Spectral methods for high-dimensional tensor data
Miaoyan Wang, December 2018

Higher-order tensors have recently received increasing attention in many fields such as neuroscience, social networks, computer vision, and genomics. One important reason for such appreciation is the effective representation of multiway data using tensor structure. Other examples include high-dimensional applications of the method-of-moment, in which higher-order statistics can naturally be represented by higher-order tensors. Methods built on tensors provide powerful tools to capture complex structures in data that lower-order methods may fail to exploit. However, tensor-based methods are fraught with challenges. Tensors are not simply matrices with more indices; rather, they are mathematical objects possessing multilinear algebraic properties. The statistical properties of tensor-based methods have yet to be studied.

The proposed research is motivated from PI's established collaborations, one on human whole-genome transcriptom project, and the other on neuroimaging studies. In these two applications and many others, scientists are interested in identifying a subset of features (genes or brain voxels) that behave similarly in subset of samples (individuals) across certain contexts (tissues or disease status). In a broader sense, it is of great importance to search for meaningful low-dimensional structure within the high-dimensional tensor data. **The PI will develop a framework–of statistical models, spectral algorithms, and relevant theory–to analyze high-dimensional tensor data.** The research will bridge the gap between practice and theory for tensor data analysis. The previous literature has advocated unfolding the tensor into a matrix and applying classical spectral methods developed for matrices. Despite the popularity of such techniques, the impact of unfolding on estimation properties is currently not well understood. Our approach goes beyond the traditional multivariate analysis: we will focus on the distribution over "objects" while the objects can be images, networks, manifolds, and tensors. **This will allow researchers to examine complex interactions among tensor entries and between multiple tensors, thereby providing solutions to questions that cannot be addressed by traditional matrix analysis.**

Intellectual Merit: The project will support the development of the PI, a young faculty, to create an inclusive and interdisciplinary working group of researchers. The PI has a unique combination of post-graduate training in statistics (2010-2015), mathematics (2016-2017), computer science (2017-2018), and genomics (2013-2018). The proposed research builds on the PI's previous high-impact work on tensor applications, published in the PNAS and the Annals of Applied Statistics [Wang et al. 2018a, Wang et al. 2018b], and theoretical work on tensor spectral properties, published in the Linear Algebra and its Applications, and Journal of Machine Learning Research W&C [Wang and Song, 2017, Wang et al. 2017]. The PI's previous and ongoing projects haven made continuous contribution to connections across diverse disciplines, and the PI strives to push the boundary of interdisciplinary research further.

Broader Impact: As one of few female faculty in her home department, PI has been striving to encourage more under-represented students into her research field. The educational components of this proposal are targeted at several different populations of students: PhD students, undergraduate students, visiting international students, AP statistics students underrepresented students at West High School in Madison. The PI plans to develop an undergraduate course on *introduction to data sciences*. This course will introduce undergraduate students to a wide range of modern exploratory analytic tools for massive data sets, and will be closely tied with data-intensive scientific domains. "Data are not just numbers, they are numbers with context." The PI will also apply effective pedagogical tools to foster the mindset of statistical thinking to students. This course will also include outreach to an AP stat class and under-represented students in Madison.