Personal, Relevant Background, and Future Goals Statement

I am motivated by a variety of issues from health to sustainable agriculture to stopping human trafficking. I found that I wanted to work on foundational aspects that could impact a wide variety of issues rather than focusing on an isolated topic. Through my education and work experience, I saw the power of mathematics to make sense of and find structure in complex data. Mathematics is the key to solving more fundamental problems in society. I expect to create my biggest impacts by developing mathematical tools that will enable others to be more productive and effective in their work.

Undergraduate Research

Following my interest in developing solutions for real problems, I immersed myself in research. I learned to apply a wide spectrum of mathematics which in turn motivated me to excel in upper division and mathematics courses. My motivation to do research and contribute to society drove me to learn and love mathematics, improving my GPA.

My first experience with research was in 2013 when I participated in the Department of Energy Summer Undergraduate Laboratory Internship at Idaho National Laboratories. There I worked on the Nuclear Risk team to modify the underlying binary decision diagram to calculate risk of failure during tree reductions. I learned about graph algorithms, and I experienced firsthand how advanced mathematics is used in software critical to our nation's safety.

During my undergraduate work at University of Hawai'i at Manoa (UHM), I worked with Dr. Brian Bingham in the Field Robotics Laboratory. My first project was developing a real-time visualization tool for monitoring deployed robots. I learned about data visualization and how to successfully write and deploy software.

Once this visualization tool was deployed, I began to work on an oil spill simulation to test robotic control algorithms. After extensive literature review, I found a set of models which could replicate the fine-grained dynamics of oil spills much better than those currently used in the robotics literature. I worked on this project over the course of a year. In that time, I worked closely with mechanical engineers and computer scientists, learning the basics of effective teamwork. I learned that academic research is a long-term process, and I contributed actively in co-authoring peer-reviewed publications.

While collaborating with Dr. Bingham, I realized that it was the underlying mathematics that was crucial to the success of the project. I wanted to experience researching mathematics further. I next worked on a project with Dr. Yuriy Mileyko for one semester. We applied Topological Data Analysis (TDA) and statistical geometry to quantify minute changes in medical images. This project introduced me to the power of TDA, an emerging field of mathematics with foundations in algebraic topology.

During this series of projects and two years of actively participating in research, I learned how advanced mathematics could be used to impact other fields. I gained experience with diverse domains, collaborating between fields and developing and deploying mathematical software. Most importantly, I experienced firsthand the power of applying topology to data, motivating me to choose this field of research as my focus in graduate school and the focus of my research proposal. I realized it has immense power, limitless potential, and many unexplored open questions.

Industry Experience

After completing my undergraduate degree, I wanted to gain firsthand experience in how mathematics is used in industry. Over two years, I worked as an analytics consultant and as a software test engineer. These two experiences gave me insight into the difficulties of implementing data analytics in industry and an understanding of how usability affects the adoption of ideas.

As an analytics consultant, I worked closely with big banks and retail firms to develop many of their modeling endeavors. I saw how advanced techniques were passed by in practice because, even though the results would be better, they were harder to use. I learned how difficult the adoption of ideas can be and how critical interpretability is for instigating enterprise-level change.

As a test engineer, I became skilled at decomposing complex systems into simple digestible components and became interested in usability issues. I enjoyed developing tests for radio-communication systems and signal decoders, which I found was very similar to constructing mathematical proofs. As I learned about information theory and signal processing, I gained an appreciation for ideas and importance of usability.

My experience in industry motivated me to return to graduate school. I wanted to learn the theory and tools from pure mathematics for application in industry. This idea is fundamental to my current research goals, and my experience in industry strengthens my foundation to take on this role.

Community Outreach

I have also served as leadership in roles broadening the understanding and use of mathematics to solve real-world problems. My first major experience in community organizing was as the President of Mathematics Club at UHM where I **organized a lecture series**. Once a month, we brought researchers from different departments and industry professionals to talk about how they used mathematics in their work. During this past summer, I was **team lead** during the Code for Good conference, where I organized and mentored a group of developers of all skill ranges to build software for nonprofits. I organized the team to build a strong community and a rich collaborative environment. Some of these mentee relationships are ongoing, and I have been asked to return next year in an organizing role.

To make tools from TDA more accessible to data scientists, I am an active contributor to open source TDA and machine learning libraries. I am a **contributing developer** to KeplerMapper, a popular open source implementation of mapper (a method from TDA whose aspects I want to study further during my graduate work). Also, I am a contributor to scikit-learn, the de facto Python machine learning library.

To increase the impact of my work, I have been developing my skills in public speaking and expository writing. I routinely participated and presented my undergraduate research at the UHM research symposiums, and more recently gave lectures on TDA at the WSU mathematics seminar. I plan to adapt these lectures for the data scientist community at large and present them at the local data science meetups. Additionally, I have written a series of blog posts about TDA with the intent of framing the tools in an accessible and interesting way for industry data scientists.

Intellectual Merit

I have already gained experience participating in key aspects of research endeavors. At UHM, after developing simulation and control algorithms, we successfully deployed the oil spill tracking robots. The research resulted in **two publications** (one forthcoming). I wrote **successful grant proposals** for both the oil spill modeling and MRI shape analysis project (both funded by the Undergraduate Research Opportunities Program). I **presented my research** at the UHM undergraduate research symposium.

Starting in Fall 2017, I began pursuing a PhD at WSU. I have already **passed my qualifying exams** ahead of schedule, while most graduate students take the exams after their first year. This allows me to immediately start focusing on research.

Additionally, I have already identified my mentors and advisors who are well qualified to train me in the field I am interested in. During my graduate degree, I will work with Dr. Bala Krishnamoorthy and Dr. Bei Wang. Both are established researchers in TDA, have extensive networks of collaborators throughout industry and academia, and have experience working on the specific aspects of my proposed research plan.

Broader impacts

While participating in research, I always made sure that I understood its purpose and that my work was contributing to society, be it helping to clean up oil spills or understanding the effects of athletic tape. Over time, I have come to realize that to make a bigger impact, I should not focus on what I can do, but how I can **enable others**. By helping others and working on the **underlying tools** they use, I can create a multiplier effect for my work that provides orders of magnitude larger impact than if I tried to solve every problem myself.

I plan to work at the **nexus between industry and academia**. Continually bringing high-powered tools from pure mathematics to the problems most prevalent to today's society. The potential for impact is massive as each useful tool created or made simpler can enable many efforts to be more successful and effective. Given my experience working in industry for two years, I am comfortable reaching out to industry workers. I will leverage this experience to communicate deep mathematics in a way that is relatable to industry professionals.

Conclusion

I was originally drawn to scientific research with the intent to understand society and to facilitate progress in solving big issues. As such, I approach mathematics within the context of understanding and addressing important problems. I am interested in applications that affect people and how people use the tools. I am fascinated by Topological Data Analysis because of its rich foundation and near limitless potential for useful applications. While this is interesting from a pure mathematics perspective, the necessity of usability and accessibility for these methods is critically important so that the tools will reach an audience broader than their creators.

I want to pursue a career in mathematics while always keeping an eye towards where it is applied. The NSF graduate research fellowship will provide me much needed support so that I can develop my research to the point where I can make a large impact. A fellowship will enable me to achieve my research goals by alleviating concerns for funding and the necessity of work outside of my research program.

Graduate Research Plan Statement

Spurred by modern communications systems, cheap sensors, and expanding digital storage, the amount of data collected is growing at exponential rates in both complexity and size. All of this data is only as useful as the insights we can discover from it. While conventional methods can extract lots of actionable information, much untapped potential for understanding and breakthroughs still remain.

Mapper is one of the main techniques from the nascent field of Topological Data Analysis (TDA). It is a form of unsupervised learning that estimates and preserves the homotopy of the underlying space. Because of this perspective, it can reveal new types of insights that are not identified by conventional data analytics techniques [6].

Mapper is a visually rich platform for exploratory data analysis that provides new insights in the form of clusters, loops, voids, and flares. Clearly, noticing that the data separates into clusters is interesting. Finding that the data lies along a long path or a loop can be equally revealing. As an example of its power, researchers discovered patterns in malaria recovery data in the form of a loop and a flare of the mapper construction [7]. Understanding these patterns led to insights for treating children with malaria, but the discovery was made in an ad hoc fashion.

While shown to be useful in a number of cases, in general mapper is very new and its characteristics are not completely understood. While extremely promising, mapper is unwieldy and discovering new insights requires extensive manual tuning and time intensive visual inspection. I want to tackle these shortcomings so that mapper and other TDA methods are more accessible and will be more useful and beneficial for society.

Automatic information extraction: For efficient use of these technologies, it is critical to formalize the identification of significant features. I will develop a framework for quantifying flares in the mapper construction. Additionally, I will develop algorithms to find and extract the features automatically. This will greatly facilitate users to leverage mapper for extracting valuable insights in a wide variety of new applications. My collaborators have begun working on an aspect of this pursuit by quantifying interesting paths in mapper [3].

Stability and Categorification: Understanding the stability of mapper is an important research focus from a theoretical perspective as well as for application. In practice, strong stability results will provide a way to automatically tune the mapper and greatly decrease the effort required to construct useful mappers. Dey et al. have extended mapper to a multiscale version which admits stability results under certain assumptions [2]. Munch and Wang (a collaborator) have used sheaves and category theory to show that mapper converges to the better understood Reeb graph [5]. Carrière et al. have characterized the statistical attributes of a 1-dimensional mapper [1].

All of these methods are interesting and promising for addressing the same problem, but none provide a holistic and clear understanding of the stability of mapper. I will explore whether the categorical representation and statistical representation can be reconciled in a way to leverage results from both. Additionally, I will explore whether the multiscale mapper stability can tell us anything about the construction of the normal mapper.

Comparing mappers: While constructing and visually comparing mappers for each year of Congress voting data, Lum et al. found interesting within-party fragmentation and reformation that conventional clustering techniques were unable to find [4]. I will develop algorithms for efficiently comparing mappers so that similar insights can be made without requiring ad hoc visual comparison.

Developing stability results should provide a metric or psuedometric, but until those are well established, I will develop heuristics for comparisons in practice. This will enable comparisons of mappers between data sets and quantification of changes over time, greatly enhancing usability of mapper in dynamic and time-variant settings.

Novel Applications: To show the utility of topological information in a variety of domains, I will explore applications to new domain areas. I will explore applications in which current methods do not use topology but have potential to benefit from this added perspective. Two such potential explications I have identified are incorporating topological information into exploitation vs exploration algorithms such as multi-armed bandits and into natural language processing tasks such as machine translation.

Intellectual merit

Exploring stability of mapper would require developing novel connections between algebraic topology, category theory, and statistical geometry. Developing algorithms for mapper comparison will enable entirely new categories of applications with the potential to unveil new insights in many fields and create breakthroughs.

Besides the questions at hand having intellectual merit, the team and I will be collaborating with are uniquely qualified for working on these problems. I have enough background in the material to begin immediately.

Broader impact

This research will advance our understanding of Topological Data Analysis, make its methods more accessible to non experts, and grow its potential for diverse applications. Advances in our understanding of mapper enable advances in other academic disciplines, industry interests, and in turn, are a direct benefit to society.

Applications of TDA have already expanded our understanding of diseases such as cancer and diabetes. Through my efforts in this research, my advances in these techniques will create a multiplier effect where they have the potential to greatly impact not just our understanding of mathematics, but also advance a multitude of fields in their efforts to address important problems and needs.

[1] M. Carrière et al. "Statistical Analysis and Parameter Selection for Mapper". In: *ArXiv* 1706.00204 (2017).

[2] T. K. Dey et al. "Mutiscale Mapper: A Framework for Topological Summarization of Data and Maps". In: *ArXiv* 1504.03763 (2015).

[3] M. Kamruzzaman et al. "Toward A Scalable Exploratory Framework for Complex High-Dimensional Phenomics Data". In: *ArXiv* 1707.04362 (2017).

[4] P. Y. Lum et al. "Extracting insights from the shape of complex data using topology". In: *Scientific Reports* 3 (2013), p. 1236.

[5] E. Munch et al. "Convergence between Categorical Representations of Reeb Space and Mapper". In: *ArXiv* 1512.04108 (2015).

[6] G. Singh et al. "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition". In: *Eurographics Symposium on Point-Based Graphs* (2007).

[7] B. Y. Torres et al. "Tracking Resilience to Infections by Mapping Disease Space". In: *PLOS Biology* 14.4 (2016), pp. 1–19.