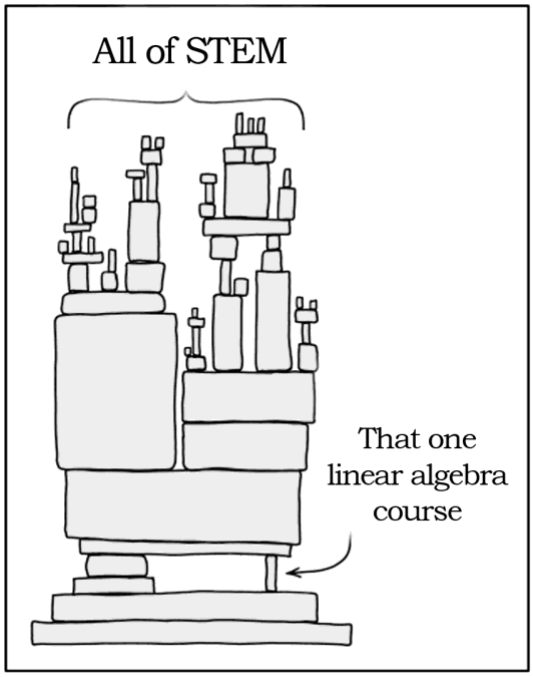
Applied Linear Algebra in Data Analysis



Source: https://twitter.com/Quasilocal/status/1664701458200121351

The above figure is not far from the truth: **linear algebra is just as important as calculus today—if not more so.**

Linear algebra serves as the first step toward high-dimensional thinking, a **prerequisite for modern data analysis**. Paired with probability theory and basic linear statistical modelling, it provides a solid foundation for tackling a wide range of real-world problems and for advancing to more sophisticated material.

In this course, our mission is to lay a solid foundation in:

- **Applied Linear Algebra:** Enabling high-dimensional thinking and the visualization of complex structures.
- **Linear Regression Models:** Arguably the starting point for most statistical modelling and an essential building block for understanding relationships in data.

We will do this through a detailed, systematic study of linear equations in the course. Many fundamental concepts in linear algebra will be developed as natural consequences of the process of understanding and solving linear equations.

Throughout this course, we will:

- Introduce fundamental concepts in linear algebra, matrix methods, probability theory, and linear regression.
- Demonstrate the practical application of these concepts through real-world scenarios in medicine and biology.
- Provide hands-on experience applying these concepts to solve problems.

Course Content

(Modules in orange cover concepts and Modules in blue are applications)

1. **Linear Systems and Matrix Operations** — foundations for modeling, understanding and solving systems of simultaneous linear equations; Matrices; Orthogonality; Matrix inverses; Signal processing: Fourier and Wavelet Transforms.
2. **Least Squares Problems** — overdetermined and underdetermined, data fitting (L_2 optimization); Signal processing: Optimal control.
3. **Linear Programming** — a special and important class of constrained optimization; Applications of linear programs in medicine.
4. **Matrix Factorization Techniques** (LU, QR, Eigen decomposition, SVD) — practical computation of both least squares and LP; Linear Dynamical Systems; Dimensionality Reduction: Principal Component Analysis; Image compression.
5. **Linear Regression and Data Applications** —modeling and inference using linear models; Applications of linear regression models; Machine learning: Cross-validation.

Assignments

- This is a hands-on course heavy on pen-and-paper and programming assignments. Students are encouraged to use Python or R for their programming assignments.
- All assignments will need to be submitted online.
- There will be several assignments in this course, and the best 5 will be used for your final grade.
- All assignments will be released after the tutorial session, and you will have a week to submit your assignments.
- You have a cumulative 5-day late submission grace period for your assignments for the entire course. If you miss the 6 PM deadline for an assignment, a submission made before 6 PM the following day will be considered 1-day late submission.
- If you use up your 5-day grace period, all subsequent late assignments will receive 0 marks.

Resources

The topics covered in this course are quite mature, so it's no surprise that there are numerous wonderful resources. Here are some resources that cover the topics discussed in this course. There are several wonderful resources:

1. G Strang, Introduction to Linear Algebra. Wellesley, *MA: Wellesley-Cambridge Press*, 1993.
2. CD Meyer, Matrix Analysis and Applied Linear Algebra. *Siam*; 2000 Jun 1.
3. S Boyd and L Vandenberghe, Introduction to Applied Linear Algebra – Vectors, Matrices, and Least Squares. Online Book.
4. CM Bishop, and MN Nasrabadi. Pattern recognition and machine learning. *New York: Springer*, 2006.
5. Gelman, Andrew, Jennifer Hill, and Aki Vehtari. Regression and other stories. Cambridge University Press, 2021.
6. McElreath, R. (2018). Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.
7. Online lectures: Linear Algebra by Prof. Gilbert Strang.
8. Selected online lectures: Matrix Methods in Data Analysis, Signal Processing, and Machine Learning, by Prof. Gilbert Strang.
9. Selected online lectures: Linear Dynamical Systems by Prof. Stephen Boyd.

Grading

The final grade for the course will consider the performance on the assignments, quizzes, mid-term, and the final exam.

Assignments	15
Quizzes	15
Mid-term	15
Final	55
Total	100