THE UNIVERSITY OF CHICAGO | **DEPARTMENT OF STATISTICS**

Alexander Strang
William H. Kruskal Instructor

alexstrang@statistics.uchicago.edu

www.alexanderstrang.com

I am an applied mathematician interested in the interaction of structure and dynamics. I address problems involving nonequilibrium thermodynamics, evolution and selection dynamics, and information flow in neural networks. To address network structure I use tools adapted from simplicial homology and spectral graph theory. To describe the structure of objective functions, I use tools adapted from random matrix theory, spectral embedding, and function approximation. I consider stochastic processes motivated by molecular biology, neuroscience, biophysics, and evolutionary game theory. Some of these projects are summarized below. My complete research profile is available at alexanderstrang.com/research.

Linking Structure to Dynamics

Any study linking structure to dynamics requires a language that describes structure, and a set of dynamical properties to evaluate. The appropriate language and properties of study depend on the application area. I largely focus on processes on networks, where structure can be effectively summarized using tools adapted from simplicial homology, and on selection and optimization, where structure is best understood as the shape of the payout or objective function.

Networks

To describe network structure, I use tools adapted from the combinatorial Helmholtz-Hodge decomposition (HHD). The HHD adapts the Helmholtz decomposition of a vector field to a decomposition of edge flows on networks. It is one of a sequence of decompositions induced by considering a network as a simplicial complex, then introducing a sequence of linear operators associated with the simplices. The simplest pair of operators are analogous to a discrete gradient and curl, so induce a decomposition analogous to the Helmholtz decomposition. The rank and nullity of the operators are widely used in persistent homology to characterize the topological properties of a graph.

The combinatorial HHD is a powerful tool for separating an edge flow into a cyclic and conservative component. In an arbitrary discrete-space continuous-time Markov chain with microscopic reversibility the log ratio of forward and reverse transition rates is analogous to the work required to cross an edge. This is an edge flow and can be used to construct an analogous physical system. Applying the combinatorial HHD to this edge flow extracts cyclic (external) energy sources from a potential function representing the internal energy of the system as in the Schnakenberg formulation of statistical mechanics.

In statistical thermodynamics, a Markov chain models an equilibrium system if all probability fluxes across pairs of forward and backward edges vanish at the steady state. Such systems obey detailed balance, are time-reversible, and admit a Boltzmannian steady state. Crucially, all equilibrium systems are energetically closed. Most life processes are not energetically closed. The continual use of energy to perform a function is the defining hallmark of active matter. As such, there is great interest in non-equilibrium systems where energy exchange maintains nonvanishing, cyclic probability currents. Such nonequilibrium systems are much harder to study than equilibrium systems since they do not admit a Boltzmannian steady state and are not time-reversible. Indeed, no closed form exists relating the steady state to the underlying energetics.

The Helmholtz-Hodge decomposition (HHD) separates the work to cross each edge into a conservative component associated with internal energy, and a rotational component associated with external energy sources. In the near equilibrium case, when driving rotation is weak, I introduced a formal expansion of the steady state distribution and currents as a Taylor series. Each term in the series satisfies a recursively defined HHD. The recursion couples perturbations to the steady state distribution and currents. I derived a nonzero lower bound on the radius of convergence of the expansion. Thus, I demonstrated that the HHD provides the natural description of systems near equilibrium. I then used the expansion to prove classical results from linear thermodynamics including Onsager reciprocity. I also used the expansion to provide the essential algebraic conditions which a system must satisfy so that its steady state is independent of driving rotation up to a chosen order and related the thermal efficiency of the system to the degree of

perturbation in the steady state. This work was the foundation of my dissertation [1].

I have since analyzed the relation between structure and dynamics in a variety of applications.

For example, neural circuits encode and transmit information via a network of coupled dynamical processes. These processes are often strongly nonlinear, may exhibit chaos, and are subject to noise. Here, understanding the relation between network structure (connectomics and synaptic weights), and network dynamics (information processing) is one of the central objectives of computational neuroscience. Over the past year, I have co-advised a master's student with Brent Doiron. The student developed model-free estimators which predict the causal relations between neurons from their recordings alone. The estimators use transfer entropy to quantify information flow. We adapted non-parametric estimators for transfer entropy and tested the relation between information flow graphs and the underlying connectomics via simulation. We aim to study these patterns in RNN's, or neural models with synaptic plasticity, where the dynamics inform the connectivity, and wiring adapts to information flow.

Real neural circuits exhibit log-normally distributed synaptic weights. It is unknown whether the distribution of weights is correlated with neuron identity, or whether it is independently sampled at each synapse. That is, it is unknown whether or not the weights are heavy-tailed since some neurons exhibit much stronger connections than others. There is no empirical method capable of directly distinguishing these cases. It is likely true that the dynamics of a network with weights organized by nodes would differ from the dynamics when the weights are not organized by the nodes. I aim to develop dynamical signatures which distinguish these cases by exploiting work I developed to describe the structure of competitive systems.

Systems of competing agents are observed across biology, social science, and AI. A system of competing agents consists of a population engaged in a series of competitive interactions. Advantage in the competitive events may be expressed by a payout function which returns the advantage one agent possesses over another given their relevant attributes. This payout function may represent a game or preference between options in a comparison task. A finite ensemble of agents defines a tournament, where edges between agents represent possible interactions, and the advantages agents posses are expressed as an edge flow. Then, the HHD extracts a component of the flow that is consistent with a ranking of the agents, and a component that represents all cyclic competitive relations. The sizes of these two components measure the degree to which the population can be organized into a hierarchy. The degree of hierarchy has significant consequences in nearly every domain where competition is studied, is a central feature used to characterize natural populations, and determines which choice problems are tractable.

I proved that, if advantage in a competitive system is determined by agent attributes, and those attributes are sampled independently, then the expected degree of hierarchy in the system is determined exclusively by the number of agents in the population, the number of competitive interactions between agents, the variance in advantage between any pair, and the correlation in advantage on edges sharing an agent. The stronger the correlation the more advantage is organized by node identity, and the weaker the correlation the closer advantage is to independent. Systems that are node-organized exhibit clearer hierarchies since the expected degree of cyclicity decreases monotonically in the correlation between adjacent edges. Those relations are fully explained in a first author paper published in SIREV in 2022 [2].

To link these results to adapting populations, I proved that populations of similar agents tend to form hierarchies. With an undergraduate student, I computed convergence rates in expectation to a perfect hierarchy by analyzing the asymptotic behavior of the correlation in advantage on adjacent edges. We showed that those rates depend on the size of the self-play gradient, the curvature of the advantage function, and whether the population concentrates at a boundary or interior of the trait space. More strongly, populations that converge to perfect hierarchy in expectation converge to hierarchy in probability. We tested these results in a series of agent-based training models adapted from a Moran process, and by analyzing the asymptotic behavior of the continuous replicator equation widely with diffusive mutation. We solved the ensuing reaction-diffusion equation and computed rates of convergence to hierarchy.

This analysis raises interesting questions regarding evolution and training in competitive systems. For example, it is widely hypothesized that biodiversity in competitive systems is maintained by cyclic com-

petition. We demonstrate the opposite, cyclic competition requires diversity, suggesting a feedback loop that may drive the rapid evolution of hierarchies after a sufficient collapse of diversity. When applied to comparison tasks, these results suggest that, in a very quantitative sense, it is easier to compare apples to apples than apples to oranges. Our work is in review at SIAP [3].

To better address random flow structures, I completed a comprehensive study of the expected structure of random edge flows. Random edge flows arise whenever we need to study ensembles, an edge flow is the result of a random or chaotic process, or is corrupted by noise. Global structure can be summarized by defining a series of measures associated with projection onto principled subspaces. Each subspace is associated with a graph operator, chosen, as in the HHD, to model a canonical family of flows. Projection onto the subspaces decomposes an edge flow into components whose sizes characterize its structure.

In collaboration with Lek-Heng Lim, I showed that the expected sizes of the components depend only on the first two moments of the flow distribution. In particular, the expected measures depend on a matrix inner product between the projector onto the desired subspace and the covariance in the flow. This inner product can be simplified when working in a natural basis that adapts to the graph topology. We use powers of the operator applied to its adjoint. The operator applied to its adjoint acts as a Laplacian for an edge graph, where adjacency in the edge graph is controlled by the overlap of motivic flows stored in the operator's columns. Powers of the Laplacian model paths of increasing length through the edge graph. when an edge flow is node-organized (a function of independently sampled node attributes), then the edge flow covariance is always a linear function of the gradient applied to the divergence. Covariances in this family only model correlation in flows on adjacent edges. By extending the family to power series, we retain the simplicity of the extant analysis, while modeling correlations to all lengths in the edge graph. The expected structure of an edge flow is invariant under the projection of its covariance matrix onto the power series family, so variations in covariance away from the family have no influence on expected structure. Moreover, edge flow distributions with covariances in the family can be easily analyzed, since the expected structure is a monotonic function of the correlations at increasing distances.

More strongly still, the projectors themselves may be expanded in powers of the operator and its adjoint. This expansion can be truncated to produce accurate approximations of expected structure using a minimal set of correlations. This analysis is closely related to the spectrum of the edge graph, since the number of distinct eigenvalues of the edge graph Laplacian determines the dimension of the power series family. Highly regular graphs have few distinct eigenvalues so may be succinctly summarized using only a few short-distance correlations. Given an ensemble of graphs with a known distribution of eigenvalues, we derived optimal approximations via the sequence of orthogonal polynomials associated with the spectral distribution. Tests on a series of graph families suggest that the number of statistics needed depends on the coefficient of variation in the edge geaph spectrum. For most graph families, the number of correlations needed scales with the diameter of the graph — an enormous complexity advantage.

In the future, I will apply these results to describe patterns of information flow in neural circuits and to distinguish node-organized neural circuits from edge-organized circuits using observed dynamics alone.

Optimization and Selection

Consider optimization and selection. In optimization, the behavior of an optimizer is determined by the shape of the objective function. In evolving populations the behavior of the population is determined by the form of the function which compares agents, and returns the advantage one possesses over the other. Examples arise in agent training tasks where the quality of an agent cannot be evaluated directly, in explicitly competitive training environments like GANs, in natural populations subject to competitive selection, and in evolutionary game theory.

My work on optimization is largely motivated by Bayesian inference, where I have completed a full research program. In particular, I have studied a family of hierarchical Bayesian models introduced by [4]. Here a signal x is sampled from a conditionally Gaussian prior, with variances drawn from a generalized gamma hyperprior. We observe noisy data b generated by applying a forward map A to x then adding Gaussian noise. The forward map is often linear but is usually underdetermined. Thus the inverse problem, recover

x from b, is not well-posed and requires prior information.

The hierarchical model described above is both expressive and computationally tractable. Since the generalized gamma distribution is not a conjugate prior, the resulting posterior is not Gaussian, and its form depends on a set of hyperparameters. Different hyperparameters produce posterior distributions with varying tail behavior, sparsity-promoting properties, and convexity. Accordingly, hyperparameter selection, and robustness to variations in the hyperparameters, are important issues. The model is tractable because coordinate ascent exploits the hierarchical structure of the priors. My work explores relations between the shape of the regularizer, as determined by the hyperparameters, the convexity of the ensuing optimization problem, the sparsity of solutions, convergence rates to those solutions, and the sensitivity of those solutions to changes in the hyperparameters.

As a graduate student, I analyzed the convergence rate of a maximum a posteriori (MAP) estimation scheme based on coordinate ascent. I proved that the scheme converged quadratically off the assumed support, but converged linearly on the assumed support. Convergence is linear on the support since coordinate ascent reduces to a reweighted least squares algorithm, and, at each stage, increasing an entry of the signal requires increasing a weight, which requires increasing an entry. The form of the update for the weight depends on the hyperparameters and is not known in closed form. I showed that, for any choice of hyperparameters, the update function can be called efficiently by solving a one-dimensional initial value problem offline. I then used the initial value problem to analyze the asymptotic behavior of the update function. These asymptotics determine the shape of the effective regularizer introduced by the hyperprior. I showed that the MAP objective function is equivalent to a regularized least squares objective, where the regularization term is a smoothed ℓ_p penalty with flexible p and smoothing determined by the hyperparameters. This work led to a pair of papers in Inverse Problems [5,6].

As a Kruskal instructor, I have collaborated with Daniel Sanz-Alonso on a series of extensions to the hierarchical framework. Those extensions follow. First, existing work only considered point estimators, so did not fully exploit the Bayesian setting. In work with a master's student, we developed a variational inference scheme that approximates the true posterior with a simpler distribution which can be used for uncertainty quantification and interval estimation. That work has been accepted at SIAM JUQ [7]. Existing work also focused exclusively on linear forward maps. In joint work with a shared PhD student, we introduced a Kalman filtering approach which uses coordinate ascent to solve a nonlinear inverse problem in the hierarchical setting. This approach introduces new regularization techniques to Kalman filtering and extends the MAP estimation methods used in the hierarchical setting to nonlinear models [8].

Since the form of the estimation problem, and corresponding estimator, depends on the prior assumptions, it is important to analyze the robustness of solutions under changes to the hyperparameters. I led two master's students through this analysis. We derived an ODE linking changes in the underlying assumptions to changes in the resulting estimator and introduced an accurate predictor-corrector method which solves the ODE. A user can then trace the path of estimators associated with a continuous hyperparameter path. The ODE evaluates the sensitivity of estimation to assumptions. It can be used for minimum selection in the non-convex regime via convex relaxation. The latter advance is crucial since the non-convex regime is strongly sparsity-promoting. By example, we demonstrated that the minima selected via convex relaxation are much more accurate than minima selected by direct solution, or by hybrid methods such as [9].

The predictor-corrector method introduced to study the sensitivity of estimation to prior assumption required careful analysis of the Hessian of the MAP objective function. The Hessian is a highly nonlinear matrix-valued function, evaluated at an estimator whose value is only ever approximated and cannot be expressed analytically. Moreover, the Hessian is extremely ill-conditioned near sparse solutions and requires careful treatment to invert. By exploiting the structure of the Hessian, I proved that, after a change of coordinates and rescaling, the Hessian is a near-to-low rank perturbation of an explicitly invertible tridiagonal matrix. Using this form, we showed that the Hessian is invertible for almost all inputs restricted to the manifold of possible MAP estimators, and, consequently, that the path-tracing ODE is well-posed. I also designed an efficiently computable preconditioner that approximates the inverse to the thresholded Hessian. This preconditioner dramatically reduces the conditioning of the Hessian by a factor of over 10^6,

and speeds iterative solvers. It enables second-order accurate optimization methods in the hierarchical framework. These methods significantly accelerate convergence on the support..

This research program produced two minisymposium talks, two seminar talks, two published papers, one accepted paper, two papers in review, and trained four graduate students.

My work on evolving populations relates the form of a payout function to selection dynamics. In collaboration with industrial partners at Lockheed Martin's Artificial Intelligence Center, I generalized an embedding scheme introduced by Google Deepmind. The embedding scheme represents the payout of a zero-sum game by embedding into a sequence of low-dimensional spaces each subject to a canonical game. Deepmind's team demonstrated that any payout function, sampled by a finite population, can be embedded pointwise. I developed a rigorous function theory which embeds the entire payout function, proved that the necessary embedding is recovered by the eigenfunctions of an integral operator whose kernel is the payout, and introduced a practical variational algorithm that recovers the embedding. I then developed a rigorous approximation theory for the algorithm, showed that it is at least as accurate as polynomial, wavelet, or Fourier expansion, and showed that all square integrable payout functions are embeddable. Our embedding scheme interpolates the pointwise approach, allows extrapolation, and allows rigorous sensitivity analysis. By exploiting state-of-the-art concentration bounds we established sufficient sampling guidelines which, when satisfied, ensure the embeddings are accurate to high probability. We are currently developing Siamese networks which directly learn the embedding and have incorporated matrix completion in order to treat asymmetric games. This work has supported two master's students, and is a central theme in a 5 day international workshop I will lead this January.

The embedding is useful since it represents the global strategic structure via a combination of simpler, easily visualized games. Thus, the embedding allows both insightful visualization and strategic feature extraction. This approach allows model reduction, and often produces interpretable, independent features. That work is in review at ICLR [10]. I am currently working with a master's student to improve feature extraction by introducing sign constraints, as in non-negative matrix factorization, and to use a sparsity promoting framework adapted from SPCA.

The embedding scheme described above provides a new latent space representation of payout functions. That representation inspires novel training approaches. Since the embedding scheme separates a payout into simpler games, each game defines an independent objective. We seek a diverse population containing individuals who excel in each independent aspect of the overall game by optimizing with respect to each objective simultaneously. We are currently testing this approach, which promises an novel alternative to gradient based optimization, self-play, genetic algorithms, or replicator based methods. We will compare our results to the rectified Nash approach introduced by Google Deepmind in [11].

Research Program: I will continue to work on problems which depend on the interaction of system dynamics and system structure. I aim to generalize tools from simplicial homology to allow more general graph complexes, to investigate applications of those tools in neuroscience, to improve strategic feature extraction, and to develop novel training paradigms which select for diverse agents.

Sincerely,

Alexander Strang

Alexander Strang
William H. Kruskal Instructor