Abstract: The explosion of available data and the revolution in computing technologies have created a critical need for both compressed representations of large, real-world data and powerful data-driven algorithms. In this talk, we will address these needs in two distinct ways: by obtaining optimal multidimensional approximations, and by designing efficient deep learning algorithms. The traditional approach to dimensionality reduction and feature extraction is the matrix singular value decomposition (SVD), which presupposes that data have been arranged in matrix format. In the first half of this talk, we will show that high-dimensional datasets are more compressible when treated as tensors (multiway arrays). We will obtain these representations using a tensor algebra under which notions of rank and the tensor SVD are consistent with their matrix counterparts. This framework yields provably optimal approximations, and we will support this theory with empirical studies. Deep neural networks (DNNs), flexible models composed of simple layers parameterized by weights, have been successful high-dimensional function approximators in countless applications. However, training DNNs (i.e., finding a good set of weights) is notoriously challenging, requiring significant time and computational resources. In the second half of this talk, we will describe two approaches for training separable DNNs, the commonly-used architecture where the weights of the final layer are applied linearly. We will leverage this linearity using partial optimization in a deterministic setting and iterative sampling in a stochastic setting. We will demonstrate empirically that both approaches yield faster convergence to more accurate DNN models and less tuning of hyperparameters. We will conclude with a discussion about new ideas to bring these two powerful data-based techniques together.

Tuesday, January 18, 2022, 10:00 am
Mathematics and Science Center: MSC W201