ASCR PI Meeting White Paper

Tensor-based Data-driven Method for Scientific Computing

Xiu Yang, PNNL

***Targeting question*:**

2. *Convergence of data- and model-driven discovery* and part of 4. *Applied mathematics for future computing directions*.

Research Goals and Impact:

1. *Data Representation Based on Tensor Decomposition*. Tensorization, or the reshaping of a huge data set into tensors, is a powerful technique in scientific computing. Once a data set has been tensorized, tensor decomposition methods, such as the *tensor-train* [1] or *high-order singular value decomposition* (HOSVD) [2], can be used to form extremely compact representations, much smaller than the original data size. By limiting the rank in a tensor decomposition, it also is possible to find an accurate approximation of the original data, which, again, reduces the size of the representation. In many U.S. Department of Energy (DOE) projects, extracting essential information from a huge data set has the potential to enhance new scientific discovery or improve physical models. Big data analysis problems exist in many DOE projects, including high energy physics, inertial confinement fusion implosion, and wind and solar energy system, to name a few. The collected data sets in these projects are extremely huge. Therefore, a concise representation of the data has the potential to dramatically reduce the storage and computational costs of using learning methods (e.g., machine learning or deep learning). New techniques, beyond existing ones, can be added to this category to make it much more powerful. For example, dictionary learning, which is another approach to compact representation, seeks a set of basis functions, and the data are identified with sparse weighted sums of this dictionary basis. By combining and scaling all three of these techniques—tensorization, low-rank, and sparsity—it will be possible to develop a method for dictionary learning that discovers sparse and low-rank decompositions of tensorized data.
2. *Scientific Discovery Based on Tensor Decomposition*. Similar to the singular value decomposition (SVD), principle component analysis (PCA), or proper orthogonal decomposition (POD), tensor decomposition methods not only provide a concise representation of the data but also help to discover critical properties of the physical systems (e.g., POD modes corresponding to large eigenvalues are important "dynamical modes"). The first data analysis applications of tensors saw the development of the Tucker and canonical decomposition (CANDECOMP) in chemometrics and the parallel factor model (PARAFAC) in linguistics [3]. Recently, tensor methods have been applied in signal/image processing and machine learning [3]. Dictionary learning is a relatively recent method in data analysis related to frame theory [4]. These ideas can be naturally extended to modeling of complex dynamical system, for example, to improve scalability of dynamic mode decomposition (DMD) through tensor decomposition and low-rank approximation or to find new machine learning mechanisms for learning the nonlinear Koopman operator. These

useful data-driven methods also can accelerate theoretical development because they provide "numerical understanding" of the physics. Moreover, the tensor decomposition method (e.g., tensor trains have been used in high-dimensional hierarchical uncertainty quantification (UQ) problems for chip design [5] and quantum chemistry [6]) can augment studies based on the large size of *high-dimensional* data (the UQ problem) or *ultra-high-dimensional* representation of the system (quantum chemistry). New development of tensor decomposition has the distinct potential to enhance different scientific research areas.

3. *Tensor for Future Computing*. The tensor based techniques have been widely used in data sciences and powerful tools based on tensor (e.g., Google's *tensorflow*) have become mainstream software for data analysis. Recently, Google released a new processing unit – Tensor Processing Unit (TPU) – designed specifically for machine learning based on tensors. In the past decade, we have witnessed the tremendous enhancement of the capability supercomputers with the development of the graphics processing unit (GPU). Similarly, it is very promising that integrating the TPU in future supercomputers will dramatically improve the efficiency of data-driven or data-based scientific discovery. Therefore, the development of efficient tensor-oriented or tensor-based scientific computing algorithms are needed for the era of supercomputing beyond the scaling limits of Moore's law.

References

[1] Oseledets, I. V. (2011). Tensor-train decomposition. *SIAM Journal on Scientific Computing*, *33*(5), 2295-2317.

[2] De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, *21*(4), 1253-1278.

[3] Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., & Phan, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, *32*(2), 145-163.

[4] Mairal, J., Bach, F., & Ponce, J. (2014). Sparse modeling for image and vision processing. *Foundations and Trends® in Computer Graphics and Vision*, *8*(2-3), 85-283.

[5] Zhang, Z., Yang, X., Oseledets, I. V., Karniadakis, G. E., & Daniel, L. (2015). Enabling high-dimensional hierarchical uncertainty quantification by ANOVA and tensor-train decomposition. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *34*(1), 63-76.

[6] Khoromskaia, V., & Khoromskij, B. N. (2015). Tensor numerical methods in quantum chemistry: from Hartree–Fock to excitation energies. *Physical Chemistry Chemical Physics*, *17*(47), 31491-31509.