Significantly Improving Lossy Compression for Scientific Data Sets Based on Multidimensional Prediction and Error-Controlled Quantization

Today's scientific simulations require a significant reduction of the data size because of extremely large volumes of data they produce and the limitation of storage bandwidth and space. If the compression is set to reach a high compression ratio, however, the reconstructed data are often distorted too much to tolerate. In this paper, authors explore a new compression strategy that can effectively control the data distortion when significantly reducing the data size. The contribution is threefold. Authors propose an adaptive compression framework to select either our improved Lorenzo prediction method or our optimized linear regression method dynamically in different regions of the dataset. This paper studies how to select them accurately based on the data features in each block to obtain the best compression quality. (3) Authors analyze the effectiveness of our solution in details using four real-world scientific datasets with 100+ fields. Evaluation results confirm that our new adaptive solution can significantly improve the rate distortion for the lossy compression with fairly high compression ratios. The compression ratio of our compressor is 1.5X~8X as high as that of two other leading lossy compressors (SZ and ZFP) with the same PSNR, in the high-compression cases. Parallel experiments with 8,192 cores and 24 TB of data shows that our solution obtains 1.5X dumping performance and 1.72X loading performance compared with the second-best lossy compressor, respectively

Reference:

Significantly Improving Lossy Compression for Scientific Data Sets Based on Multidimensional Prediction and Error-Controlled Quantization