Mathematical Writing

Javier Larrosa

November 7, 2019

Introduction

Mathematicians have developed over the centuries the universal *language of Mathematics*. The main advantage of this language is its *conciseness* and *precision*, which is absolutely needed to communicate technical and scientific knowledge. Its main disadvantage is that it may be dry and hard to digest. However, good writing makes the reading simpler and more pleasant. Good writing allows most of the intellectual energy of the reader to be devoted to understanding the technical issues and not to be wasted on deciphering the writer's mind. Good writing includes choosing the best *notation* and deciding the right *concepts* to define, so that more complex concepts become more natural.

Engineering students learn over the years many non-trivial concepts through mathematical language. Some examples are *definitions*, *theorems* and *properties* from *Math*, *Physics* and *Chemistry* classes. So, they develop the skill of *reading* and *understanding* Mathematics. However, they do not practice as much the skill of *writing*. If you are going to write technical documents containing original ideas you are going to need this skill as well.

As it happens with all other languages, an idea can be expressed in many different ways, and some alternatives may be better than others. As one can imagine, the very concept of good writing is fuzzy and there is no secret recipe for becoming a good writer. Mastering the art of mathematical writing just requires a lot of practice.

Mathematical writing requires careful thinking and a lot of revisions. Our first advise is that a good writer must be willing to consider a change of notation or the introduction of a new intermediate concept if she believes that it will improve clarity. Good writing is a sign of respect towards the readers who are giving their precious time to the writer. All the time spent on improving the writing is time saved by the readers. Be aware that digesting bad writing is a very unpleasant task (every teacher agrees on that), so a bad writer is hardly going to be a good impression even if the content of the document is interesting.

Example 1. *Consider the following three different ways to say the same thing,*

1. *Let C be a set of cities. The population size of a city $c \in C$ is noted $size(c)$. The*

average size of cities in C is,

$$m = \frac{\sum_{c \in C} size(c)}{|C|}$$

2. *Suppose we have a set of n cities each one identified with an index $1 \leq i \leq n$. The population of the i-th city is noted s_i. The average size of the cities is,*

$$m = \frac{1}{n} \sum_{i=1}^{n} s_i$$

3. *Let C be a set of cities where each city is a pair $c = (id, s)$ with id being its name and s being its population. The average size of the cities in C is,*

$$m = \frac{\sum_{(n,s) \in C} s}{n}$$

where $n = |C|$.

Technically all of them are correct (as well as many other variations that you can easily think of). In this simple case, preferring one or another is mainly a matter of taste (personally, I prefer the second one which I find simpler and visually more pleasant). However, if you are writing a document that talks about many other things besides city size and you need to define many other things besides average size one of the options may stand on top of the others. For instance, think about what would happen if we also care about the country where the cities are located and we need country-based definitions (e.g. a given country largest city, or the smallest city of the country with the largest average size among its cities). As more concepts are needed, some of the approaches may produce more complex definitions, properties,... while others may look simpler and more natural. Sometimes we can anticipate and make good notation decisions from the very beginning. Other times we need to take blind decisions and reconsider them as we make progress in our document.

In this lecture we will review the basic rules of math writing and practice the process of going from an unprecise idea in our head to a faithful precise written math specification.

Mathematic Expressions

Mathematic expressions are made of symbols and operators.

Symbols

A mathematic symbol is a letter in a given style. The usual symbol styles are:

- lower case italics: x, y, p, q

- lower case boldface: $\mathbf{x}, \mathbf{y}, \mathbf{p}, \mathbf{q}$
- upper case italics: P, Q, R
- upper case boldface: $\mathbf{P}, \mathbf{Q}, \mathbf{R}$
- Calligraphy: $\mathcal{P}, \mathcal{Q}, \mathcal{R}, \mathcal{X}$
- Greek alphabet (lower and upper case) $\alpha, \beta, \Delta, \Gamma$

Warning: Plain text is never a valid style in mathematics because it is important to distinguish between the English and the Math. (e.g. It is wrong to write that n is the number of elements in set C; it is right to write that n is the number of elements in set C). One exception to this are some well established functions like exponential, logarithm, sinus, etc. Those are usually written in plain text.

Warning: In English, you can write a word in plain text, *italics* or **boldface**, lower or UPPER case and it is still the same word. In math writing, if you define a symbol (e.g. n) the definition does not extend to other fonts or styles (e.g. $\mathrm{n}, \mathbf{n}, N, \mathcal{N}$...). In other words, you cannot say that the size of a given set is n and later on use $\mathbf{n}$ as a synonym because they are different symbols.

Warning: A follow up of the previous warning is that it is advisable not to use the same letter in different, but similar, styles to denote different things, because it may be confusing. For example, it is a bad idea to denote the size of a set of numbers m and their mean m. The rule does not apply between lower and upper case. For instance, it is not unusual to name a set P and an arbitrary element p. However, I would avoid naming a set C and an arbitrary element c because they look much more similar.

Warning: It is important to be consistent with the notation. For instance, I have seen many papers about logic that use upper-case letters to denote formula (e.g. F, G) and many others that use lower-case greek letters (e.g. α, γ). The two options are okay, but never mixed in the same paper.

Upper case has greater *notational entity* than lower case, and calligraphic upper case has greater notational entity than upper case. Additional **notational entity** can also be obtained using boldface. A notation of greater entity is to be used to name more complex elements. For instance, we can name an integer x, a set of integers S, and a set of sets $\mathcal{P}$. It would be awkward to have a set s and use N to denote its size.

Sometimes we need to name several mathematical objects that are very related. To emphasize their relation one can use the same letter with a **modifier**. Modifiers are sub-indexes, super-indexes, accents and priming. These elements can be very useful in some situations, but one should use with care.

Example 2. *I recently read a paper in which the authors defined a problem P which was intractable, so they studied several approximation techniques that were in fact algorithms that solved exactly relaxations of the problem. In the paper they defined these relaxations and named them as $\hat{P}$, $\tilde{P}$, $\check{P}$,.... Subsequently, they defined slices of the relaxations as $\hat{P}_{ii'}$, $\tilde{P}_{ii'}$,... When I was reading the paper, every time that I encountered one of this symbols (in* *the text, inside definitions, inside algorithms,...) my brain had to take a break in the reading and spend some energy deciphering the symbol. As a result, the reading was unpleasant and more tiring than needed.*

It is a good idea to use single letters to name things. For complex definitions (specially functions) words are sometimes a good idea, because they are easy to memorize (e.g. sin for *sinus*, max for *maximum*,...). However, Computer Scientists use words too often (probably because of our habit to name variables and algorithms). So it is good advise to restrict our notation to single letter symbols unless there is a good reason against it.

Some usual naming conventions:

- scalars a, b, x, y, p, q
- indexes i, j, k
- vectors (i.e, fixed-size ordered sequences) $\vec{v} = (v_1, v_2, \ldots, v_n)$.
- In Computer Science we often talk about arrays and lists, which are finite (but not necessarily fixed-size) ordered sequences. Usual notation for arrays is $s = (s[1], s[2], \ldots, s[n])$, and for lists is $l = (l_1, l_2, \ldots, l_n)$
- sets, multisets $S = \{x_1, x_2, \ldots, x_n\}$
- tuples $P = (A, B)$
- functions f, g, h
  - $f : A \longrightarrow B$ (that means $x \in A$, then $f(x) \in B$)
  - $f : A \times A \longrightarrow B$ (that means $x, y \in A$, then $f(x, y) \in B$)

Operators

- Set definition $\{\ |\ \ \}$

  For instance, the set of prime numbers can be defined as,

  $$\{x \in \mathbb{N} |\ \forall y \in \mathbb{N}\ 1 < y < x,\ x \mod y \neq 0\}$$

  where $\mathbb{N}$ is the set of natural numbers and *mod* is the modulo of integer division.

- Minimum of a set $\min S$

  $$\min_{i=1}^{n} \{v_i\}$$

  $$\min_{s \in S} \{s\}$$

  min is undefined if the set is empty. A filtering conditions $f()$ can be added as,

  $$\min_{s \in S, \text{s.t.} f(s)} \{s\}$$

- ...

The Grammar of Mathematics [1]

It may surprise you that mathematical expressions, which are made of symbols and operators, follow the same rules of grammar as English. In an expression, mathematical symbols correspond to different parts of speech. For instance, below is a perfectly good complete sentence.

$$1 + 1 = 2$$

It reads *one plus one equals two*. The symbol $=$ acts like a verb. Below are a few more examples of complete sentences.

$$3xy < -2$$

$$5x \in \mathbb{R}$$

$$9 - s \neq t$$

Can you identify the verb in each one of them?

On the other hand, expression,

$$5x^2 - 10y$$

is not a complete sentence because there is no verb. It should be treated as a noun. When you write math, as well as when you write English, your prose is made of sentences. When you write definitions, theorems, etc, you should take this into account. For instance,

$$P = \{n \in \mathbb{N} | \ \forall_{m \in \mathbb{N}, 1 < m < n}, \ n \mod m \neq 0\}$$

is a sentence. It reads as: *P is a set of natural numbers such that n is in P if and only if every natural number in the range $2..n-1$ does not divide n.* Observe that the sentence is clear and concise as a mathematical expression, but its direct translation into English is a little long and sloppy.

Observation: The equality sign, $=$, is sometimes used as a definition, sometimes as a theorem, and sometimes just an indication of a small step in a reasoning. When you write you should be aware of which is the use you are making. If your are not, probably your writing is not very clear.

Math Talk

Because natural language is easier to read but more likely to contain ambiguities, technical and scientific documents normally combine both English and Math. Good writing means using both languages in parallel and often repeating the same concepts so that the reader can benefit from the best of each world. Redundancy up to a reasonable extent is beneficial because the mathematical notation provides precision while an accompanying natural language description makes it easier to read it and check that the concept has been properly understood. How to effectively combine Math and English will be the topic of another lecture. Here we will focus on the use of Mathematics.

[1] Parts of this section are taken from *A Guide to Writing Mathematics* by Dr. Kevin P. Lee

Examples

In the following we present several examples on how to write definitions. The examples are to be worked in the classroom. Starting from an informal definition given with an example, the goal is to end up with a good written mathematical definition. The students should struggle with finding a good notation, good definitions and finally, a good presentation of the definitions in English.

Example 1

Consider pairs that represent an element and some associated measurement (for instance the first element may be a student identifier and the second its grade at a given exam). Such pairs can be denoted as (a, b). Many pairs can be represented as a set P (note that there are no repetitions). Define,

- The set of failing students,

  $$\{(a, b)|\ (a, b) \in P, b < 5\}$$

  which is similar but not completely equivalent to,

  $$\{a|\ (a, b) \in P, b < 5\}$$

  Can you see the difference?

- The highest score,

  $$\max_{(a,b)\in P} \{b\}$$

- The student with the highest score,

  $$\text{argmax}_{(a,b)\in P}\{b\}$$

  argmax returns an arbitrary element in case of ties.

- The average grade,

  $$\frac{\sum_{(a,b)\in P} b}{|P|}$$

  Note that average grade requires P not to be empty.

Example 1 (variation)

Consider a set of n. Without lost of generality, we assume that each student is identified by an index i in the range $1 \leq i \leq n$. Let r_i denote the grade obtained by student i in a exam. Define,

- The set of failing students,

$$\{i |\ 1 \leq i \leq n, r_i < 5\}$$

- The highest score,

$$\max_{1 \leq i \leq n} \{r_i\}$$

- The student with the highest score,

$$\text{argmax}_{1 \leq i \leq n} \{r_i\}$$

  argmax returns an arbitrary element in case of ties.

- The average grade,

$$\frac{\sum_{1 \leq i \leq n} r_i}{n}$$

Example 2

Suppose that we are analyzing two values associated to a set of cities: their **taxation index** and their **public services index** (health, education, safety,...). Both indexes range from 0 to 10. Regarding taxes, 0 (respectively 10) corresponds to a city with an extremely low (respectively high) taxation load. Regarding services, 0 (respectively 10) represents a city with an extremely low (respectively high) quality of public services.

Generally speaking, we can say that **good cities** are those with a good services index (even if taxes are high). But also cities with low taxes (even if public services are bad). What clearly defines a **bad city** is having high taxes and a bad public health system. In the following we are going to introduce notation and make related definitions.

It is reasonable to assume that we can index the cities. If the number of cities is n, we can use indexes such as $1 \leq i \leq n$ to denote cities.

Let (t_i, h_i) be the taxation and services index of city i (with $0 < i \leq n$). Thus, our data is a multi-set $C = \{(t_1, h_1), (t_2, h_2), \ldots, (t_n, h_n)\}$ (note that a set is not a choice, because there may be repetitions, and a sequence does not seem appropriate because the order seems irrelevant).

If we are interested in measuring how good are cities we can define the **efficiency** of a city as $e_i = h_i/t_i$ (oops! now I realize that I should have indexes in the range 1..10 ("Ooops! inside Ooops" now I realize that I am using the word index in two different contexts, so may be I should call the public services and tax *measurements*). An alternative definition of efficiency would be $e_i = h_i - t_i$. We may think that one city is better than another if it is more efficient. But this way of comparison does not seem very useful if the cities are very different, so maybe I need some distance measure between cities (Ooops! if we want to compare cities, it may be good to add notation so that I can refer to cities $c_i = (t_i, h_i)$).

- The distance between two cities $d(c_i, c_j) = |t_i - t_j| + |h_i - h_j|$

- City c_i is better than or equal to city c_j (with similarity threshold α), noted $c_i \succeq^\alpha c_j$, if $e_i \geq e_j$ and $d(c_i, c_j) \leq \alpha$

Alternatively, we could take a softer approach,

- We say that city i is better than or equal to city j, noted $c_i \succeq c_j$, if $t_i \leq t_j$ and $h_i \geq h_j$
- We say that city i is better than city j, noted $c_i \succ c_j$, if $c_i \succeq c_j$ and $c_i \neq c_j$
- Note that $\succeq$ defines a partial order
- The set of superior cities is,

$$Sup(C) = \{c_i \in C | \; forall \; c_j \in C such \; that \; c_j \neq c_i, \quad c_j \not\succeq c_i\}$$

- Consider subsets of C that correspond to the cities of different countries. This is a **partition** of C. Given the cities of two different countries $P, Q \subset C$, we say that P is better than Q if every city of Q is dominated by some city of P
  - for all $c_i \in Q$ there is some $c_j \in P$ such that $c_j \succeq c_i$
  - (a better definition) $Sup(P \cup Q) = Sup(P)$

Example 2 (variation)

Rephrase everything without using indexes.

C is a multi-set of pairs, where the first component of the pair is the taxation index and the second component is the health index.

Given a city $c \in C$, its taxation index will be noted $t(c)$ and its health index will be noted $h(c)$.

- Given two cities $b, c \in C$, we say that city b is better than or equal to city c, noted $b \succeq c$, if $t(b) \leq t(c)$ and $h(b) \geq h(c)$
- The set of superior cities is,

$$Sup(C) = \{c \in C | \; forall \; b \in C such \; that \; b \neq c, \quad b \not\succeq c\}$$

Observation: probably this notation is better than the other if we do not need to use $t(c)$ and $h(c)$ too much.

Example 3

Permanent of a matrix, applications in counting problems

Example 4

Pseudo-tree

Example 4

Consider that we have a historical database of a population of mice over several years. These individuals live on a semi-supervised environment, so we have no information about their mating habits, but we know an approximation of their date of birth. We are concerned about the possible parents of each individual. Clearly, given an individual not all pairs are potential parents. For instance parents must be one male and another female and their date of birth must reasonably close (for mating purposes) and also with respect to their children (a child must be born when both parents are fertile). Even among pairs that satisfy the previous conditions, some of them are more likely than others given their phenotypes (e.g. a short-tailed mouse is more likely to be descendant of two short-tailed mice) [2]. Therefore, we assume that we have a measure of how likely it is that an individual is a descendant of a given pair of individuals in terms of genetic traits. We are interested in definitions such as the set of possible genealogical trees ending in a given individual, or the most likely genealogical tree for a given individual.

We define some preliminary notation. The range of integers between a and b is noted $I_{a,b}$. Given a directed graph $G = (V, A)$,

Let's assume that the population size is n and each mouse is identified by an index i (with $1 \leq i \leq n$). The gender of individual i, which can be male (m) or female (f) is noted $gdr(i)$. We will consider that time is discrete. Accordingly, the date of birth of individual i is an integer noted $dob(i)$.

The fertility period of individual i, noted F_i, is the interval from $dob(i)+K$ to $dob(i)+K+K'$ where K and K' are constants that depend on mice biology.

(Comment: I used function names $gdr(\cdot)$ and $dob(\cdot)$ aiming at easy memorization. An alternative would have been noting gender as $g(i)$ or g_i, and date of birth as $d(i)$ or d_i. A non-recommended alternative would be gdr_i and dob_i because inside a mathematical expression they may look as multiplications (like $g \times d \times r_i$)

Parenthood Graph

The *parenthood graph* $G = (V, A)$ is a directed acyclic graph that represents *possible* parents of individuals. Graph G has two types of vertexes $V = M \cup P$. Nodes in the set M are indexes and correspond to mice (i.e, $M = \{1, 2, \ldots, n\}$). Nodes in the set P are pairs of indexes and correspond to potential mating pairs. More precisely, a pair (j, k) is in P if

[2] We will make a sloppy interpretation of likeliness since it is not the purpose of this document to enter the complex world of genetics and probabilities

individual i is male, individual j is female and they shared a fertility period,

$$P = \{(j,k)|\ \ gdr(j) = m, gdr(k) = f, F_j \cap F_k \neq \emptyset\}$$

The graph G has two types of arcs $A = B \cup C$. The set B contains arcs from nodes in M to nodes in P, and the set C contains arcs from nodes in P to nodes in M. There is an arc in B linking each node i in M with each node in P containing i. Formally,

$$B = \{(i,(j,k))|\ i \in M, (j,k) \in P, i = j \text{ or } i = k\}$$

Arc $(i,(i,k))$ indicates that ... The set C links parents to potential children (individuals born during the fertility period of both parents),

$$C = \{((j,k),i)|\ (j,k) \in P, i \in M, dob(i) \in F_j \cap F_k\}$$

Possibilistic Genealogy Graph

Graph G represents the set of possible parenthoods over the whole population. Now we are concerned about the possible genealogy trees ending up with a given individual i, which is the portion of G that represents potential ascendants of individual i.

The *possibilistic genealogy graph of mouse* i, noted $G^i = (M^i \cup P^i, B^i \cup C^i)$ is a sub-graph of G defined as follows:

- M^i contains all the possible ancestors of i. Formally, those $j \in M$ such that there is a path from j to i in G. Note that i is trivially in M_i
- P^i contains all the mating pairs that could have generated an possible ancestor of i. Formally, those (j,k) such that: (j,k) is in P, j and k are in M^i and there is a path from (j,k) to i in G
- Arc $(p,(j,k))$ is in B^i if p and (j,k) are in M^i and P^i, respectively, and $(p,(j,k))$ is in B
- Arc $((j,k),p)$ is in C^i if p and (j,k) are in M^i and P^i, respectively, and $((j,k),p)$ is in C

Most Likely Genealogy Tree

A genealogy tree of mice i represents one of the many possible mating combinations starting with first generation mice and ending in i. Formally, it is a sub-tree of G^i that (1) contains node i which is its root, (2) if it contains a node j and j has the root of T^i is i..