

# Notes on Projects in Mathematics at EPFL

A. C. Davison

January 17, 2019

## 1 Introduction

There are three types of mathematics projects during BSc and MSc studies at EPFL. The master project (PDM) is the capstone to the MSc degrees and involves full-time work for an entire semester, which is taken to encompass the exam period and thus contains 17 weeks. Semester projects comprise 9–10 credits and entail 5–6 weeks of work full-time, or around 2 days/week through the semester. Mini-projects are an integral part of certain courses, and may for example correspond to full-time work for around 1–2 weeks. For master and semester projects, it is essential to work regularly and systematically during the entire semester.

A possible timetable is:

**Week 1** Choice of project, sketch of work to be undertaken. Obtain reading list (and maybe data) and begin reading and note-taking.

**Weeks 2–4** Clarify structure of problem and steps needed for project. Preliminary data analysis, if any. Continue reading, take careful notes, and start bibliography.

**Weeks 3–7** Plan structure of report, draft and type early chapters. Continue reading, programming, and data analysis. First results.

**Weeks 7–14** Bulk of work, plus writing in parallel.

**Week 8** (Maybe) intermediate oral presentation (see Becker and Keller-McNulty, 1996, for example) and written report. Parts of the report will be blank, but the structure should be clear, and some chapters and the bibliography should be drafted. Adjust plan and structure, and maybe workrate, in light of feedback.

**Week 13 (or week 15 for PDM)** Only final results and conclusions should remain.

**Weeks 14/15 (or week 17 for PDM)** Final oral presentation and penultimate version of the written report. Minor changes only before the final version of the report is submitted.

The oral presentations and written reports to be presented in weeks 8 and at the end are compulsory, but only the final version counts for credit—the idea is to give feedback early enough to fix any problems and to ensure that a good final mark is merited. However,

**work is given a failing mark, if the presentation—including spelling, grammar, syntax, figures, tables—is unacceptable and/or if the referencing is inadequate.**

## 2 Structure

A typical structure for a report is:

- Front page, with title, author and date.
- Summary: a brief, preferably notation-free, account of the setting and main findings.
- Introduction: background, motivation, context and purpose of the work. Outline of chapter contents.
- Several chapters, usually divided into sections and subsections. Avoid very short sections or subsections.
- Discussion: summary of the main findings, their limitations and possible generalisations; what could have been done with more time.
- Acknowledgements: briefly thank anyone who helped.
- Appendices, if any.
- Bibliography.

## 3 Writing

### 3.1 General

*The difficulty of literature is not to write, but to write what you mean; not to affect your reader, but to affect him precisely as you wish.*

*Do not write merely to be understood. Write so that you cannot possibly be misunderstood.*

*It takes hard writing to make easy reading.*

*When I say writing, O believe me, it is rewriting I have chiefly in mind.*

Robert Louis Stevenson did not have mathematics projects in mind when he wrote the lines above, but they are pertinent to any sort of writing. If you know his most celebrated work, *Treasure Island*, you will have experienced how well he applied them. The first two quotations highlight the importance of precision, of choosing and arranging your words (and, in the case of mathematics, your notation) carefully. Doing this well helps not only your reader but also improves the quality and clarity of your own thought. The last two quotations emphasise that writing well is difficult, so you should start early and be prepared to rewrite often. Ask a friend to read your work when you are satisfied with it, and revise it in light of any criticism. Your default reaction to criticism should be that if your chosen reader does not follow your argument, it should be clarified.

Strunk and White (1979) and Gopen and Swan (1990) are short and very useful guides to writing. The first contains much valuable advice on producing clear and simple English. The second points out that readers are more likely to grasp your meaning if information is where they unconsciously expect to find it, and has practical suggestions on achieving this. A good start is to find an author whose writing you find clear and concise, and to analyse it with a view to emulating it. Don't worry about style; as Matthew Arnold said:

*Have something to say, and say it as clearly as you can. That is the only secret of style.*

Here are some general comments:

- Before you start, consider who you are writing for, and keep this person in mind as you write. This might be one of your class-mates: someone with your mathematical background but without specific knowledge of your topic. In your professional life this person might be a client, a colleague, your boss, or a company board. Whoever it is, your language and the background you assume must be chosen suitably.
- The purpose of writing is to get ideas from your head to your reader's head as directly and precisely as possible, and you should avoid anything that impedes this. Many problems are caused by the use of the passive voice, long words where short ones will do, circumlocutions and elegant variation, heavy or unnecessary notation, and excessive use of abbreviations.
- Incorrect punctuation can obscure the simplest sentence. Read your work aloud to hear where the natural pauses are; this tells you where to place commas, semicolons, etc.
- Good scientific writing in English uses the active voice as much as possible. Using the passive voice typically results in indirect, meandering and unclear sentences. 'Table 1 shows that ...' is simpler and more direct than 'On looking at Table 1 we see that ...', and 'A performs better than B' is preferable to 'A has a better performance than B'
- Check every word, sentence and paragraph and ask: is this essential? Is it clear? Is it in the right place? Can it be shortened without detracting from clarity, for example by rearranging the sentence? Deleting words, especially adverbs, will often strengthen a text. Likewise positive statements are typically stronger than negative ones.
- Check the logic of your argument carefully, using words like 'hence', 'thus' and 'therefore' to indicate the reasoning.
- Say things once. If a section is entitled 'Group theory', there is no need for its first sentence to be 'This section is about group theory.'
- Look out for changes in level, such as explaining elementary notions in detail and then assuming knowledge of more advanced ideas.
- Keep footnotes to a minimum (ideally of zero). If something is important enough to be in the document, it should almost certainly appear in the text.
- Use a spell-checker (and a grammar-checker if your grammar is weak).
- Note the punctuation of e.g., i.e., etc. and cf, and check what they mean.
- For guidance on writing reports, see Ehrenberg (1982) (a bit simplistic), and relevant sections of Chatfield (1988). Higham (1998) and Krantz (2016, Ch. 1, 2)<sup>1</sup> are helpful on writing mathematics. Standard books on English writing style (e.g., Strunk and White, 1979; Gowers, 1986; Williams and Bizup, 2017) are also valuable.

---

<sup>1</sup><https://arxiv.org/abs/1612.04888>

## 3.2 Reproducibility

The essence of scientific writing is that the reader (perhaps a later version of yourself) should be able to reproduce the results, so give enough detail to allow this. Usually there is no need to give every single line of long routine computations or standard proofs, but the main steps should be fully explained. If providing sufficient detail disrupts the flow of the text (e.g., code for some task, or a proof that need not be in the text proper), then place it in an appendix or somewhere accessible online.

If your project contains numerical work, reproducibility will be enhanced if you use R Markdown or another environment that allows you to integrate writing and computation. Of course you should comment your code adequately.

## 3.3 Mathematics

Mathematical text is just text (sentences, paragraphs, sections and chapters) that uses symbols as well as words to convey ideas. The usual rules of grammar, syntax and structure apply:

- sentences with mathematics should be punctuated appropriately. Displayed equations should be punctuated correctly—e.g., if a display terminates a sentence, it should finish with a period (aka full stop).
- When writing mathematics in English it is incorrect to precede every displayed equation with a colon or a comma; if in doubt, look up a guide to English punctuation.
- Symbols should not start sentences.
- Symbols in mathematical writing conventionally use an italic font;  $x$  (`$x$` in `TeX`) not `x` (`x`). Check the entire text to ensure that this convention is maintained.
- Certain phrases that have made their way into mathematical writing are the written equivalents of a speaker saying ‘um’ every few words. Examples are ‘note that’, ‘It is important to note that’ and ‘is given by’. Usually the first two can be cut and the third can be replaced by ‘is’. An exception is the use of ‘Note that’ to avoid starting a sentence with a symbol (as in ‘Note that  $a > 0$  but  $b < 0$ ’).
- The use of ‘we’ (as in ‘Hence we see that ...’) often puzzles readers of mathematical texts. ‘We’ refers to the reader and the author, who are imagined to be working their way through the argument together. Since the reader will not (usually) perform numerical calculations, it is often preferable to refer to tables and figures directly: ‘Table 1 was constructed by simulating .... It shows that ...’.
- Unless you have a very strong reason not to use it, conventional notation will make life easier for the reader. For example, use  $t_1, \dots, t_n$  for times, rather than  $b_1, \dots, b_n$ .
- If you use  $\cdot$  or  $.$  for multiplication, there may be confusion with the decimal point or with scalar product; it may be better to use  $\times$ .
- Arrange brackets in a consistent order, such as  $\{[( )]\}$  or  $\{[( )]\}$ , iterated as necessary. This makes mathematical expressions easier to parse and to check.
- Particularly if you are using material from several sources, beware of changes of notation and look out for notational clashes. In particular, don’t use  $|$  to define sets, as in  $\{a \mid a > 0\}$ , if you also use  $|$  for conditioning in probability statements.

- Check that every symbol you use is defined on first appearance, and remind the reader of its meaning from time to time. For example, if  $W$  is a waiting time defined in Section 2, you might write ‘the waiting time  $W$ ’ the first time you use it in each subsequent section, but just refer to  $W$  thereafter within that section.
- Avoid abbreviations, which lead to sloppy writing. Sentences such as ‘MLE for a GLMM may be performed using the BFGS, NR, CG or EM algorithms, but MCMC is an alternative’ force the reader to spend mental energy on expanding the abbreviations (if she/he knows them at all). Likewise a.s., i.i.d., d.f. and the special symbols  $\exists$  and  $\forall$  increase the cognitive load and thus should be avoided.
- When referring to whole numbers, it is conventional to write ‘one, two, . . . , ten’, and 11, 12, . . . subsequently: ‘ten sets each with 12 elements’.
- Often it is clearer to use the word ‘unity’ when referring to the number one: writing ‘the elements of the vector  $a$  must sum to one’ invites the question ‘one what?’, whereas ‘the elements of the vector  $a$  must sum to unity’ leaves no room for doubt.
- Long or important mathematical expressions should be displayed (i.e., shown on a separate line). Short formulae should generally be left in the text, but must not be more than one line high and not contain reduced-size type. For example  $\frac{dy}{dx}$  should not be left in the text, but should be written  $dy/dx$  or should be displayed on a separate line. Also  $\begin{pmatrix} a \\ b \end{pmatrix}$  must not be left in the text; write  $a!/\{b!(a-b)!\}$  if this must be inline. Otherwise the line spacing becomes ugly.
- Use zeros preceding decimal points: 0.2 not .2.
- Use:  $a/(bc)$  or  $a(bc)^{-1}$ , not  $a/bc$ ;  $x_1, \dots, x_n$ , not  $x_1, x_2, \dots x_n$ ; and  $\sum_{i=1}^n$ , not  $\sum_1^n$ .
- For fractions in displays, use `\dfrac{a}{b}` rather than `\frac{a}{b}`.

### 3.4 Franglais

- When writing in English, Francophone students should avoid French literary style, particularly the over-use of the passive voice, elegant variation and synonyms.
- English has a large vocabulary. Often the translation of a French word most obvious to a French-speaker is Latinate, but there is a pithier and better Anglo-Saxon alternative. Thus, for instance, ‘use’ is preferable to ‘utilise’.
- Some common ‘faux amis’ are:

French term	faux ami	correct English
un ensemble	ensemble	set
une estimation	estimation	estimate
une explication	explication	explanation
expliciter	to explicit	to detail/to explain/to make explicit
exposer	to expose	to present
noter	to note	to denote
un paragraphe	paragraph	section
pathologique	pathologic	pathological
permet de	allows to	allows <i>somebody/something</i> to <sup>2</sup>
une problématique	problematic	<i>no exact equivalent, but could be</i> issue
proposer	to propose	to give
sensible	sensible	sensitive

- In English, Table 1, Figure 2, Chapter 3, etc., are proper nouns and therefore have upper-case initial letters, but are referred to as ‘the table’, ‘the figure’, or ‘the chapter’ when improper nouns.
- Colons (:), semi-colons (;), exclamation marks (!) and question-marks (?) are not separated from the previous word in English. Using `babel{english}` in L<sup>A</sup>T<sub>E</sub>X may help.
- Data are plural; dataset is singular. Thus ‘The data are from ...’, but ‘The dataset is from ...’.

### 3.5 Referencing

- Referencing has two purposes:
  - to give appropriate credit to others, and
  - to allow the reader to check your sources or to look at background material.

The first is essential to avoid charges of plagiarism, and the second is key to reproducibility. Thus you should give references to definitions, theorems, or anything you take from elsewhere. Often it will be enough to begin a section with a sentence such as ‘The definitions and theorems in this section are taken from Coles (2001, Chapter 4)’, but in other cases you will need to reference main results individually.

- Be sure to point out any original material in your project—credit where credit is due!
- Use the L<sup>A</sup>T<sub>E</sub>X `\label` and `\ref` constructions when referring to equations, sections, figures, etc. within your document; it is then much easier to modify your work, and being precise will help your reader. Referring to ‘equation (2)’ or ‘Section 2.1’ rather than to ‘the equation above’ or ‘the previous section’ saves the reader from wondering which of the previous equations or sections is meant.

### 3.6 Bibliography

- Every reference given in the text should appear in alphabetical order in the list at the end of the document. The list should contain only those references cited.
- References to books should be to the latest edition; a page, section or chapter number is nearly always necessary. References to books of papers should include title of book, editor(s), first and final page numbers of paper, where published and publisher.
- Complete lists of authors and editors should be given, with all their initials.
- Google Scholar is useful for finding things, but untrustworthy on the details, so *always* check the details of the original sources. If you cite something, you can be questioned on it, so you should also read the (relevant parts of the) source.
- Check the citation details (author name spelling and initials, title of paper, journal name, volume and pages, etc.) very carefully the first time you enter a reference in your database, to avoid needing to find it again later. Use `{de Haan}`, `{US}`, etc. to force L<sup>A</sup>T<sub>E</sub>X to get names correct or to insist on upper-case where needed.
- Use BibTeX from the start: you will save time overall. There are plenty of packages for bibliography management (e.g., `Jabref` or `BibDesk`).

---

<sup>2</sup>The verb ‘allow’ requires a noun or pronoun immediately after it, as in ‘This equation [subject] allows [verb] us [pronoun] to see [infinitive] ...’. For more details, check ‘infinitive complement’ in an English grammar book.

- Active citation, for instance of Coles (2001), and passive citation (Coles, 2001) use `\citet{}` and `\citep{}` respectively. Check also `\citeyear`, `\nocite`, etc.
- Use an **author-date** or similar citation style in which it is easy to remember what the reference concerns. The BibTeX default style [5] forces even an expert reader to check the list of references repeatedly, whereas seeing Coles (2001) or [Col2001] immediately rings bells for him/her.
- The citation and bibliography styles **natbib** and **CUP** are used in this document—if you use the files from <http://stat.epfl.ch/page571.html>,<sup>3</sup> you should have a bibliography like that on page 9.

### 3.7 Figures and tables

- Every table and figure should be referred to in the text, and should appear as soon as possible after where it is first mentioned.
- Place tables and figures at the tops of pages, so that the text is not broken up, and so that the reader knows where to find them.
- The caption should explain the contents of a figure or table in enough detail that the reader need not delve into the text, but interpretations should appear in the text. There is no need for the text to repeat a description of the contents of a graph (lines, symbols, etc.) or table (standard errors, etc.) given in its caption.
- If you are using `\label` within a float (i.e., a **table** or **figure**), place the `\label` after or inside the `\caption`; otherwise the labelling goes awry.
- Figures:
  - All the panels of a graph (including axis labels) should be large enough to be read easily—try and give the graph a shape that will use the page space well.
  - Check that all the axes are labelled correctly, including the units of measurement. Verify that the caption and the graph agree, that every line and symbol is described correctly (e.g., if the caption says it’s dotted, it’s dotted), and that all lines or symbols in the graph are described in the caption.
  - If the axes of a graph are the same (e.g., both axes show the interval  $[0, 1]$ ), then it should be square.
  - Use a consistent style (same size margins, same font for labels, etc.).
  - If you shrink a figure from full-page size to much smaller, then it will be hard to read when printed—the lines will be too thin and the axis labels too small. Make figures at the size and shape at which you intend to print them in the final document.
  - Cleveland (1993, 1994) is good on graphs and graphics; Tufte (1983, 1990) are also worth reading, but are less practical.
- Tables:
  - Use as few horizontal and vertical lines as possible; they should not dominate the numbers.
  - Check that they cannot be simplified, e.g., by multiplying all the numbers in a column by  $10^3$ , or by representing probabilities 0.01–0.99 as percentages 1–99.

---

<sup>3</sup>Check link

- A common error is to give far too many digits—does the reader need to see all the digits in 3.14159? Perhaps it can be rounded to 3.1, or even to 3.
- If a table contains simulation results, their standard errors should also be given. A compact way to do this is to put a phrase such as ‘The largest standard error for the figures in column 1 is 0.01’ in the caption.
- If a table contains negative numbers, use `$-1$` to get minus signs rather than `-1`, which will give hyphens.

### 3.8 L<sup>A</sup>T<sub>E</sub>X tips

- Don’t try and improve on L<sup>A</sup>T<sub>E</sub>X’s layout, fonts etc. A lot of effort went into them, and almost any change you make will worsen readability, as well as costing you time you might spend improving the contents of your work.
- In addition to the `amsmath` and other standard packages, there should be a definition of fonts and encoding, such as

```
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc} % accents
\usepackage{lmodern} % fix problems with typewriter font
```

- Packages over-write each other, so the order in which they are loaded matters.
- Do not use `\` to end a paragraph, since the next line is not properly indented. A new paragraph should be preceded by one or more blank lines in the `.tex` file.
- Using `~` ties things together. For example, if you write `Section~\ref{sect2}`, then T<sub>E</sub>X will give ‘Section 2’ on a single line, while `Section \ref{sect2}` risks it giving ‘Section 2’.
- Adding `\` after italic or emphasised text or some punctuation marks helps to get the spacing right; for instance, `etc.` and `etc.\` give *etc.* and *etc.* respectively. Owing to the period, L<sup>A</sup>T<sub>E</sub>X thinks that a sentence ended after the first *etc.* and makes a slightly longer space; the `\` avoids this.
- Hyphens - (– in T<sub>E</sub>X), n-dashes – (--), m-dashes — (---), and minus signs – (\$-\$) have different uses. Hyphens are used to join two words, or in the double-barrelled name of a single person (e.g., non-user, Barndorff-Nielsen); n-dashes are used in ranges of numbers or to join the names of two different people (1–7, Neyman–Pearson); m-dashes are used for punctuation, either to make a parenthetical point — like this — or to divide two halves of a sentence, in place of a colon; and minus signs are used in mathematics (e.g.,  $-2$ ).
- Mathematical operators should appear in an upright font, so use `\min`, `\max`, `\exp`, `\log`, and so on, not `min`, `max`, `exp`, `log`, etc., in the T<sub>E</sub>X; we should see  $\max(a, b)$  not *max(a, b)*. Otherwise the reader may wonder whether  $\log(x)$  means  $l \times o \times g(x)$ .
- If you need special operators, such as `\var` and `\cov`, put `\renewcommand{\var}{\rm var}` or `\DeclareMathOperator*\cov{\cov}` at the start of your document or in a separate input file. Then you can easily change them without searching the entire document.
- A similar comment applies to abbreviations of words. If KL is an abbreviation for Kullback–Leibler, for example, write  $v_{\text{KL}}$  (`$v_{\rm KL}$`); even better, define a macro `\renewcommand{\KL}{\rm KL}` and write `$v_{\text{KL}}`. Writing  $v_{KL}$  could raise the question whether you mean  $K \times L$  or the  $(K, L)$  element of an array  $v$ .



- Lists such as  $x_1, \dots, x_n$  (`$x_1, \ldots, x_n$`) use `\ldots`, but sequences of binary operators such as  $x_1 = \dots = x_n$  (`$x_1 = \cdots = x_n$`) use `\cdots`.
- Use `\mathrm{d}` (`\mathrm{d}`) rather than `d` (`$d$`) in derivatives and integrals. Then, should you wish to do so, you can write  $dD(d)/dd$  or  $\int D(d) dd$  without confusion.
- For ironing out T<sub>E</sub>X and L<sup>A</sup>T<sub>E</sub>X problems, see for example Kopka and Daly (2003), Lamport (1994), Knuth (1994), Oetiker *et al.* (2018) or just search online.<sup>4</sup>

## 4 Marking scheme

A suggested marking scheme is given at the end of this document. Note particularly that:

- an interim report whose presentation is inadequate will be returned without comment. It is then up to the student to re-read the suggestions above and to improve their report enough that comments can be given on its contents and structure without wasting time on details of presentation;
- a final report whose presentation is inadequate will be failed without discussion and with no opportunity given for improvement.

The above comments apply particularly but not only to the quality of the English (or French) grammar and expression. It is not our job to teach you how to write grammatically; you should have learned this before arrival at EPFL, or at the Language Centre before starting your project.

## References

- Becker, R. A. and Keller-McNulty, S. (1996) Presentation myths. *The American Statistician* **50**, 112–115.
- Chatfield, C. (1988) *Problem-Solving: A Statistician's Guide*. London: Chapman & Hall.
- Cleveland, W. S. (1993) *Vizualizing Data*. New Jersey: Hobart Press.
- Cleveland, W. S. (1994) *The Elements of Graphing Data*. Revised edition. New Jersey: Hobart Press.
- Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer.
- Ehrenberg, A. S. C. (1982) Writing technical reports or papers. *The American Statistician* **36**, 326–329.
- Gopen, G. D. and Swan, J. A. (1990) The science of scientific writing. *American Scientist* **78**, 550–558.
- Gowers, E. (1986) *The Complete Plain Words*. Revised edition. London: Her Majesty's Stationary Office.
- Higham, N. J. (1998) *Handbook of Writing for the Mathematical Sciences*. Second edition. Philadelphia: SIAM.
- Knuth, D. E. (1994) *The T<sub>E</sub>Xbook*. Reading, Massachusetts: Addison–Wesley.

---

<sup>4</sup>e.g., <https://en.wikibooks.org/wiki/LaTeX>

- Kopka, H. and Daly, P. W. (2003) *A Guide to L<sup>A</sup>T<sub>E</sub>X*. Fourth edition. Reading, Massachusetts: Addison–Wesley.
- Krantz, S. G. (2016) *A Primer of Mathematical Writing*. Second edition. Providence, Rhode Island: American Mathematical Society.
- Lamport, L. (1994) *L<sup>A</sup>T<sub>E</sub>X: A Document Preparation System*. Reading, Massachusetts: Addison–Wesley.
- Oetiker, T., Partl, H., Hyna, I. and Schlegl, E. (2018) *The Not So Short Introduction to L<sup>A</sup>T<sub>E</sub>X<sub>2</sub><sub>ε</sub>*. <https://tobi.oetiker.ch/lshort/lshort.pdf>.
- Strunk, W. and White, E. B. (1979) *The Elements of Style*. Third edition. New York: MacMillan.
- Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, Connecticut: Graphics Press.
- Tufte, E. R. (1990) *Envisioning Information*. Cheshire, Connecticut: Graphics Press.
- Williams, J. M. and Bizup, J. (2017) *Style: Lessons in Clarity and Grace*. Twelfth edition. Boston: Pearson.