
Error Controlled Compression of Scientific Data with Low-Rank Approximations

George Turkiyyah*¹

¹KAUST – Saudi Arabia

Abstract

The truncated singular value decomposition and its various tensor generalizations have long offered a simple and practical mechanism for compressing data stored in 2D or higher-order tensors. While the resulting low-rank approximations provide strong guarantees on the quality of the resulting reconstruction, they do not always provide sufficient compression as needed by the increasingly large scientific data volumes being produced in large-scale spatio-temporal simulations. In this work, we introduce a second-stage compression where the singular vectors themselves are further compressed, adding a multiplicative effect on the overall compression ratios achieved. The intuition of this second-stage compression is that singular vectors associated with smaller singular values can be compressed using lower accuracies without compromising the overall accuracy targeted in the compression. Specifically, we show that if the singular vectors are compressed with an accuracy inversely proportional to their singular values, the overall accuracy reached can be readily controlled. We show the effectiveness of the resulting algorithm on sample representative datasets from climate, weather, and computational fluid dynamics, where the proposed algorithm achieves higher compression ratios than other state-of-the-art algorithms for the same target root mean square accuracies. The strength of the method appears to lie in its combination of global feature-level SVD-based compression in the first stage and local entry-level interpolation-guided compression in the second stage, leveraging the complementary strengths of both approaches in a unified pipeline.

*Speaker