Project Proposal Details

Group 40

1. Group Information

Team Members: Jaeho Kim, Daniel David, Ana Vizitiv
Group Number : 40

2. Topic Selection

Title:
Linear Algebraic Approaches to Neuroimaging Data Compression: A Comparative Analysis of Matrix and Tensor Decomposition Methods for High-Dimensional Medical Images

Research Question:
How can matrix and tensor decompositions be applied to optimize image compression by reducing memory usage while preserving image quality?

This study seeks to analyze and compare various linear algebra-based image compression methods, focusing on understanding how techniques like Singular Value Decomposition (SVD), Principal Component Analysis (PCA), and Higher-Order Singular Value Decomposition (HOSVD) can balance data compression efficiency with image fidelity. The research aims to identify the most effective mathematical frameworks for managing high-dimensional data in image processing applications, providing insights into the trade-offs between compression rate and quality retention.

Background and Motivation:

Image compression has become increasingly critical in modern data management, particularly in specialized fields like medical imaging where data quality directly impacts diagnostic accuracy. While traditional compression methods have served general-purpose images well, the unique challenges posed by high-dimensional medical imaging data, particularly in neuroimaging, demand more sophisticated mathematical approaches. Linear algebraic methods, with their ability to capture complex data relationships and reduce dimensionality while preserving essential information, offer promising solutions to these challenges.

While existing compression techniques have made significant strides in managing neuroimaging data, current methods often struggle to maintain an optimal balance between compression efficiency and image fidelity, particularly for multi-dimensional datasets. The application of advanced linear algebraic methods, specifically matrix and tensor decompositions, remains understudied in the context of medical image compression.

Tensors are mathematical objects that generalize scalars (single values), vectors (1D arrays), and matrices (2D arrays) to higher dimensions; they are represented as multi-dimensional arrays and are used to describe relationships between sets of data in spaces of arbitrary dimensions, often in physics, engineering, and machine learning (Hackbusch, 2012). Tensors provide a natural framework for representing multi-dimensional neuroimaging data, offering powerful tools for decomposition and compression (Kolda & Bader, 2009). These methods have shown particular promise in medical imaging applications (Zhou et al., 2013), where they can effectively capture the complex spatial and temporal relationships in neuroimaging data while enabling significant dimensionality reduction (Cichocki et al., 2016).

Neuroimaging techniques, such as Functional Magnetic Resonance Imaging (fMRI), construct data as multi-dimensional images. These are often represented as three-dimensional (3D) volumetric images that capture spatial information about brain structure, or as four-dimensional (4D) datasets that concatenate time-series data to reflect neural activity across multiple volumes over time (Smith, 2004). These datasets are inherently large and complex, presenting significant challenges for storage and management.

The substantial size of fMRI datasets makes it difficult and costly to store multiple sets of data on a single endpoint. This issue is exacerbated in clinical and experimental contexts, where the need to retain imaging data from numerous subjects adds to the logistical and financial burden of managing such extensive datasets (Van Horn & Toga, 2014; Wardlaw et al., 2007). For instance, large-scale studies like the NeuroGrid Stroke Exemplar trial generated approximately 12,000 scans, highlighting the challenges of storing, sharing, and analyzing these datasets across multiple sites (Wardlaw et al., 2007).

Advancements in neuroimaging techniques, such as higher spatial and temporal resolutions, have further increased the dimensionality and complexity of imaging data. Early diffusion MRI (dMRI) studies captured diffusion along six directions to infer neural fiber orientations, whereas modern approaches now resolve over 512 directions, significantly expanding dataset sizes (Van Horn & Toga, 2014; Wardlaw et al., 2007). With neuroimaging dataset sizes doubling approximately every 26 months since 1995 and exceeding 20GB per study by 2015, efficient compression methods are essential to address the growing demands of storage and analysis (Van Horn & Toga, 2014).

As newer forms of human brain imaging generate increasingly high-dimensional data, it becomes necessary to identify suitable compression methods. Both traditional matrix-based approaches and recent tensor-related methods offer potential solutions for managing these datasets efficiently. Surveying these techniques is critical to ensure that the exponential growth of neuroimaging data does not hinder its clinical and experimental applications (Calhoun & Sui, 2016; Dinov, 2016; Poldrack & Farah, 2015).

Calhoun, V. D., & Sui, J. (2016). Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biological psychiatry: cognitive neuroscience and neuroimaging*, 1(3), 230-244.

Dinov, I. D. (2016). Volume and value of big healthcare data. *Journal of medical statistics and informatics*, 4.

Hackbusch, W. (2012). Tensor Operations. In: Tensor Spaces and Numerical Tensor Calculus. *Springer Series in Computational Mathematics*, vol 42. Springer, Berlin, Heidelberg.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3), 455-500.

Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., & Arnfred, S. M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3), 938-947.

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, 526(7573), 371-379.

Smith S. M. (2004). Overview of fMRI analysis. *The British journal of radiology*, 77 Spec No 2, S167–S175.

Van Horn, J. D., & Toga, A. W. (2014). Human neuroimaging as a “Big Data” science. *Brain imaging and behavior*, 8, 323-331.

Wardlaw, J. M., Bath, P., Sandercock, P., Perry, D., Palmer, J., Watson, G., ... & Farrall, A. (2007). The NeuroGrid stroke exemplar clinical trial protocol. *International Journal of Stroke*, 2(1), 63-69.

Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), 540-552.

3. Project Scope and Objectives

a. Analyze how matrix-based compression methods extend to tensor operations in higher dimensions.
b. Compare theoretical compression ratios achievable through linear transformations across dimensions.
c. Evaluate computational complexity of linear algebraic operations in higher dimensions.
d. Develop a mathematical framework for understanding dimensionality's impact on storage efficiency for compressed data, demonstrated through their application to medical images.
e. Compare obtained efficiency insights to the current usage in the field.

4. Approach and Methodology

1. <u>Literature Review:</u>
    - **Matrix-Based Compression Techniques**: Study current image compression algorithms, particularly those based on matrix decompositions such as SVD, QR, and LU Factorization.
        i. Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., & Arnfred, S. M. (2006). Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3), 938-947.
        ii. Zhang, L., & Wu, X. (2006, January). An efficient lossless compression algorithm for fMRI data volume. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (pp. 3093-3096). IEEE.
        iii. Theis, F. J., & Tanaka, T. (2005, September). A fast and efficient method for compressing fMRI data sets. In *International Conference on Artificial Neural Networks* (pp. 769-777). Berlin, Heidelberg: Springer Berlin Heidelberg.
        iv. Cai, W., Feng, D., & Fulton, R. (1998, November). Clinical investigation of a knowledge-based data compression algorithm for dynamic neurologic FDG-PET images. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286) (Vol. 3, pp. 1270-1273). IEEE.

v. Cohen, M. S. (2001). A data compression method for image time series. *Human brain mapping*, 12(1), 20-24.
vi. Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of neuroscience methods*, 264, 47-56.
vii. Liu, S., Bai, W., Zeng, N., & Wang, S. (2019). A fast fractal based compression for MRI images. *IEEE Access*, 7, 62412-62420.

- **Tensor Decomposition Methods**: Investigate methods for decomposing tensors (such as HOSVD) and their applications in multi-dimensional data.

i. Hackbusch, W. (2012). Tensor Operations. In: Tensor Spaces and Numerical Tensor Calculus. *Springer Series in Computational Mathematics*, vol 42. Springer, Berlin, Heidelberg.
ii. Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3), 455-500.
iii. Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), 540-552.
iv. Oseledets, I. V. (2010). Approximation of 2^d\times2^d matrices using tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 31(4), 2130-2145.
v. Zhang, L., Zhang, L., Tao, D., Huang, X., & Du, B. (2015). Compression of hyperspectral remote sensing images by tensor approach. *Neurocomputing*, 147, 358-363.
vi. Tyrtyshnikov, E. E. E. (2003). Tensor approximations of matrices generated by asymptotically smooth functions. *Sbornik: Mathematics*, 194(6), 941.
vii. Wang, H., & Ahuja, N. (2005, June). Rank-R approximation of tensors using image-as-matrix representation. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Vol. 2, pp. 346-353). IEEE.
viii. Gurumoorthy, K. S., Rajwade, A., Banerjee, A., & Rangarajan, A. (2009). A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases. *IEEE Transactions on Image Processing*, 19(2), 322-334.

- **Dimensional Expansion in Linear Spaces**: Review the mathematical foundations and medical applications of scaling matrix-based methods to higher-dimensional spaces using linear transformations.
    i. Ahmadi-Asl, S., Caiafa, C. F., Cichocki, A., Phan, A. H., Tanaka, T., Oseledets, I., & Wang, J. (2021). Cross tensor approximation methods for compression and dimensionality reduction. *IEEE Access*, 9, 150809-150838.
    ii. Chang, S. Y., & Wu, H. C. (2022). Tensor quantization: High-dimensional data compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8), 5566-5580.
    iii. Lalgudi, H. G., Bilgin, A., Marcellin, M. W., Tabesh, A., Nadar, M. S., & Trouard, T. P. (2005, April). Four-dimensional compression of fMRI using JPEG2000. In *Medical Imaging 2005: Image Processing* (Vol. 5747, pp. 1028-1037). SPIE.

2. Mathematical Analysis:
    - **Matrix Decomposition in Higher Dimensions**: Explore how matrix decompositions like SVD and PCA can be adapted for tensors, and analyze their effectiveness in dimensional scaling.
    - **Tensor Analysis**: Examine how tensors represent multi-dimensional data and develop the mathematical tools needed to extend linear transformations to higher dimensions.
    - **Linear Transformations Across Dimensions**: Investigate how transformations (such as rotations, scaling, and translations) behave as they are extended from 2D to 3D and beyond.

Mathematical Methods:

The core of our methodology relies on linear algebra and tensor operations, including the following:

1. **Matrix Decompositions**:
    - **Singular Value Decomposition (SVD)**: Used to represent matrices as the product of three matrices, providing insights into data compression.
    - **Principal Component Analysis (PCA)**: A dimensionality reduction technique that identifies the directions (principal components) of maximum variance in the data.
    - **Eigenspace Representations**: Decomposing matrices using eigenvalues and eigenvectors, which is crucial for many dimensionality reduction techniques.

2. **Tensor Operations**:
    - **Tensor Decomposition** (HOSVD): Generalizes matrix SVD to higher-dimensional tensors, enabling the compression and analysis of multi-dimensional data.
    - **Multi-linear Transformations**: Extend linear transformations to tensors, using concepts like Kronecker products and tensor reshaping.
    - **Kronecker Products**: Generalize matrix multiplication to higher-dimensional data, allowing tensor decompositions to be performed more efficiently.

3. **Supporting Concepts**:
    - **Vector Space Theory**: Essential for understanding how data can be represented in multi-dimensional spaces and how tensor decompositions operate within these spaces.
    - **Linear Operator Theory**: Provides the framework for understanding linear transformations on vectors, matrices, and tensors.
    - **Matrix Approximation**: Techniques like low-rank approximation are key for reducing data size without losing critical information, particularly when scaling up to higher dimensions.

5. Resources and Tools

<u>Data Sources</u>

1. **Open Scientific Data**: The medical images that we will use to demonstrate the storage efficiency of compressed data will be sourced from open data sources including, but not limited to, the below:
    - Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., ... & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *Elife*, 10, e71774.
    - Royer, J., Rodríguez-Cruces, R., Tavakol, S., Larivière, S., Herholz, P., Li, Q., ... & Bernhardt, B. C. (2022). An open MRI dataset for multiscale neuroscience. *Scientific data*, 9(1), 569.

<u>Data Types</u>

1. **2D Data:** Images can be represented as matrices (grayscale or color) where each element corresponds to pixel values.
   - Matrix representations of images
   - 2D linear transformations

2. **3D Data:** Tensors are used to represent image volumes, for instance, videos or multi-spectral images, where the third dimension represents depth or time.
   - Tensor representations
   - 3D linear operators

<u>Tools and Software</u>

1. **Primary Tools**
   - Python with NumPy for matrix operations
   - SciPy for linear algebra computations
   - Jupyter Notebooks for mathematical analysis

6. Project Timeline

Week 1 (Nov. 11 - 17)

- Literature review of linear algebra in compression
- Study of matrix decomposition methods

Week 2 (Nov. 18 - 24)

- Analysis of 2D compression using matrix methods
- Extension of matrix methods to tensor operations

Week 3 (Nov. 25 - Dec. 1)

- Mathematical analysis of dimensional scaling
- Development of theoretical framework

Week 4 (Dec. 2 - 8)

- Documentation of mathematical proofs
- Final paper writing and revision

7. Potential Challenges and Solutions

Challenges

1. **Mathematical Complexity**
    - Complex tensor calculations
    - Higher-dimensional linear algebra concepts
    - Proof development for dimensional scaling
2. **Theoretical Framework**
    - Bridging the gap between matrix and tensor operations
    - Establishing clear relationships between dimensions

Solutions

1. **Mathematical Approach**
    - Simplify Models
        - Break down tensor operations into manageable matrix operations
        - Use systematic dimensional analysis
        - Develop clear mathematical notation for cross-dimensional analysis
    - Utilize Computational Tools (such as Tensorflow, Pytorch, [Columbia University HPC](), etc.)

2. **Analysis Strategy**
    - Focus on fundamental linear algebraic properties
    - Establish clear mathematical connections between dimensions
    - Use progressive complexity building from 2D to higher dimensions

SIGNATURES: Ana Vizitiv, Daniel David, Jaeho Kim