

Explanatory Notes for 6.3900

Shaunticlaire Ruiz

Fall 2024

Contents

0	Prerequisites - Explanatory Notes	2
0.1	Multi-variable Calculus (MIT's 18.02)	2
0.2	Vectors and Matrices (MIT's 18.02)	3
0.3	Linear Algebra (18.06)	4
0.4	Programming (6.100A, 6.1010)	5
0.5	Algorithms (6.1210)	5
0.6	Probability	6
0.7	Notation: Sets	7
0.8	Notation: Numbers and functions	9
0.9	Notation: Vectors Spaces	10
0.10	Optional	11

CHAPTER 0

Prerequisites - Explanatory Notes

This course assumes knowledge of several topics. Here, we'll outline them: hopefully, this will make it easier to get up to speed if you have a gap in your background, and to know what you're looking for.

This is designed to be somewhat comprehensive, so if you've taken a class, you can likely skip the corresponding section.

If a class has its own prerequisite, then we assume the understanding of that class as well. For example, we assume you know multi-variable calculus, so single-variable is assumed as well.

0.1 Multi-variable Calculus (MIT's 18.02)

You will need most of the differential aspects of multi-variable calculus:

- The concept of partial derivatives and how to find them
- The **multivariable** chain rule
- An intuition for the gradient as the **direction** of greatest increase in a function, and the **magnitude** of that increase.

It is helpful to be able to visualize a surface created by a function. You won't need to memorize the shapes of specific surfaces; just the 3D intuition in general.

You should also be able to imagine "zooming in" to that surface, and seeing it "locally" as a **plane**, just like how we zoom in to a one-variable function, and see the **tangent line**.

Sometimes, in this class, we will also do derivatives **numerically**, where we approximate the derivative with finite steps, of the form

$$\frac{dy}{dx} \approx \frac{\Delta y}{\Delta x} \quad (1)$$

You should also be comfortable with some basic ideas of "infinity": what happens in the "limit as we approach infinity", for example.

You will not need double/triple/line integrals, curl, divergence, or greens/stokes theorem.

0.2 Vectors and Matrices (MIT's 18.02)

You will need an understanding of vectors:

- You need to know what a vector is, with two interpretations:
 - an **ordered list** of numbers
 - an object in some "space" with **magnitude** and **direction**
- You should know that the **length** or **dimension** of a vector is just how many numbers(or **elements**) that vector contains.
- You should know that a **scalar** is just a number, or in some perspectives, a 1-element vector.
- You need to be able to **add** or **subtract** pairs of vectors. You should also be able to **scale** them (in other words, multiply by a **scalar**).
- You should know how to take the **derivative** of a vector \vec{v} , with respect to a scalar x , in the form

$$\frac{d\vec{v}}{dx} \quad (2)$$

You will also need to understand the **dot product**, and its intuition as the **similarity** between vectors.

You will need to understand matrices:

- You should know what a matrix is, with three perspectives:
 - a 2D grid of numbers (in a "rectangle")
 - an **ordered list** of equal-length vectors.
 - a **transformation** of vectors.

- You should understand the **dimensions** of a matrix, and the common notation (# of rows \times # of columns)
- You need to be able to multiply two matrices, or a matrix times a vector.
 - You need to understand when you are able to multiply matrices, based on dimensions.
 - The dimensions of the new matrix after multiplication.
 - How to do the calculation of multiplying matrices by hand.

You should understand the **determinant**: both how to calculate it, and the intuition behind it.

Finally, you should know what a matrix **inverse** is, and that a **zero-determinant** matrix has **no inverse**.

0.3 Linear Algebra (18.06)

Currently, linear algebra is a **new** prerequisite.

You need all of the concepts mentioned in the "vectors and matrices" segment.

You should also understand **independence** between vectors, the **rank** of a matrix, and how to take the **transpose** of a matrix.

Linearity is a really nice (and important!) property, where a function doesn't get in the way of some simple operations: **addition** or **scalar multiplication**. The order doesn't matter.

Definition 1

For **linear** function/operator \mathbb{L} ,

Addition (of any kind) has the same effect, before or after the function.

$$\mathbb{L}(x + y) = \mathbb{L}(x) + \mathbb{L}(y)$$

Multiplication by a **scalar** also has the same effect.

$$\mathbb{L}(3z) = 3\mathbb{L}(z)$$

Example: The **derivative** is **linear**:

$$\frac{d}{dx}[f + g] = \frac{d}{dx}[f] + \frac{d}{dx}[g] \quad (3)$$

$$\frac{d}{dx}[10h] = 10 \frac{d}{dx}[h] \quad (4)$$

Often, we talk about **operators**. They're like functions, where they have an input and an output. But sometimes, we use this word when we have **another** function as an input.

Definition 2

An **operator** *often* takes in a **function** as an input, and gives another **function** as an output.

Example: The **derivative** is also an **operator**. If you input $f(x) = x^2$, the output is $\frac{d}{dx}[f(x)] = 2x$: another **function**.

An operator doesn't have a really "unique" definition in math: it's used for convenience.

Thus, we can call the derivative a **linear operator**.

The **visual** intuition of a matrix as a spatial transformation is useful in this class.

It would be helpful to understand **nullspace**, **column space**, and **vector spaces** in general.

A good reference for intuition is the linear algebra series by YouTube channel, 3blue1brown! Each video averages 11 minutes, and has been helpful for many past students.

0.4 Programming (6.100A, 6.1010)

For 6.100A:

You should be familiar with object-oriented programming in **Python**: you will be implementing various classes and simple algorithms in this class.

You should have a basic understanding of **time complexity** and big-O notation. Ease with reading basic pseudocode would be helpful as well.

For 6.1010:

Either 6.1010 or 6.1210 are counted as a prereq: they are not equivalent, and neither is individually required to understand the course, but either will make your work in this class much easier.

6.1010 offers more coding experience, which makes the process of implementation much smoother.

Misc:

Prior understanding of numpy is not mandatory, but would be very helpful. Pytorch would also be helpful, but is not used until much later in the course.

0.5 Algorithms (6.1210)

Either 6.1010 or 6.1210 are counted as a prereq: they are not equivalent, and neither is individually required to understand the course, but either will make your work in this class much easier.

6.1210 is about **algorithms**, and thus makes it easier to understand our discussions of different algorithms throughout this class.

Concepts of dynamic programming, complexity, and reading/writing pseudocode are all helpful for this course.

0.6 Probability

You don't need to have taken a full probability course, but there are some core concepts you must understand:

- **Probability** (or chance) is the relative **frequency** of a particular outcome, if you were to run many trials - specifically, it is the proportion of those trials that gave this particular outcome.

For example, if $p = .4$, then you should expect to have that event occur 40 times for every 100, on average.

- Probabilities p are between 0 and 1:
 - If $p = 0$, the event **will not** occur
 - If $p = 1$, the event **will definitely** occur.
 - If $p = .5$, the event has a 50% chance of happening if you try once.
 - Any other probability between 0 and 1 will give some corresponding percentile chance of occurring ($100 * p\%$)
- You can represent probability of event A as

$$P(A) \tag{5}$$

Treating P as a function that returns the **probability**.

- If you want two events to both occur, you can write that with an **and** statement:

$$P(A \text{ and } B) = P(A \cap B) \tag{6}$$

- If you want at least one of two events to occur, you can write that with an **or** statement:

$$P(A \text{ or } B) = P(A \cup B) \tag{7}$$

- The **conditional probability** is the probability of an event **given** that another event has already occurred:

$$P(B|A) \quad (8)$$

This is read as "The probability of B **given** A".

- We can use this to get the probability of two events at the same time.

$$P(A \text{ and } B) = P(A|B)P(B) \quad (9)$$

- Two events are **independent** if knowing the outcome of one event does not affect the odds of another. You can write this as

$$P(A|B) = P(A) \quad (10)$$

You can equivalently use the next bullet point's definition.

- The chance of two **independent** events both happening is their odds multiplied.

$$P(A \text{ and } B) = P(A)P(B) \quad (11)$$

- The **sum** of the probabilities of all outcomes **must add** to 1: otherwise, there's a chance of getting none of the listed outcomes.
- The chance of a particular event not occurring is called the **complement**, and has a chance of $1 - p$.

0.7 Notation: Sets

There are some common definitions and notations you should be familiar with (though they may be introduced if necessary).

If you understand an equation, move on to the next one: each explanation is mostly basic.

Definition 3

An **element** is a single object in a collection of objects.

This definition is often linked to **sets**.

Definition 4

A **set** is a collection of distinct elements with no given order: if you shuffle the elements in a different order, you have the same set.

- We can **define** a set by listing out its elements. For example, this says, "The set A

contains the numbers 1, 2, and 3"

$$A = \{1, 2, 3\} \quad (12)$$

- Shows that an element is **in** a set. The following says "x is an element of the set A".

$$x \in A \quad (13)$$

- Shows that an element is **not in** a set. The following says "x is **not** an element of the set A".

$$x \notin A \quad (14)$$

- Natural numbers**, or the counting numbers.

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\} \quad (15)$$

- Real numbers**, or all of the numbers on the number line (including the full space between integers).

$$\mathbb{R} \quad (16)$$

- We can also define a set by starting with another set, and then listing a **restriction**:

The following says, "Include each natural number n".

$$\{n \in \mathbb{N}\} \quad (17)$$

This seems redundant, but now, we can choose to only include natural numbers **larger than 10**:

$$\{n \in \mathbb{N} \mid n > 10\} = \{11, 12, 13, 14, \dots\} \quad (18)$$

- A set A is a **subset** of B if all elements in A are contained in B. B is then said to be a **superset** of A.

For example, the set of natural numbers is a **subset** of the set of real numbers: every natural number is also a real number.

Here, this says " \mathbb{N} is a subset of \mathbb{R} ".

$$\mathbb{N} \subseteq \mathbb{R} \quad (19)$$

Notice that this symbol looks similar to \leq : that's intentional! This is because the two sets could be the same set, while one is a subset of the other.

- The previous symbol allowed for the two sets to be the same. But if we know they aren't, we can use a **proper subset**.

Here, this says " \mathbb{N} is a **proper** subset of \mathbb{R} ".

$$\mathbb{N} \subset \mathbb{R}$$

(20)

Since we know that some real numbers are not natural numbers, they can't be the same.

0.8 Notation: Numbers and functions

- **Sum notation**: adding up elements in a sequence. For example:

$$\sum_{n=1}^6 n^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 \quad (21)$$

- **Product notation**: multiplying elements in a sequence. For example:

$$\prod_{n=1}^5 n = 1 \times 2 \times 3 \times 4 \times 5 \quad (22)$$

- **Rounding up**: round up real numbers. The following says, "round 2.5 up to the nearest whole number"

$$\lceil 2.5 \rceil = 3 \quad (23)$$

- **Rounding down**: round down real numbers. The following says, "round 2.5 down to the nearest whole number"

$$\lfloor 2.5 \rfloor = 2 \quad (24)$$

- **Function** notation: shows the name of the function, the set of inputs, and the set of outputs.

For example, this below says, "the function f takes real numbers as inputs, and outputs natural numbers."

$$f : \mathbb{R} \longrightarrow \mathbb{N} \quad (25)$$

- If you want to get the **maximum** or **minimum** output of a function, you use the function with the corresponding name: max or min.

For example:

$$\min_{x \in \mathbb{R}} x^2 = 0 \quad (26)$$

$$\max_{x \in \mathbb{R}} \sin x = 1 \quad (27)$$

Below the max or min declaration you can denote the domain over which to find the maximum or minimum, respectively.

- Sometimes, you don't want the minimum or maximum output: you want to know the **input** that gives you the minimum or maximum output. If the domain can be inferred from context, it may be omitted.

So, you pick an **argument** (input variable) and get the **argmax** or **argmin**

The following says, " $x = 1$ **gives you** the minimum output for $f(x) = (x - 1)^2$ ".

$$\arg \min_x (x - 1)^2 = 1 \quad (28)$$

The following says, " $f(x) = 0$ **is** the minimum output for $f(x) = (x - 1)^2$ ".

$$\min_x (x - 1)^2 = 0 \quad (29)$$

Make sure to keep track of the **difference** between min and argmin, or max and argmax!

0.9 Notation: Vectors Spaces

Here, we'll build up some notation for representing sets of vectors, by representing them as ordered sequences.

- Often, we care about **ordered sequence** of numbers. Maybe you want to return the entire sequence.

We start with ordered pairs of numbers: you can represent every pair of elements from two sets with \times .

For example, here we have "every pair of two natural numbers":

$$\mathbb{N} \times \mathbb{N} = \{(1, 1), (1, 2), (2, 1), (2, 2) \dots\} \quad (30)$$

Notice that this can be used to fill in an grid of numbers: with real numbers, you can fill in the whole space with no gaps. Note that since sets do not contain duplicate elements, $(1, 1)$ is not included twice.

- If you want **more than two** elements, you can simply use more **crosses** \times .

For example, here is every trio of natural numbers:

$$\mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{(1, 1, 1), (1, 1, 2), (1, 2, 1) \dots\} \quad (31)$$

- Here, we introduce a shorthand, because writing every cross (example: $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$) can get tiring.

We can compress multiplication with **exponents**, so we'll do the same here:

$$\mathbb{R}^5 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \quad (32)$$

- Because one of our perspectives on **vectors** is as "an ordered list of numbers", we can represent all of our desired vectors using this notation.

In general, the set of all length-n vectors can be represented as

$$\mathbb{R}^n \quad (33)$$

0.10 Optional

Here, we list concepts that could be **helpful** for understanding this course more easily, but are entirely **not required**, as we'll be teaching what we need.

- Tensors
- Convolution
- Examples of Optimization (Least-Squares, etc.)
- Markov Chains
- Probability and Statistics (Expectation, Variance, Distributions...)
- Mathematical maturity (from upper-level math courses, etc.)
- Matrix Calculus