CS 6962 Decomposition Techniques for Data and Computational Science
Fall Semester 2023

Instructor: Chris Johnson, Ph.D.

Time: M,W 1:25 - 2:45 p.m.

Place: WEB 2760

Office: 4692 WEB

Office Hour: Mondays between 3:00 - 4:00 p.m. and by appointment

Phone: 801-587-7875 (Deb Zemek, Assistant)

Email: crj@sci.utah.edu

Class web page: `my.eng.utah.edu/~cs6962`

Co-Instructor: Timbwaoga Aime Judicael Ouermi (TAJO), Ph.D.

Office Hour: Wednesdays between 3:00 - 4:00 p.m. in WEB 2807

Email: touermi@sci.utah.edu

Course Overview:

"We are drowning in data and starving for knowledge." - John Naisbitt

Researchers in a variety of fields collect measurements, observe data, perform simulations and use a wide range of techniques to describe, classify, analyze and draw conclusions from these data. Selecting appropriate techniques and understanding their advantages and disadvantages is an important component of data analysis. In particular, in this age of big data, large data sets provide distinct challenges given that many of the general techniques for small data sets do not scale to larger problems and are often prohibitively expensive from a computational perspective.

In this class, we will survey several data decomposition techniques for data and computational science applications including:

- Principle Component Analysis
- Independent Component Analysis
- Singular Value Decomposition

- Generalized Singular Value Decomposition
- Tensor Factorization
- Signal Fraction Analysis
- Non-Negative Matrix Factorization
- Probabilistic Matrix Factorization
- Low Rank Approximation Methods

and evaluate their strengths and weaknesses for a variety of applications.

Prerequisite: Linear Algebra

Course Goals: Upon completion of this course, the student should:

- Know a wide variety of algorithms and techniques for data decomposition.
- Have seen and discussed examples of the use of data decomposition techniques applied to problems in a variety of fields.
- Know where to locate data decomposition software resources and references.
- Have completed a data decomposition project.

Assignments: There are two main types of assingments for this course. One is in the form of in between class homework that will primarily consist of applying the techniques learned in class to analyze data sets. The second will be in the form of a data decomposition project. This project can involve data you have collected from a simulation and/or experiment or development of a new software tool(s).

Late Assignments: Assignments submitted late will receive a ZERO. Every student is allocated one (1) ‘late pass’, which they may use on any assignment. A late pass gives the student one (1) extra week to turn in the assignment without penalty. Other exceptions to the late policy will only be made on a case-by-case basis for legitimate cause (unexpected visits to the hospital, etc.). Evidence of the cause is required (i.e. doctors note).

Languages: For this course we will primarily use Matlab and Python along with additional open source software tools.

Grades: Final course grades will be computed according to 70% Homework and 30% Final Project.

Incompletes: As the project is due by the end of the semester, in past similar project-based courses, it has turned out that some people do not wisely schedule their time and do not finish their projects. They then want to take an incomplete and finish the project sometime in the Spring. I only give incompletes very rarely and only for truely unusual circumstances (death in the family, etc.), so **please** work to finish your final project on time.

Schedule for CS 6962

Week 1. Class mechanics, motivation and overview of application problems

Week 2. Linear algebra background

Week 3. Singular value decomposition and principal component analysis

Week 4. SVD, PCA, and Robust PCA

Week 5. Independent component analysis and generalized SVD

Week 6. Guest lectures

Week 7. Nonnegative matrix factorization

Week 8. Fall break

Week 9. Low rank approximations

Week 10. Guest lectures

Week 11. Probablistic matrix factorization

Week 12. Tensor decompositions

Week 13. Case studies

Week 14. Project presentations

Week 15. Project presentations

Note: During the semester, we will have multiple guest lectures on selected topics in decomposition methods.

Decomposition Techniques for Data and Computational Science Project

The decomposition project can be (1) from simulation and/or experimental data you have or (2) development of new software tools or extending existing software tools.

It is your responsibility to pitch your project at the appropriate level. Challenge, but do not exhaust, yourself. Please ensure that even if you underestimate the difficulty of your project, you will have something to hand in by the due date (choosing too difficult a project is not a valid reason for an incomplete).

Due dates:

Project description **due October 4**.

Project progress report **due October 30**.

Project presentations will be on **November 29**, **December 4**, and **December 6**.

Final project write up is **due December 13**.

On **October 4** your project design report is due. This should be a well thought out, well-written 2-3 page description of your proposed project. It should outline any necessary background, specifically what goals you plan on accomplishing, and what you will need to do in order to accomplish your goals. You will also need to include what software/hardware you plan to use, and what you intend to hand in (i.e. what are the "deliverables"). See below for details about the project design report.

On **October 30** your project progress report is due. This report should contain a description of how much of your project is completed at this point and what still remains to be done. This is the time to make modifications (which you must justify) and present a timeline for completion of the project. See below for details about the project progress report.

You will present your final projects on Wednesday, **November 29**, Monday, **December 4** or Wednesday, **December 6**. A final project sign-up sheet will be handed out in class for you to schedule a time. Presentations should typically take approximately 15-20 minutes. Your final project report is due on **December 13**.

Project Design Report

Please hand in your Project Design Report by **October 4** (or sooner). It should contain the following information.

Student Name:
Project title:

- Give an overview of the project.
- Why is this project important and/or interesting?
- What are the objectives of the project? What are the questions you want to answer?
- What would you like to learn by completing this project?
- What data will you be using for your project?
- List the hardware and software you will be using.
- What is your project schedule? What have you done thus far and what will you have to do to complete this project? Be as specific as possible.
- When the project is completed, how *specifically* can we evaluate how successful it is?

Project Progress Report

Please hand in your Project Progress Report by **October 30** or sooner. It should contain the following information.

Student Name:

Project title:

- Estimate the percentage of the overall project you have completed thus far.
- What have you completed?
- Create a list of what still needs to be done on the project and estimate the effort each item will take to complete.
- Have you had to make any changes in your project description? If so, please list and justify the changes.
- Any additional information?

Project Final Report

You will be required to hand in your Project Final Report on **December 13**. Your final report should contain the following information.

Student Name(s):
Project title:

- Provide a brief description of your project and how to run it if it is not self-explanatory.
- Outline what you learned from doing this project.
- If you have not accomplished all the goals of your project, or if you have exceeded them, describe how the finished project differs form the description in your project design.
- Evaluate your project: how successful do you think it was? What are the strengths and weaknesses of your project?
- Provide additional comments useful in evaluating your project.

References

1. R. Baraniuk, "Compressive sensing," IEEE Signal Processing Magazine, pp. 118-–124, July 2007.

2. S. Brunton and J.N. Kutz. Data-Driven Science and Engineering, 2nd edition, Cambridge University Press, 2022.

3. E. Candés, L1-magic toolbox , `http://www-stat.stanford.edu/~candes/l1magic`.

4. A. Cichocki, N. Lee, I.V. Oseledets, A-H. Phan, Q. Zhao, D. Mandic. Low-Rank Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Problems: Perspectives and Challenges PART 1. arXiv:1609.00893, `https://arxiv.org/abs/1609.00893`, 2016.

5. Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, Shun-ichi Amari. Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation, Wiley, 2009.

6. J. Demmel. Applied Numerical Linear Algebra, SIAM Press, 1997.

7. D. Donoho, "Compressed sensing," IEEE Transactions on Information Theory 52 , 1289—1306 (2006).

8. L. Elden. Matrix Methods in Data Mining and Pattern Recognition, SIAM Press, 2007.

9. F. Emdad. High Dimensional Data Analysis: Overview, Analysis, and Applications, VDM Verlag, 2008.

10. G. Golub and C. Van Loan. Matrix Computations. Johns Hopkins University Press, 2012.

11. by Trevor Hastie, Robert Tibshirani, Martin Wainwright. Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press, 2015.

12. A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," `http://mlsp.cs.cmu.edu/courses/fall2012/lectures/ICA_Hyvarinen.pdf`.

13. M. Kirby and L. Sirovich, "Application of the Karhunen–Loeve procedure for the characterization of human faces," IEEE Transactions on Pattern Analysis and Machine Intelligence 12 , 103–108 (1990).

14. J. Nathan Kutz, Data-Driven Modeling and Scientific Computation: Methods for Complex Systems and Big Data", Oxford University Press, 2013.

15. by J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor. Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems, SIAM Press, 2016.

16. J. Shlens, "A tutorial on principal component analysis," `www.sci.utah.edu/~chris/PCA-Tutorial-Shlens-2014.pdf`.

17. D. Skillicorn. Understanding Complex Datasets: Data Mining with Matrix Decompositions, Chapman and Hall/CRC, 2007.

18. J. Stone. Independent Component Analysis. MIT Press, 2004.

19. L. N. Trefethen and D. Bau, Numerical Linear Algebra, SIAM Press, 1997.

20. M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience 3 , 71–86 (1991).

21. J. Wright and Y. Ma. High-Dimensional Data Analysis with Low-Dimensional Models. Cambridge University Press, 2022.

22. X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," `www.optimization-online.org`, 2009.