the top and bottom boundaries are set to rotate at specified rates. The left shows a three-dimensional visualization of the deformation in the material, while the right shows cross sections at fixed y value, revealing complex spatio-temporal shear banding dynamics. The simulation was performed on a $768 \times 768 \times 192$ computational grid.

Numerical methods for elastoplasticity

Many materials of scientific and engineering relevance exhibit elastoplastic behavior, which consists of reversible, elastic deformation below a yield stress, and irreversible, plastic deformation above this threshold. One prominent example is bulk metallic glass (BMG), a class of metallic alloy with an amorphous atomic-scale structure rather than the crystalline arrangement common to most metals. This disordered structure gives BMGs many promising characteristics, such as high tensile strength, high corrosion and wear resistance, and the ability to be thermally processed like plastics. However, BMGs are not widely used in applications as they often fail along a narrow localization of stress known as a shear band, a poorly understood instability. Shear banding has been studied in two dimensions, but there are few numerical characterizations of shear banding for realistic three-dimensional samples due to large computational expense.

In collaboration with Prof. Chris Rycroft at Harvard University, I developed a three-dimensional projection algorithm for simulating elastoplastic deformation in the quasi-static limit [8]. The algorithm exploits an analogy between the constitutive equations for quasi-static hypo-elastoplasticity and the Navier-Stokes equations for incompressible fluid flow [9], and is similar to Chorin's projection method [10, 11]. I designed a parallel implementation of the method in C++ and Open MPI, enabling efficient large-scale simulation of shear banding in three dimensions. Using the method, I was able to examine uniquely three-dimensional shear banding dynamics, such as in friction welding experiments (Figure 2).

A significant step in the algorithm is the solution of an elliptic partial differential equation (PDE) for the stress projection. The geometric multigrid algorithm often provides highly efficient solutions of linear systems that arise from discretizing such elliptic PDEs. We developed a parallelized, three-dimensional geometric multigrid solver in C++ and Open MPI capable of solving for arbitrary datatypes at each grid point, which we plan to release as open source software.

The deformation of BMGs is well-described by the shear transformation zone (STZ) theory [12], a first-principles continuum theory derived through statistical mechanics. A significant element of the STZ theory is an *effective temperature* field that characterizes local structural disorder. Physically accurate initial conditions for the effective temperature are difficult to construct, because it is not known how to connect the field to a microscopic configuration. Molecular dynamics (MD) – whereby individual atoms are simulated according to Newton's laws – provides an alternative simulation approach that does not require the introduction of a new continuum field, but incurs a higher computational expense, and as such is limited to microscopic systems and unrealistic loading rates.

One way to find initial conditions for large-scale continuum simulation would be to perform both simulations with the same geometry and to quantitatively match outputs. Attempts have been made in this direction [13], but a significant impediment is a difference in boundary conditions. Due to finite size effects, MD simulations typically employ Lees-Edwards boundary conditions, where periodic images of the domain are manipulated to impose shear stresses. By contrast, continuum simulations generally employ moving parallel plate boundary conditions, and naively implementing Lees-Edwards conditions poses computational challenges, such as grid misalignment between the periodic images and the primary domain.

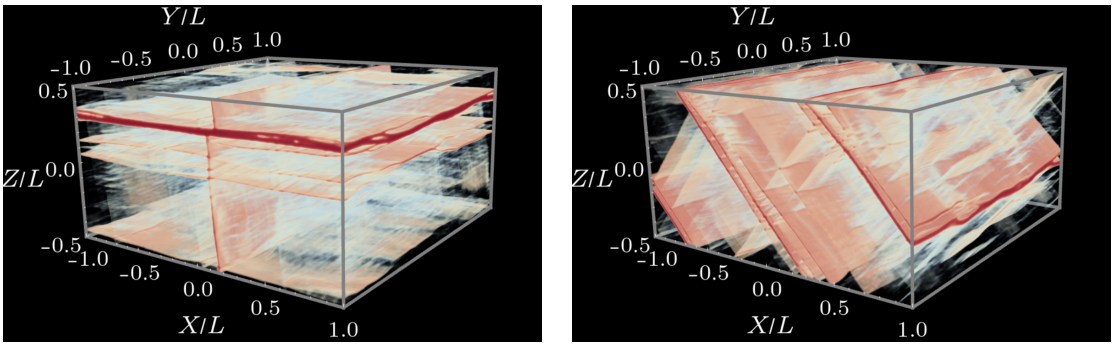

Figure 3: Three-dimensional continuum simulations of shear banding dynamics with Lees-Edwards conditions (left) and pure shear conditions (right). Boundary conditions are periodic in all directions in both cases, and shear stresses were imposed by explicitly deforming the problem domain. Both simulations were performed on a $512 \times 512 \times 256$ computational grid.

To address this issue, I developed a coordinate transformation methodology that decouples continuum boundary conditions from material deformation, enabling precisely matching boundary conditions between MD and continuum simulation [14]. A fixed reference domain is mapped to a deformed domain through a time-dependent transformation, and the transformation is chosen to impose material deformation. The stress projection is modified accordingly to incorporate transformation-dependent terms, and the resulting elliptic problem becomes significantly more complex. I leveraged a meta-programming scheme to automatically generate the linear system for an arbitrary transformation, so that boundary conditions can be specified through a single 3×3 matrix. Using this framework, I was able to study shear banding dynamics at continuum scale in a random, inhomogeneous environment subject to Lees-Edwards and pure shear boundary conditions (Figure 3). A two-dimensional variant of the method is now being used by Prof. Michael Shields at Johns Hopkins University to quantitatively match MD and continuum simulation.

Optimization and machine learning

Machine learning and data-driven methods are becoming increasingly prevalent across all fields of science and engineering, but many questions remain unanswered in both theory and practice. A consistent theme in my research has been feedback between dynamical systems theory and machine learning. I am particularly interested in how continuous limits and proof techniques from dynamical systems can give insight into machine learning and optimization algorithms, and how data-driven methods can be used for dynamical systems modeling and control.

In a project focused on applying dynamical systems theory to machine learning algorithms, I worked with Prof. Jean-Jacques Slotine at MIT to study distributed stochastic optimization algorithms from a continuous-time perspective, making use of stochastic differential equations [15]. I focused on consensus-based distributed architectures inspired by Elastic Averaging SGD (EASGD) [16], an algorithm commonly used to parallelize training of deep networks, but which had few formal guarantees. Using dynamical systems tools such as contraction analysis [17] and synchronization theory, I proved new convergence results in the convex setting, providing justification for the use of EASGD in practice. I categorized a trade-off between *noise magnitude* due to the stochastic approximation of the gradient and *synchronization* of the individual optimization trajectories due to distributed coupling. Using this categorization, I was able to explain the dependence of the generalization error of deep networks on several hyperparameters.

In a separate work, I collaborated with Dr. Vikas Sindhwani and Dr. Stephen Tu at Google Brain to devise a statistical learning methodology for determining the stability of a dynamical system from trajectory data alone, obviating the need for explicit knowledge of the dynamics [18]. Complex control policies obtained through deep reinforcement learning have led to a recent surge in laboratory performance of modern robotic platforms, but practitioners will be hesitant to deploy these policies in the real world without guarantees of safety. While powerful techniques such as Lyapunov theory [19] exist for determining the stability and

safety of a given dynamical system, these methods assume that the dynamics is known, which makes them inapplicable to many systems of modern interest.

To address this problem, we formulated a general optimization procedure that captures several notions of learning stability. Adapting existing theory for classifiers [20], we proved that function approximation can be used for stability estimation, and categorized the sample complexity for several function classes of interest, such as Reproducing Kernel Hilbert Space predictors and deep neural networks. We showed that stability can be learned efficiently in practice, and moreover that verification of stability can be used for downstream tasks such as adaptive control.

Condensed matter physics and quantum chemistry

Density functional theory (DFT) [21] is a prominent method for computing ground-state electronic properties of condensed matter systems. In recent years, hybrid functionals, which mix a fraction of the Hartree-Fock exchange-correlation functional with standard exchange-correlation functionals, have shown impressive success in matching computations to experimentally-determined electronic properties. The Hartree-Fock exchange can be efficiently computed with chemically-motivated Gaussian basis sets, but these bases do not provide a systematic way to increase accuracy of the solution. Classical methods from approximation theory and numerical analysis – such as finite differences, finite elements, and discontinuous Galerkin methods – offer a compelling solution, as it is possible to rigorously quantify the gains in accuracy obtained by increasing the grid size or resolution. However, calculation of the exchange operator is prohibitively expensive, because its expense scales poorly with the grid size, and because the operator is computed hundreds or thousands of times as a subroutine in an eigenvalue calculation. Without an efficient way to compute the exchange operator, hybrid functionals cannot be effectively used with these principled, mathematically-motived methods.

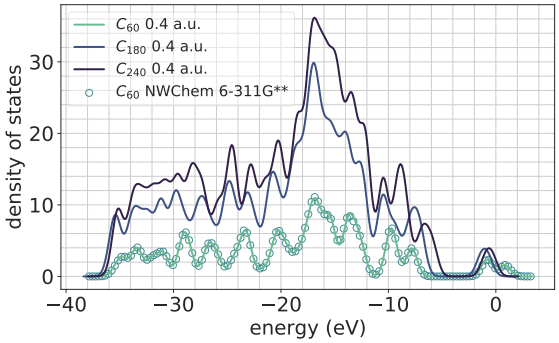

Figure 4: A density of states calculation for complex many-electron molecules. Our method (solid) matches existing approaches (e.g. a Gaussian basis set calculation using NWChem, open circles for C_{60}) and can scale to large systems such as C_{240}, which has 480 electrons.

Based on prior work in the plane wave basis [22], I developed an efficient algorithm for computing the Hartree-Fock exchange on a real-space grid in collaboration with Prof. Amir Natan at Tel Aviv University [23]. The operator is projected onto the span of occupied states and a small number of low-lying unoccupied states, so that its computational expense is primarily set by the number of electrons in the system, rather than the grid size. We implemented the method in parallel using OpenMP in the open source electronic structure code PARSEC [24]. Our method obtains a 100× speedup in comparison to calculation of the full operator, and can be used to efficiently calculate the Hartree-Fock energies, dipole moments, and polarizabilities of complex molecules with essentially no loss of accuracy (Figure 4).

Within Hartree-Fock theory, the energy of the highest occupied molecular orbital is often interpreted as the ionization potential of the molecule, and the energy of the lowest unoccupied molecular orbital (LUMO) is often interpreted as the electron affinity. Despite this common interpretation, there is a large discrepancy between LUMO energies computed with different bases, and it is not clear that larger bases will lead to a higher-accuracy calculation. Because a real-space grid provides a clear way to increase the basis size, we used our method to analyze the dependence of unoccupied state energies and wavefunctions on the grid size [25]. We showed that for neutral molecules, the LUMO is quantitatively predicted by a particle in a spherical well as the grid size tends to infinity. For molecules with unoccupied states with negative energies, small grid sizes produce a confinement effect that leads to physically incorrect positive energy values.

References

[1] Nicholas M. Boffi and Jean-Jacques E. Slotine. Implicit regularization and momentum algorithms in nonlinear adaptive control and prediction. *Neural Computation (accepted)*, 2020.

[2] Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167 – 175, 2003.

[3] Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. Characterizing implicit bias in terms of optimization geometry. volume 80 of *Proceedings of Machine Learning Research*, pages 1832–1841, Stockholmsmässan, Stockholm Sweden, 2018. PMLR.

[4] Navid Azizan, Sahin Lale, and Babak Hassibi. Stochastic mirror descent on overparameterized nonlinear models: Convergence, implicit regularization, and generalization. *International Conference on Machine Learning Generalization Workshop*, 2019.

[5] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[6] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *J. Mach. Learn. Res.*, 17(1):1334–1373, 2016.

[7] Nicholas M. Boffi, Stephen Tu, and Jean-Jacques E. Slotine. Regret bounds for adaptive nonlinear control. *Learning for Dynamics and Control (submitted), arXiv:2011.13101*, 2020.

[8] Nicholas M. Boffi and Chris H. Rycroft. Parallel three-dimensional simulations of quasi-static elasto-plastic solids. *Computer Physics Communications*, 257:107254, 2020.

[9] Chris H. Rycroft, Yi Sui, and Eran Bouchbinder. An Eulerian projection method for quasi-static elastoplasticity. *Journal of Computational Physics*, 300:136–166, 2015.

[10] Alexandre J. Chorin. A numerical method for solving incompressible viscous flow problems. *Journal of Computational Physics*, 2(1):12–26, 1967.

[11] Alexandre J. Chorin. Numerical solution of the Navier-Stokes equations. *Mathematics of Computation*, 22(104):745–762, 1968.

[12] James S. Langer. Shear-transformation-zone theory of plastic deformation near the glass transition. 77(2):021502, 2008.

[13] Adam R. Hinkle, Chris H. Rycroft, Michael D. Shields, and Michael L. Falk. Coarse graining atomistic simulations of plastically deforming amorphous solids. *Physical Review E*, 95(5):053001, 2017.

[14] Nicholas M. Boffi and Chris H. Rycroft. Coordinate transformation methodology for simulating quasistatic elastoplastic solids. *Phys. Rev. E*, 101:053304, 2020.

[15] Nicholas M. Boffi and Jean-Jacques E. Slotine. A continuous-time analysis of distributed stochastic gradient. *Neural Computation*, 32(1):36–96, 2020.

[16] Sixin Zhang, Anna Choromanska, and Yann LeCun. Deep learning with elastic averaging SGD. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, page 685–693, Cambridge, MA, USA, 2015. MIT Press.

[17] Winfried Lohmiller and Jean-Jacques E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998.

[18] Nicholas M. Boffi, Stephen Tu, Nikolai Matni, Jean-Jacques E. Slotine, and Vikas Sindhwani. Learning stability certificates from data. *Conference on Robot Learning (CoRL)*, 2020.

[19] A. M. Lyapunov. *The general problem of the stability of motion (in Russian)*. PhD thesis, University of Kharkov, 1892.

[20] Nathan Srebro, Karthik Sridharan, and Ambuj Tewari. Smoothness, low-noise and fast rates. *Neural Information Processing Systems*, 2010.

[21] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.

[22] Ivan Duchemin and François Gygi. A scalable and accurate algorithm for the computation of Hartree–Fock exchange. *Computer Physics Communications*, 181(5):855 – 860, 2010.

[23] Nicholas M. Boffi, Manish Jain, and Amir Natan. Efficient computation of the Hartree–Fock exchange in real-space with projection operators. *Journal of Chemical Theory and Computation*, 12(8):3614–3622, 2016.

[24] http://real-space.org/.

[25] Nicholas M. Boffi, Manish Jain, and Amir Natan. Asymptotic behavior and interpretation of virtual states: The effects of confinement and of basis sets. *The Journal of Chemical Physics*, 144(8):084104, 2016.