

Maymester 2024 MATH347

Linear Algebra for Applications in Data Science

An exploration of the mathematics and art of data science

Course syllabus

Times/Credits	Daily, PH385, 5/15 - 5/31, 11:30AM-2:45PM, 3 Credit Hours
Instructor	Sorin Mitran (e-mail: mitran@unc.edu)

Motivation

Artificial intelligence, data science, machine learning are buzzwords arising in academia, business, politics. What do they really mean? What is the knowledge foundation of all the excitement? This is where mathematics comes in. The entrepreneurial mindset is to resolutely transform an idea into action that changes society. The mathematical mindset is to rigorously distill observation, intuition, and aesthetics into an idea. The mathematical idea explored in this course is simple to state:

“What can I build by simple combination of some multidimensional objects?”.

After choosing n objects each characterized by m numbers, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$, the simple combination studied is to resize each object by scale factors x_1, x_2, \dots, x_n and then add them together,

$$\mathbf{b} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_n \mathbf{a}_n.$$

The above *linear combination* formula is about as complicated as the mathematics gets to be in the course, but leads to a treasure trove of applications: balancing a chemical reaction, determining market equilibrium, finding genetic inheritance, analyzing social interactions, identifying faces in a crowd. Solutions to all these problems are found by linear combinations, and linear algebra provides the rigorous framework to determine answers to questions such as:

- can all objects of interest be reached by linear combination?
- what type objects cannot be reached by linear combination?
- if an object cannot be reached by linear combination, what's closest to it?

- are there special objects that do not get significantly changed by linear combination?
- can objects be more insightfully be described by a different linear combination?

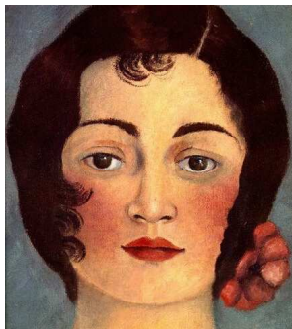
Mathematical concepts are introduced to precisely frame each question above (e.g., *range*, *null space*, *least squares*, *eigenvectors*, *change of basis*), but the technical terms should not cloud the essential simplicity of the questions.

A common feature of applications is that very many numbers are required to describe an object, such that m is very large, while n , the number of objects we wish to combine is small. Once the limits of linear combinations in such cases are determined in *linear algebra*, the natural question is to ask if some other way of combining objects, formally described by some yet unspecified non-linear function f as

$$\mathbf{b} = \mathbf{f}(\mathbf{a}_1, \dots, \mathbf{a}_n),$$

is more powerful. This is the main question within *data science*. It turns out that the relevant mathematics is more complicated and incomplete, and linear algebra is again a useful guide, as exemplified by deep neural networks that serially link several linear combinations by simple nonlinear functions.

The applications of linear algebra and its role as a foundation for data science *arguably* make the subject of greater relevance to today's society than topics such as calculus. There is also an art, a certain aesthetic to statement of linear algebra problems captured by a symbiosis of notation, definitions and understanding of concepts. The course reinforces this link with many examples from the world of art, hopefully leading to an appreciation of the essential unity of the three examples of human ingenuity presented in the table below.



Portrait of a Woman in White, 1930
Frida Kahlo

$$\min_{\mathbf{b}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$$

The least squares problem
David Hilbert, ~1909



Romanian Rhapsody No. 1,
George Enescu

Though it might not be readily apparent all the above involve linear combinations!

Course goals

This Maymester course is intended as a rapid introduction to concepts from algebra that are most useful to data science. Upon course completion students:

- will understand what can and cannot be obtained by a linear combination of objects;
- will recognize the principal problems within linear algebra, i.e.,
 - change of basis (solving a linear system),
 - best approximation (least squares),
 - invariant directions (eigenvectors);
- will become proficient in organization of vector and matrix manipulations by hand;
- understand the role of the most useful matrix factorizations (**LU**, **QR**, SVD, eigendecomposition);
- be exposed to topics within calculus with close links to linear algebra and data science;
- gain the basic practical coding skills in [Julia](#) needed to solve linear algebra problems;
- be exposed to applications of linear algebra outside of the realm of the physical sciences, with an emphasis on examples from art, biology and medicine, and the social sciences.
- understand the role of linear algebra within the wider topics of algebraic structures and data science.

Target Audience & pre-requisites

The course is suitable for a general audience comfortable with basic mathematical abstraction concepts and willing to learn basic coding. There is currently a formal course pre-requisite of MATH232 (Calculus of one variable II), but this is normally waived; connections to calculus concepts are mentioned, but are not required in order to understand linear algebra.

Honor Code

Unless explicitly stated otherwise, all work is individual. You may discuss various approaches to homework problems with students, instructors, but must draft your answers by yourself.

Course policies and organization

- Class attendance is required. Students must bring a laptop that conforms to the [minimal CCI requirement](#) to each class and the final examination.

- Homework is assigned every two lessons. The last third of each daily meeting during this Maymester course is used to start the homework with assistance of Instructor. Completion of the homework should require no more than two additional hours outside of class time. Each homework will consist of 10 theoretical questions (1 point each), and an application to realistic data (8 points)
- Homework is to be submitted electronically through Canvas. Late homework is not accepted.
- Three fifteen-minute quizzes are given on days of lessons 4,7,10 as a self-diagnostic to test basic comprehension of definitions and simple operations.
- The final examination will consist of a first, closed-book part with questions similar to those on quizzes that test understanding of basic concepts, followed by a two-hour, open-book part in which students will use course concepts and laptops to solve a practical problem of complexity similar to a homework application.

Accessibility resources and services. The University of North Carolina at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in barriers to fully accessing University courses, programs and activities.

Accommodations are determined through the Office of Accessibility Resources and Service (ARS) for individuals with documented qualifying disabilities in accordance with applicable state and federal laws. See the ARS Website for contact information: <https://ars.unc.edu> or email ars@unc.edu.

Counseling and psychological services (CAPS). CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: <https://caps.unc.edu/> or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.

Title IX resources. Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Reports can be made online to the EOC at <https://eoc.unc.edu/report-an-incident/>. Please contact the University's Title IX Coordinator (Elizabeth Hall, interim – titleixcoordinator@unc.edu), Report and Response Coordinators in the Equal Opportunity and Compliance Office (reportandresponse@unc.edu), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (gvsc@unc.edu; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.

Grading

Required work

- Homework: 4 assignments, 4 x 18 points = 72 points.
- Final examination: 40 points = 10 True/False quiz questions, 3 x 6 point application questions

Mapping of point scores to letter grades

There is no “grading on a curve”, but opportunities to make up for missed work are provided.

Grade	Points	Grade	Points	Grade	Points	Grade	Points
A+	101-112	B+	86-90	C+	71-75	D+	56-60
A	96-100	B	81-85	C	66-70	D-	50-55
A-	91-95	B-	76-80	C-	61-65	F	0-49

Course materials

Course topics

The course is organized around six basic **questions**, each discussed in two days of the Maymester course schedule. Each question leads to specific **mathematical concepts** listed below. In each lesson the mathematical concepts are applied to **realistic data**, chosen from a variety of fields.

COM. *What tools are needed to work with linear combinations?* **Vectors, matrices, matrix operations, norm, scalar product. Images, electroencephalograms, musical phrases** as vectors and their transformation.

VEC. *What is the mathematical framework for questions about linear combinations?* **Algebraic structures, vector spaces and subspaces, vector set span, range and null spaces, linear dependence, matrix rank, orthogonal matrix.** Data redundancy and **the facial recognition problem.**

THM. *Can we classify objects as reachable or unreachable by linear combination?* **Fundamental theorem of linear algebra, rank-nullity theorem, singular value decomposition. Painter style and motifs,** bases for a large dimensional space.

LSQ. *How close can we get to an object by linear combination?* **Gram-Schmidt algorithm, projection, least squares, data fitting.** Data compression, simplification of complex models from **structural engineering** (reduced-order systems).

BAS. *What happens if we change the objects we combine?* **Linear systems, coordinates, change of basis, Gauss elimination, LU-factorization, determinants.** Construction of bases by greedy data approximation to distinguish **painter style.**

EIG. Are some objects left essentially unchanged by linear combination? Eigenvalues, eigenvectors, characteristic polynomial, repeated eigenvalues (algebraic and geometric multiplicities), the Schur decomposition, spectral expansion, rank-1 expansions. Musical phrases, mechanical vibrations.

Textbook

Class notes specially drafted for this course will be provided as “live” TeXmacs documents (.tm file extension) that contain code for carrying out linear algebra operations and data analysis. Notes are posted prior to class time. Prior to semester start Lesson00, Lesson01 are posted for prospective students to form an idea on course content and approach. Notes are also available for download in Portable Document Format (.pdf file extension) for offline study and printing., as well as Extensible Markup (.xhtml file extension) for web browsing.

Lesson	Topic	LiveDoc	PDFs	Slides	Data
00	SFT	lesson00.tm	lesson00.pdf		
01	COM	lesson01.tm	lesson01.pdf	slides01.pdf	
02	COM	lesson02.tm	lesson02.pdf	slides02.pdf	
03	VEC	lesson03.tm	lesson03.pdf	slides03.pdf	
04	VEC	lesson04.tm	lesson04.pdf	slides04.pdf	ECGData.mat
05	THM	lesson05.tm	lesson05.pdf	slides05.pdf	
06	THM	lesson06.tm	lesson06.pdf	slides06.pdf	
07	LSQ	lesson07.tm	lesson07.pdf	slides07.pdf	
08	LSQ	lesson08.tm	lesson08.pdf	slides08.pdf	
09	BAS	lesson09.tm	lesson09.pdf	slides09.pdf	
10	BAS	lesson10.tm	lesson10.pdf	slides10.pdf	
11	EIG	lesson11.tm	lesson11.pdf	slides11.pdf	

The above notes are also gathered into a traditional [textbook](#).

Homework

Homework consists of direct application of concepts discussed during each lesson, and is mostly completed during class time. Homework is drafted using the integrated mathematical editing and computation facilities of TeXmacs. A tutorial template is provided (hw00.tm) to familiarize students with basic editing and computation procedures.

Nr.	Topic	Problems	Data	Solution
00	tutorial	hw00.tm hw00.pdf		sol00.tm sol00.pdf
01	COM	hw01.tm hw01.pdf	eeg.mat	sol01.tm sol01.pdf
02	VEC	hw02.tm hw02.pdf	faces.mat testfaces.zip	sol02.tm sol02.pdf
03	THM	hw03.tm hw03.pdf		sol03.tm sol03.pdf
04	LSQ	hw04.tm hw04.pdf		sol04.tm sol04.pdf

Software

Modern public domain software systems allow efficient, productive formulation and solution of mathematical models. A key goal of the course is to familiarize students with these capabilities, by extensive use of two applications:

1. [TeXmacs](#), a scientific editing platform, used for preparation of live lessons and to draft homework assignments.

Mac OS installation steps.

1. Download the latest (currently 2.1.4) disk image file TeXmacs-2.1.4.dmg to your Downloads folder.
2. Open Finder, navigate to Downloads, right-click on disk image file and open it. Accept “Unknown Developer” warning.
3. Drag the TeXmacs file to your /Applications folder.
4. On your Desktop, right click the TeXmacs disk image and Eject.
5. In Finder, navigate to Applications, right-click TeXmacs, and Open. You'll get a warning message: “TeXmacs can't be opened”. Close the warning window. Click top-left Apple Icon, go to System Preferences, Security & Privacy. A “TeXmacs was blocked” message should appear, select “Open Anyway”. Close TeXmacs.

Windows installation steps.

1. Download the latest (currently 2.1.4) TeXmacs-2.1.4-installer.exe and execute it.
2. [Julia](#), an open-source numerical and graphical computation package.

Mac OS installation steps.

1. Open /Applications/Utilities/Terminal and execute instructions

```
curl -fsSL https://install.julialang.org | sh
sudo mkdir /usr/local/bin
sudo ln -s ~/.juliaup/bin/julia /usr/local/bin/julia
sudo cp ~/.juliaup/juliaupself.json /usr/local/
```

2. Close Terminal

Windows installation steps.

1. Search for the cmd app and launch it. Execute instruction

```
winget install julia -s msstore
```

3. Extend Julia environment by packages for plotting, linear algebra, interactive notebooks.

1. On MacOS launch Terminal, in Windows launch cmd. Launch julia

```
julia
```

2. In Julia press] key to enter package management environment, and the following packages

```
add Printf Latexify PyPlot LinearAlgebra Revise InteractiveUtils Plots
```

3. Keystroke Ctrl-C to exit package management and then type exit() to exit Julia.

4. Julia plugin for TeXmacs.

Common step. Right-click the zip file [julia.zip](#) and download linked file to your Downloads folder.

Mac OS further installation steps.

Open Terminal and execute instructions:

```
cd ~/Downloads  
cp julia.zip /Applications/TeXmacs.app/Contents/Resources/share/TeXmacs/  
cd /Applications/TeXmacs.app/Contents/Resources/share/TeXmacs/pl  
unzip julia.zip
```

Windows further installation steps.

Open cmd and execute instructions:

```
cd %HOMEPATH%  
copy Downloads\julia.zip %AppData%\TeXmacs\plugins  
cd %AppData%\TeXmacs\plugins  
tar -xf julia.zip
```

5. Launch TeXmacs, select menu Tools->Update->Plugins. Insert Julia session Insert->->Session->Julia

Students are requested to install latest versions of the above two applications prior to first class.

Tutorials

Software usage is introduced gradually in each class, and class participation should be sufficient to gain enough familiarity to effectively use these tools. Some additional resources are also readily available for further study if desired:

- TeXmacs
 - <http://www.texmacs.org/tmweb/help/tutorial.en.html>
 - <https://www.youtube.com/watch?v=mlcqGRv7xhc>
- Julia
 - <https://julialang.org/learning/>
 - <https://www.youtube.com/user/JuliaLanguage/playlists>